

---

# Structuring the Information Commons

---

Open Standards and Cognitive Security



EDITED BY

Scott David, R.J. Cordes, and Daniel A. Friedman

---

# Structuring the Information Commons

---

Open Standards and Cognitive Security



EDITED BY

Scott David, R.J. Cordes, and Daniel A. Friedman

Published by The Cognitive Security & Education Forum (COGSEC) on behalf of the Verified Information Exchange Environments (VIE) Program hosted by University of Washington's Applied Physics Lab's Information Risk and Synthetic Intelligence Research Initiative (IRSIRI), as a part of their participation in the National Science Foundation's "Convergence Accelerator"



[www.cogsec.org](http://www.cogsec.org)



[www.apl.uw.edu/vie](http://www.apl.uw.edu/vie)

The views expressed in chapters within this publication are the chapter authors' alone. They do not necessarily represent the views of editors, VIE Program participants, partners, or funders. The views and findings within this book do not imply endorsement or represent the views of any government or government-affiliated organization, private corporation, university, college, school, institution, or other entity foreign or domestic, private or public.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). You, the reader, are free to copy and redistribute the material, unmodified, in any medium or format for any non-commercial purpose, so long as attribution to COGSEC.org and VIE Program is given. This is a summary of, but not a substitute for, the license itself, found at:

[www.creativecommons.org/licenses/by-nc-nd/4.0/legalcode](http://www.creativecommons.org/licenses/by-nc-nd/4.0/legalcode)



This digital edition, 2022

Includes works cited or consulted for each chapter and appendices where relevant. Citations are in the format chosen by the author, metadata was often produced using reference management software and may contain errors. Choices between American and British English and format, style, and structure vary by chapter, in the interest of preserving the voice of interdisciplinary and international contributors. All images used were generated by the writers, public domain, or attributed under fair use. Some images may be blurred.

ISBN: 978-1-7364269-3-7

# CONTENTS

*Acknowledgements*      *i*

*Editors' Foreword*      *ii*

*I. Narrative Information Management*      *1*

*II. Digital Rhetorical Ecosystem Analysis*      *48*

*III. Swarmcheck*      *104*

*IV. How to Build a Truly Modular System or Organizations*      *116*

*V. Active Inference in Modeling Conflict*      *123*

*VI. An Active Inference Ontology for Decentralized Science*      *179*

*VII. The Synthetic Intelligence Guild*      *223*

*VIII. Elements Related to Maturity of Function in Markets*      *262*

*IX. Estimating Return on Impact of Misinformation Intervention*      *293*

*X. Tracking Public Sensemaking through Rhetorical Annotation of Image Memes*      *310*

*XI. TrustFinder*      *351*

## Acknowledgements

*Thank you to the companies iProov and MATTR for the ongoing feedback, to the communities of practice and event organizers that provided venues for discussion, and to the contributing consultants, academics, professionals, and organizations which provided recommendations, concerns, and writing.*

# Editors' Foreword

There is a tragedy of the commons in the modern global information environment. High quality "information" resources often feel scarce, even as the "data" artifacts of online interactions continue to grow exponentially. The paradox of this global information value scarcity in the face of rapid data expansion is partially due to the unreliability of mechanisms for the sharing of context and meaning of communications on the Internet. Data and communication "cut and paste" affordances have become so easy and accessible that the state of digital discourse can be described as a context-stripping game of data "telephone," where the original context of communication fades away in a potentially infinite series of forwarded communications, repostings, and links. The resulting vacuum of context and meaning of online communications is vulnerable to ignorant negligence and intentional exploitation as naked data is re-used in new contexts yielding new information, but without mechanisms to reveal to the ultimate recipient that any context-shift has taken place.

The absence of mechanisms for the conveyance of context and meaning online has resulted in inadvertent and intentional enclosures (and spoliation) of portions of the "information commons". Events with deleterious effect on the information environment invisibly occur whenever the assertion of context and meaning by a limited number of actors are not subject to a degree of structure, scrutiny, challenge, examination, or counterpoint commensurate with the complexity of the issue at hand.

In historical “commons” involving rivalrous physical assets, the enclosure activity of users (e.g., the harvesting, taking, etc.) is more readily apparent and measurable than in the digital setting. In physical “commons”, the “tragedy” of the commons was due to enclosures and overuse of “the spoils” of rivalrous assets (whether in the form of fish or forest or game or grazing). When one party takes (or encloses) the “spoils” of a rivalrous physical asset, that same asset is no longer available for another party to take and enjoy. The absence of the desired physical asset is typically quite obvious, taking the form of measurable reduction in production levels of depleted farmland, collapsed fisheries, or deforestation.

By contrast, information is intangible, and is generally non-rivalrous, as it can be duplicated infinitely, and the value of the same morsel of information can be enjoyed by multiple parties, unlike a fish or rabbit. For intangible informational assets, a significant source of the functional equivalent of physical “rivalrousness” is the derogation of information value for one party when another party swaps context for a given communication without the knowledge of the ultimate recipient of the communication. Stated simply, the contributions of an opportunist (intentional) or fool (accidental) spoils the informational agora through their false, intentionally misleading, or poorly informed assertions of context and meaning that improperly “enclose” information (by wrapping it in bad context or meaning) to the detriment of the cognitive security of other participants.

Enclosure activity in the information commons is less obvious than similar activity taking place in physical asset commons. Such invisible context or meaning swapping might be done intentionally (as in the case of “dis-information”) or accidentally/negligently (as in the case of “mis-information”). In either event the result can be helpfully analyzed as a form of “enclosure” or “spoiling” of the information commons, affected by the intentional or negligent imposition of a local functional pairing of context or meaning with a given communication of data. Of course, since standards of care and “best practices” have yet to be established for the online activities of pairing data with context and meaning, it is not possible to objectively evaluate whether a given action of pairing is being pursued maliciously, negligently, or with good intention.

In the vast gray area of online interactions, various tactics of rhetoric and persuasion are being practiced for myriad subjective “good” and “bad” reasons (and every motivation in between those two extremes), which compounds the ambiguity of communications and fuels broader existential crises in governance, identity, community, and culture. Reliable and trustworthy communication is a foundation for all these domains. We need metrics of performance to document our expectations for such “context swapping” activities in ways that can be conveyed and shared. Since data accompanied by context and meaning equals information; sustainable, scalable, and successful co-management of the “information commons” (in order to

optimize the functional, aesthetic, and other positive qualities of the pairings of data with context and meaning) requires tools for reliable and transparent preservation and conveyance of data, context, and meaning of online communications.

Indeed, a well-meaning individual attempting to develop an informed opinion on any topic of the day will encounter a deluge of both possibly relevant and contradicting information from good-faith sources, opportunists, flawed thinkers, and threat actors alike. Some of the intentional “context and meaning” swapping online is considered anti-social, unethical, or illegal, while other similar activity is accepted business practice. This flood of information is generally accessed through legally-permitted tools which are designed for the optimization of dwell-time and emotional engagement (and other forms of “rhetorical payload” that entrain user context and meaning preferences and attention) creating commercially exploitable network effects that incidentally reinforce echo chambers, amplify outrage, and reward discord. The speed and ease with which resharing can occur, coupled with the ability and incentives to mask or pervert identity, source, and meaning, exacerbates these network effects, contributing to an increasingly hostile and polarized public discourse. This situation has been titled an “Infodemic” by the World Health Organization, “Information Bankruptcy” and “Failing Trust Ecosystem” by the Edelman Trust Barometer’s 2021 report, and the “Decade of Decreasing Trust” in the Atlantic Council’s 2021 GeoTech Commission Report. Without timely intervention, this situation will only degrade, as deep-fake technology, augmented reality, and digital influence operations methodologies are rapidly becoming more sophisticated and accessible. Given the well over 2 million peer-reviewed articles published per year, growth in reliance on preprints, and the millions of other reports, datasets, and other media being published yearly, not even experts whose professional responsibilities include managing and evaluating information flows are immune from this infodemic.

Recognizing the nature of the online global information challenges associated with “context stripping” of data is a necessary prerequisite to designing, developing and implementing effective interventions. Our suggestion, explored in this volume, is that focus be placed on shared context and meaning as a key element of co-management of the information commons (and as foundational to effectively addressing stubbornly persistent security, privacy, and accountability challenges online). This suggestion broadens attention from data security (a zero-sum setting) to also include context sharing (a non-zero sum setting), and invites consideration of how various historical arrangements (such as the commons, markets, cooperatives, mutual insurance arrangements, granges, and guilds) that have previously been successful in the context of managing assets in the physical world might be applied for co-management and governance of information risks in online settings.



In fact, while this problem of volatility and declining trust in the global information environment initially appears to be novel, when examined as a market undergoing growing pains from rapidly increasing interaction volumes, decentralization, and competition, the setting is revealed to be simply a new instance of an old problem with novel features. Traditionally, solutions to problems of this kind have involved the establishment of shared exchange protocols, tools, and local rule-setting. The historical forms listed above (commons, markets, guilds, etc.) each present examples of how these shared elements have been organized and operated in different historical and cultural settings that gave rise to varying information risk-related challenges. Whatever the differences, these types of organizational solutions all tend to foster one very attractive emergent effect in particular: increased trust when engaging in transactions. That trust manifests in many ways, whether that be the ability to assume you will find a buyer for what it is you're selling (liquidity), receive what is expected (standards), have recourse in the case you don't receive what you expect (judicial function), and can feel confident in your investments of time, money, and attention (risk analysis).

In Fall of 2021, the University of Washington Applied Physics Lab's Information Risk and Synthetic Intelligence Research Initiative (IRSIRI), as a participant in the United States National Science Foundation's (NSF) "Convergence Accelerator", launched an experimental, crowd-sourced research program which framed the global information environment as a proto-market for information in need of exchange protocols and local standards setting in order to de-risk, reduce volatility, and increase trust in digital information transactions. The program aimed to create the conceptual foundation for information exchange-houses, or Verified Information Exchange environments (VIEs), which would increase trust in digital information exchanges through the use of an eclectic ensemble of community-based data standards and information sharing, curation, and research tools with consideration for business, operations, legal, technical, and social (BOLTS) use-cases.

The program made use of the recently formalized methodology known as "catechism-based project management", which applies question-based, versionable project documentation in order to allow for comparability among projects and the ongoing maintenance of alignment of the organization and operation of work being done among participating teams and program stakeholders at scale. This methodology was originally implemented by the Defense Advanced Research Projects Agency (DARPA) and has since been adapted for generalized use. A blurring of the boundaries between the program and participating teams allowed for a free flow of ideas and direct feedback - resulting in many direct collaborations and successful project catechisms (each of which documents a new and independent vector of collaboration among teams). The result of this facilitation and coordination is that in less than a year, insights from a vast selection of theoretical literature and

the experience of professionals across disciplines and domains was synthesized into a body of work to inform tool development.

This volume presents all work submitted and approved for publication through the VIE program in the initial year of the project (during Phase 1 of the NSF Convergence Accelerator Track F in 2021), in order of their submission and approval. Short summaries of the works in this volume are offered below:

**I. Narrative Information Management.** A growing number of fields are engaged explicitly with information management and related system dynamics, and with expanding interaction volumes, few fields can ignore informational challenges completely. Chapter 1, "Narrative Information Management", presents insights from numerous information-oriented disciplines, including command and control systems research, knowledge management, and library science, and explores the information management needs of other domains in an effort to synthesize a comprehensive list of common information management features.

**II. Digital Rhetorical Ecosystem Analysis: Sensemaking of Digital Memetic Discourse.** In the modern operating environment teams and organizations need support for realtime, flexible, scalable analysis of inputs of heterogeneous media data (text, audio, video, and image) that lend meaning and context to communications. Tools that can offer that support require will feature the integration of powerful qualitative frameworks (rhetoric), deep metaphors (ecology), and modern computational affordances (e.g., automated annotation of symbols or narrative). Chapter 2, "Digital Rhetorical Ecosystem Analysis", integrates rhetorical, ecological, and computational approaches in order to propose a new model for analyzing digital content and making inferences regarding hidden states, and to propose the foundation for a supervisory control and data acquisition (SCADA) system for multimedia memetic discourse.

**III. Swarmcheck: Crowdsourced Argument Checking for Improving Rational Public Discourse.** Online discussions are plagued by low-quality information, and both organized and organic bad-faith behavior. The resulting frustration and eroded trust poses risks to public health and national security. Chapter 3, "Swarmcheck: Crowdsourced Argument Checking for

Improving Rational Public Discourse”, discusses recommendations for the development of a system for structuring discussion online, rendering reasoning more transparent and facilitating the reuse of the resulting informational structures across contexts. In addition, it proposes a multi-phase collaboration between the VIE program and the company Swarmcheck, with the intent of enabling interoperability between resulting systems, and among the systems-of-systems they intend to support.

**IV. How to Build a Truly Modular System or Organization.**

Designing for modularity has the potential to enhance organization and operation of systems beyond the field of computer science. However, modularity may be difficult to attain and plan for. Chapter 4, “How to Build a Truly Modular System or Organization”, provides background and insights into the potential gains associated with the implementation of systems-engineering principles in order to achieve and assess modularity within complicated systems.

**V. Active Inference in Modeling Conflict: A Framework for Modeling Conflict in Business, Operations, Legal, Technical, and Social Contexts.**

With significant advances in weapons systems and the introduction of global defense pacts, the risk calculus of triggering an official declaration of war has changed, compelling state and non-state actors to pursue conflict through alternative means. The resulting complex threat surfaces are not always well modeled by existing frameworks for conflict. Chapter 5, “Active Inference in Modeling Conflict”, applies the cognitive modeling framework called “Active Inference” to formalize and model conflict in terms of multiscale processes of communication, trust, and relationship management among multiple agents, which invites consideration of additional and alternative means for intervening and managing conflict.

**VI. An Active Inference Ontology for Decentralized Science: from Situated Sensemaking to the Epistemic Commons.**

The prospect of achieving and maintaining broad systems of structured, decentralized sensemaking and scientific progress with the inclusion of nontraditional organizational structures is as alluring as it is challenging to implement. Chapter 6, “An Active Inference Ontology for Decentralized Science”, explores

how distributed autonomous organizations and similar decentralized structures could be organized and operated to produce traditional research products and participate seamlessly in the larger science, technology, and research ecosystem while avoiding perverse structural incentives and maintaining and enhancing accountability and transparency.

**VII. The Synthetic Intelligence Guild: A Social Technology for a Digital Bazaar.** While large amounts of information are now available to us at any moment, the abundance comes with great costs. Institutions originally developed under a paradigm of information scarcity are now faced with mission-disrupting overabundance; this existential organizational change when accompanied by bad-faith actors and online opportunists, has resulted in individuals being left to themselves to navigate the deluge of information and find trusted sources. Chapter 7, “The Synthetic Intelligence Guild”, proposes the revival of trade guilds, adapted for modern, digital affordances. Specifically, it explores the requirements for the development of underlying infrastructure for a trade guild concerned with Synthetic Intelligence, intended to cultivate tradecraft related to decentralized sensemaking and other information crafts.

**VIII. Elements Related to Maturity of Function in Markets: An Initial Exploration of the Applicability of Market Mechanisms for Solving Challenges in the Information Environment.** The VIE Program’s discussions and interviews with experts on mechanism design and economics featured a recurring theme of “maturity of function” in markets. Such maturity can be measured through a variety of metrics, all of which ultimately focus on a particular market’s capability, as a generalizable system, of coordinating commercial and non-commercial exchange under the pressure of increasing interaction volume. Chapter 8, “Elements Related to Maturity of Function in Markets”, describes some of the key elements gleaned from these discussions that are associated with a market’s maturation journey while managing exchange in the context of exponential increases to interaction volumes, extracted from resources recommended by, and from discussion with, experts across multiple fields.

**IX. Estimating Return on Impact of Misinformation Intervention: An Initial Exploration of the Use of the Business Case for Estimating Return on Investment of Intervention and Incentivizing Information Sharing.** Dense social networks allow for adversarial narrative influence campaigns to have broad impacts with limited effort, creating substantial interest in measuring the impact of various forms of interventions on deterrence. However, there has been limited work considering how to quantify the return on investment in misinformation response activity. Chapter 9, “Estimating Return on Impact of Misinformation Intervention”, explores the potential approaches and challenges of quantifying intervention efforts, including from the perspectives of auction theory, market design, interorganizational public relations coordination, and information management.

**X. Tracking Public Sensemaking through Rhetorical Annotation of Image Memes.** While there is substantial interest in digital misinformation intervention and deterrence, for example through the use of machine learning and natural language processing, a sharper situational awareness, deeper understanding of belief structures, and greater connectivity among teams and organizations is needed to address our information-related crises. Further, despite the increasingly important role of image memes in public discourse, they have proven to be a very difficult class of artifacts to collect, classify, and analyze in aggregate. Chapter 10, “Tracking Public Sensemaking through Rhetorical Annotation of Image Memes”, the challenges of collection and analysis of image memes in the context of interorganizational teams are explored, and requirements for alleviating these challenges are offered.

**XI. TrustFinder: A Community-Based System for Finding Trusted Sources and Evaluating Claims.** The creation of fully operational exchange houses for semantic information products is a daunting challenge for a variety of reasons. Among the most notable of these challenges are (i) granularization and modularity, (ii) meeting both the user and the information products “where they are” in order to avoid the need for universal adoption of standards and storage, and (iii) being able to meet the needs of disparate use-cases across the knowledge economy. Chapter 11, “TrustFinder”, synthesizes the theoretical background, frameworks, and use-case requirements extracted

from discussions with experts and written contributions to the VIE program, in order to present a design document for a sociotechnical system "TrustFinder". TrustFinder makes use of market-inspired controls and open standards, web and document annotation affordances, argument representation frameworks, and crowdsourcing design principles in order to harness the work of global research communities, structuring the environment in order to allow for clearinghouses for conditional exchange of bundles of granular and modular information products.

The work published within this volume represents only a fraction of the program's impact and development. The volume provides glimpses into the kinds of developments that are contributing to the rapid evolution of the information environment. The submission of project catechisms to the VIE program team did not obligate participating teams to publish through this volume - as such, some work influenced by the program has been or will be submitted for publication in journals or as book chapters elsewhere. Further, some projects moved beyond short term publication and gained a life of their own and are now independently proposed initiatives elsewhere or within VIE program's proposed second phase of work. The editors and contributing authors hope and intend that the publication of this volume will help contribute to the efforts being undertaken in myriad communities to measure and improve the levels of trust, authenticity, structure, and meaning in online communications.

# Chapter I

---

## Narrative Information Management

R.J. Cordes, Shaun Applegate-Swanson,  
Daniel A. Friedman, V. Bleu Knight,  
& Alexandra Mikhailova

### Abstract

---

There are many areas of research defined by their interest in information dynamics related to facilitating organizational sensemaking, such as knowledge management, information management, and library science, and many more areas of research, disciplines, and even hobbies which are facing information-related challenges. While all may be concerned with very similar challenges, lack of information exchange and common ontology between these areas may be causing silos, missed opportunities, and potentially even friction among areas. In this paper, we address the need for synthesis and exchange of knowledge, tools, and approaches among various fields by proposing Narrative Information Management (NIM) as a unifying term and framework for the fundamental features and challenges of facilitating collective sensemaking. Through this framework, we offer an initial common set of features of impactful information systems found in literature on information-focused disciplines, such as knowledge management, and explore what insights and ad-hoc solutions may be found in an eclectic set of fields facing information challenges, including personal finance, ancestry research, hybrid cloud infrastructure security, translational neuroscience, and genomics. Finally, we offer recommendations for future research.

---

Narrative Information Management was originally published in the COGSEC 2021 volume "Narrative Information Ecosystems: Conflict and Trust on the Endless Frontier".

# Introduction

When the brain cannot reduce the complexity of the environment, it reduces the complexity of the strategy used to make sense of it [1–7]. This difficulty in reducing the complexity of a given information environment is often referred to, depending on context, as either data overload [8,9], reference overload [4], information overload [5,9,10], or, more broadly, as cognitive overload [3,11,12]. The volume, density, and structural complexity of information has impacts on cognition beyond increasing time-to-insight [1,3]. Unfortunately, simply providing more information as a basis for improving decision-making and sensemaking may make outcomes worse rather than better [3,7]. When an individual is exposed to potentially relevant yet contradictory information at a rate inconsistent with the time and effort required to integrate, and does not have access to appropriate tools, a trusted network of experts, or domain-specific training, they may withdraw from their role in the environment or experience anxiety and reduced ability to manage stress, set priorities, make decisions effectively, and detect logical inconsistency [1,3,6,8,13–16]. Failures of individual cognition and decision-making can lead to cascading errors in systems, highlighting the importance for understanding the nature of these informational pathologies and how to avert them in modern settings [17].

In this paper, we highlight the need for synthesis and exchange of knowledge, tools and approaches among various fields concerned with addressing these sensemaking challenges through the framework of Narrative Information Management (NIM). First, we present a broad summary of the challenges faced by information-centered disciplines such as knowledge management. Following this summary, we consider the value of using NIM as a unifying category of features, or functions, within information systems used or designed by these disciplines. We then synthesize a set of common features which contribute to effective NIM systems and consider how they can be understood from a NIM perspective. Next, in the interest of discovering additional feature needs and requirements which may not be well-recognized within information-centered disciplines, we explore an eclectic selection of disciplines that, while not primarily focused on information dynamics, are increasingly experiencing informational challenges. These disciplines include retail finance, amateur ancestry research, genomics, neuroscience, and hybrid cloud infrastructure security. In each of these areas, insights about requirements and the domain-specific challenges and ad hoc solutions for NIM are considered. Finally, we conclude with a discussion assessing common features found and discovered amongst the discussed domains and with recommendations for future work on NIM.



# The Past and Present of Solutions to Cognitive Load

Throughout human history, solutions designed to reduce cognitive load and facilitate individual and organizational action have emerged as a response to increases in local information complexity. Broadly, human action-oriented sensemaking can be seen as a type of narrative inference, where individuals are able to act appropriately to the extent that they have identified the story they are in and role they play [18,19]. Domain-specific approaches to sensemaking have also been developed. In economics for example, mechanisms for externalization, abstraction, and communication of financial information emerged in response to the numerous explosions in economic complexity caused by the opening of new trade routes [20–22]. In science and scholarship, changes to methodology and tools for research and the maintenance of doctrine have traditionally followed paradigm shifts in science as well as sociotechnical changes such as increased volume and accessibility of research publications (e.g., such as those caused by the introduction of the printing press) [2,23–25]. Changes to the scientific process and research methodologies are not just lagging indicators of change to publication systems – historically, the development of information management systems has resulted in shifts in how information is synthesized and communicated. For example, the first reference management systems and formalized cartographic procedures were generated at the Library of Alexandria and funded by its stakeholders in order to process and exploit an unprecedented flow of information and new discoveries [26–28]. Finally, in military operations, documentation and intelligence processes and tools have consistently been adapted and updated in response to increased complexity in geopolitics and mobility in the battlespace [29–32].

The introduction and continued development of digital communication and storage technologies have caused changes in the accessibility, communication, structure, presentation, and production of information at a historically unprecedented rate [33,34]. The challenges and opportunities presented by these new technologies have illuminated the need to reduce cognitive load and facilitate sensemaking. The need for research in this domain will only continue to grow as these technologies develop and increase in informational complexity and volume in the coming years. Nearly 60 zettabytes (60 trillion gigabytes) of data were created in 2020 and the expectation is that the amount of digital data created between 2021 and 2025 will greatly exceed the cumulative amount created since the advent of digital storage [35,36]. Data sets alone and in any size can overwhelm analysts if data are ambiguous, inaccurate, structurally complex, or require specialized analysis. Additionally, transdisciplinary projects for small teams as well as larger organizations require groups of analysts to come to a shared operational understanding of the topic, potentially involving

significant data engineering, modeling, and analysis. For example, with over 7,000 peer-reviewed scientific and engineering articles and countless preprints, datasets, and other relevant materials being published each day, academics and researchers are prone to a state of information overload without the presence of big data dilemmas [37–40].

Unlike past paradigm shifts in information dynamics, where only certain groups such as generals, government officials, or employed scholars were faced with significant demands for adaptations to these changes [26,27,29,31,41], broad adoption of digital information technologies implies that the majority of organizations and citizens, outside the context of any particular discipline, are now in need of tools to overcome challenges related to managing streams of digital information and reducing informational complexity [16,42–46]. Now in the throes of the COVID-19 pandemic, not even children are spared of the need to spend additional effort on narrative sensemaking [47]. The timelessness of challenges related to sensemaking, paired with their distinctly-different application across sectors, means that research addressing information overload has the potential to become siloed and disconnected due to differential usage of keywords, citations, and types of deployed systems [42].

There are already many formalized fields of research which focus on how to design and implement systems, protocols, and procedures to store, manage, communicate, synthesize, curate, and search digital information to help manage the cognitive load of users. Significant examples of interacting fields and topics include knowledge management, information management, and library science [42]. Modern organizations operating in information-rich environments look to these information-centered fields for the solutions that they influence, design, and implement in the interest of reducing cognitive overload. For different users in different scenarios, such sensemaking tools might assist in maintaining situational awareness, facilitating reduction in information complexity, navigating users toward effective action, or the creation, sharing, use, attribution, synthesis, and management of intelligence and knowledge products. As the volume and structural complexity of the available or presented information increase, systems in this category tend to shift from a facilitating role to being essential to operations. In such cases, the usefulness of a given system can be related to its efficiency in helping users meaningfully aggregate data, develop understanding, and navigate toward action, as opposed to simply being tied to the provision and access of information [1,48–51].

Knowledge management, information management, and library science are representative examples of fields which have information dynamics as a primary focus; however, these are not the only fields concerned with information dynamics [42]. There are many other areas of research, disciplines, and even hobbies which require attention to theory and implementation of information-related systems and

data-rich processes. Solutions for domain-specific or even generalized sensemaking may arise within these areas, potentially drawing from the literature within the fields listed above, or using tools reflecting these fields. However, this relationship may be one-sided between information management in the general cases, and domain-specific applications: various fields may draw tools and frameworks from the informational sciences, but rarely translate their feedback or requirements back to the informational sciences. This disconnectedness may cause failures to communicate insights and implementations across areas of theory and practice, leading to further siloing, confusion, and disconnection [42,52]. Recent analyses have suggested that even the fields which share information dynamics as a primary focus show only partial bibliographic and theoretical overlap, reflected by divergent ontologies and professional scope [42,53].

The fields and specializations which are primarily focused on how to design and implement systems, protocols, and procedures to store, manage, communicate, synthesize, curate, and search digital information are numerous and divergent, and have been for centuries. For example, by 200 AD the Roman Army had formalized many roles associated with management of information, including *interpretes* (interpreters who worked to archive translations of written and vocal communications), *librarii* (archivists), *notarii* (secretaries and records managers), *exactii* (recorders and scribes), *exceptores* (short-hand recorders and scribes), *frumentarii* (messengers and information collectors), *quaestionarii* (human source development), and *spectulatores* (information collectors), each representing a formal discipline with its own specialized training [31]. By roughly 1100 AD, the storage, access, synthesis, sharing, and curation of documents, records, and knowledge held within libraries was considered a formal science in China with overlapping sub-disciplines [54]. As noted earlier, the introduction and development of digital storage and communications technologies has meant that modern organizations and individuals are contending with increasing information-related challenges. As sensemaking processes diverge across fields, there is a higher potential for divergent ontologies to develop and siloed practices to occur. It may be time for synthesis and generalization of the underlying sets of challenges and requirements within these myriad domains in order for research and solutions to become more easily discovered and integrated, as well as to prevent redundant research [51]. Here we offer a brief summary of 3 categories of divergent, information-centered fields and areas of research.

**Meta-Information Fields.** The term meta-information fields is used here to describe the category of fields which are concerned with information flows and use in general, with no defining interest in any particular field. In this category are the fields of (1) knowledge management, (2) information management, (3) information engineering, (4) records management, (5) document

management, (6) archive management, (7) reference management, (8) data, information, and sensor fusion systems, and (9) information resources management. For example, knowledge management refers to the design, implementation, and study of processes and systems related to creating, sharing, using, attributing, synthesizing, and managing the knowledge and information of a group or organization in order to improve situational awareness, decision making quality, knowledge transfer between organizational components, and productivity [42,46,55].

**Interdisciplinary Information Fields.** The term interdisciplinary information fields is used here to describe the category of fields which are concerned with the provision and design of information systems which are intended for use in some common category of disciplines. In this category are the fields of (1) library science, (2) intellectual capital management, (3) relationship management systems, (4) decision support systems, (5) case management systems, (6) situation awareness systems, and (7) intelligence management. For example, library science is primarily focused on providing features and insights for the management of documents within organizations whose primary purpose is to lend and manage information resources [56], and intelligence management is concerned with the protocols and procedures that facilitate situational awareness and the creating, sharing, using, attributing, synthesizing, and managing of relevant intelligence products and information streams in law enforcement, military and intelligence, and manufacturing and industrial settings [57–60].

**Application-Focused Information Fields.** The term application-focused information fields is used here to describe the category of fields which are concerned with the provision and design of information systems which are intended for use in a specific discipline. In this category are the fields of (1) command and control systems, (2) intelligence, surveillance, and reconnaissance systems, (3) intelligence fusion systems, (4) asset management systems, (5) supervisory control and data acquisition (SCADA), (6) security management, (7) business intelligence systems, and (8) learning management systems. For example, an asset management system is a set of protocols and procedures tied to software which facilitates situational awareness of, decision making related to, and the planning and controlling of financial

assets, relationships between assets, and asset-related activities [57], and SCADA researchers are primarily interested in providing information tools to organizations which have to remotely monitor and intervene in mechanical or industrial systems [58].

Instead of focusing on simply storing, moving, reading, and writing bytes of data, these information-centered fields are concerned with the facilitation and meaningful direction of data-transfer. A formidable gap exists between the raw syntactic inputs provided by information databases and the semantic or action-oriented representations that an end user might expect to receive as a result of an interaction with the system [51]. Even records and archive management, which might rightfully be assumed to be primarily about storage processes, are equally concerned with the nature of access and user dynamics [59–62]. This focus on facilitating semantic interactions with humans helps distinguish these areas from disciplines like computer science and from meta-disciplines such as information science, which may include within their scope both consideration for use-cases and practical aspects of digital transfer and transformation of information [63,64]. It also reflects one of the earliest maxims from the oldest of the information-centered fields, library science: “Libraries are for use” [54].

Systems that are influenced, designed, and implemented by information-centered fields have disparate use-cases; however, many integrated sensemaking systems can be generalized, or reduced to parts that can be generalized. We identified several features commonly used in information management across domains, such as search, curation, situational awareness, and predictive analytics. While essential within subdomains, these common features already represent generalized areas of research of their own, rather than generalizations of the ensemble of features, emphasizing the exponential expansion of domain-specific information burdens. Here, in addition to the various other framings for integrated sensemaking, we propose Narrative Information Management (NIM) as a term to both unify the common features of these many information-centered disciplines and provide a lens through which to consider their requirements and development. Where narrative information in other situations may refer specifically to the information contained in a given narrative, for example a book or self-reported experience [65], we intend for NIM to refer to the management of information in the facilitation of narrative sensemaking.

# Narrative Information Management

Narrative has received many definitions, and in some cases these definitions contradict [66]. Where there is consensus, there is often some ambiguity regarding scope that parallels analogous debates in memetics (e.g., what isn't a meme/narrative?, is this a single meme/narrative or a cluster?, is this a meme/narrative or a component of one?) [66]. However, even where narrative has been labeled a “buzzword”, there is agreement that it practically represents story, patterns of expectation, plot, and sequence patterns, that it is encoded and decoded through stories, images, symbolism, and metaphor, and that this encoding represents internalization which impacts how humans integrate, store, compress, and communicate information and navigate moral, physical, and social terrain [67–69]. Many examples exist of narrative-driven approaches in various domains attempting to differentiate from scientific-, evidence-, or data-driven approaches, usually focusing on the use of what would traditionally be defined as a “story”, such as the use of fictional or real accounts of events in order to influence behavior as opposed to leaning on data or evidence [70,71]. Attempts to define narrative usually provide similar differentiations between narrative and other forms of communication, some in poetic fashion:

*“Science explains how in general water freezes when (all other things being equal) its temperature reaches zero degrees centigrade; but it takes a story to convey what it was like to lose one’s footing on slippery ice one late afternoon in December 2004, under a steel-grey sky.”*

[72]

However, the line between these forms of communication (science and story) present in the quote above is inherently subjective [66,73] and there is a reasonable argument to be made that the scientific explanation is simply a narrative constructed from interpretations of scientific data and that the explanation through story is a narrative constructed on common experience and metaphor [74]. Further, raw data in any sufficient volume fails to communicate anything meaningful without visualization, descriptive statistics, and presentation—all of which are used to allow different components of the data to “tell a story” [75]. “Nobody walks into a bookstore and asks for a narrative” [66] but it could be argued that nobody walks into a book store without one, as one has to have internalized some set of stories about book stores and what they provide in order to consider shopping there as an option.

While certain disciplinary approaches have been interpreted as being "free" from narrative (e.g., objectivity in the sciences), it has been argued that these are professional narratives about objectivity that serve to reduce cognitive load and facilitate sensemaking in complex, information-rich environments; although such simplifications may not always be helpful [76]. It has also been argued elsewhere that formal documents such as instruction manuals, medical records, project documentation, and historical documents being categorized as narrative-free or not being meaningful in the construction of narrative is largely up to interpretation, presentation, and context—especially where these kinds of media create expectations for navigating the world and taking action [32,65,73,77,78]. Broadly, action-oriented sensemaking can be seen as a type of narrative inference, where individuals are only able to act appropriately when they have identified the story they are in and role they play [19,78]. Frameworks from cognitive science, such as active inference, are increasingly considering psychological, cultural, and narrative aspects of individual decision-making [79,80]. In such frameworks, narrative inference is cast as an ongoing process by which agents estimate hidden environmental states (variables that are not directly observed but bear strongly on how observations change through time). Estimation of narrative state variables can reduce uncertainty about future outcomes. For example, knowing that one is watching a movie in the romance genre as opposed to horror, would reduce one's uncertainty about the relationship status that the characters might be in at the end of the film and what actions they may or may not take.

While narrative frameworks and approaches have been dismissed by some as too theoretical, passing fads, or superfluous cognitive layers [81], their utility should not be underestimated. A core function of the human brain is the detection of event boundaries in order to construct and maintain episodic memory [82,83]. Studies have shown that areas of the brain related to narrative comprehension are active when segmenting events [82,84,85], indicating that narrative structure is not an extraneous layer that we apply to experience, but instead anchors our perception of reality. This has led some to synthesize features of episodic and semantic memory as a single area or subcategory referred to as "narrative memory" [86,87]. Similar work on narrative comprehension has led others to characterize large portions of human sensemaking as a function of "narrative intelligence" [88,89]. If the brain's sensemaking about the world is, at its core, structured around narrative, and if knowledge management and similar systems aim to scale sensemaking from individuals to groups, then the role of narrative in developing shared understandings cannot be dismissed. Further, if the study of narrative provides tools and frameworks for communication, reduction of cognitive load, and extraction of meaning, then narrative study may be of use in generalizing aspects of systems which facilitate meaningful communication.



## Features of NIM Systems

Below, we describe features common to the systems and processes employed by information-centered fields, which generally reduce cognitive load and facilitate sensemaking, thereby helping to manage and communicate narrative.

### Managing Information Gaps

Discovering and handling information gaps is a key feature of many information systems for a number of reasons. In learning management systems, finding and filling knowledge gaps is not just a challenge, but often the reason for their implementation—as learning management systems assist learners in discovering and managing prerequisites to new competencies [90]. In knowledge and intellectual capital management, knowledge and resource gaps are seen as a primary challenge but also as an opportunity to build new knowledge [91]. When making decisions under uncertainty in industrial, commercial, military, and intelligence settings, command and control, information fusion, business intelligence, intelligence management, and decision support, systems are used to rapidly identify where more information is needed or where information needs to be verified or integrated cautiously [92,93]. In archive, records, and document management systems, the faster a document can be identified as missing or missing pieces, the more likely it is that it can be repaired or found [94–96].

Narrative itself has been described as a “dynamic system of gaps”, where well-structured written stories manage information gaps strategically and efficiently—to build suspense, to prompt the reader to focus their attention on details, and maintain engagement [97]. Narratives help form expectations for patterns in and across classes of systems and event sequences, acting as a tool which helps facilitate the agent in directing their attention to areas needing further investigation, where to expect surprise or uncertainty, or where they will simply have to cope with the absence of information [32]. Frameworks built from research on narrative and scenario structure have been used to define and frame information expectations, project documentation, and document annotation needs [32], and could be broadly applied to any system which manages information gaps. For example, signals about gaps in expectations within the lifecycle or typical “stories” of a document's use and transformations can reveal potential tampering [98] or help to identify linked documents that may be missing [96]. In addition, media communicating personal experiences, case studies, or reports of types of professional tasks and encounters can also be used in a variety of use cases, such as helping to fill gaps in tutorials and formal descriptions as well as to help contextualize events or use of knowledge [99–101].



## Facilitating Situational Awareness

Situational awareness is an explicit and primary feature of interest within the domains of command and control, situation awareness systems, intelligence management and fusion systems, security management systems, SCADA, and sensor fusion systems, but due to divergent ontologies, often goes unmentioned in areas such as knowledge and information management. For example, in records, document, intellectual capital, and archive management, knowing who should have access and who has access to documents or materials is a vital feature [102]. Moreover, in knowledge and information management systems user awareness of potential bias in curation systems helps manage expectations [103].

There is a general consensus that multiple factors are necessary to reliably measure situational awareness [104–106], and these factors could be reduced to a smaller set of key components when considering the agent's goal orientation within a given operating environment. The factors to consider in measurement of situational awareness include (1) perception of the components and processes within an operating environment—that the agent can recognize the phenomena, agents, or collections of agents which are relevant to the current situation [105–108] (2) awareness of the spatial, mechanical, and abstract relationships between environmental components [48,108], (3) temporal awareness—awareness and knowledge of sequences of events occurring within the operating environment and in past scenarios [105,108] (4) communicability—how easily the information about the environment can be synthesized and communicated to others [105,107,109] and (5) projection and prediction—how well an individual can synthesize and fuse information about the situation and tie it to similar cases in order to project what is likely to happen next [104–106,108,110].

The use of narrative frameworks in facilitating, measuring, and understanding situational awareness in myriad contexts requires no exhaustive argument, as this has already been done elsewhere over the course of the last 40 years [111–116] however, a brief summary of insights is warranted. The study of narrative comprehension is robust due to the varied research interests which include it as a key measure, such as the cognitive development of young children [117,118], empathy and theory of mind development in adolescents [119], reading comprehension in educational settings [120], and cognitive decline due to disorder or aging [121].

Reframing situational awareness under the same umbrella as narrative comprehension would allow both areas to benefit from generalization and otherwise siloed research. Situational awareness research tends to prioritize raw knowledge of the environment, as opposed to filtering and comprehension in complex information-rich environments [108]. Given that narrative comprehension consists

of components which are nearly identical to those of situational awareness, provides frameworks and ontology (e.g., plot, setting, character archetypes) for comprehension of those components, and intends to address many of the same challenges posed to situational awareness [106,108,122,123] the likelihood of benefit from generalizing the challenges and requirements of situational awareness within narrative frameworks is quite high.

## Providing Descriptive and Explanatory Information

The provision of descriptive and explanatory information about systems of interest is essential. Rapid provision of descriptive information is an area of rich overlap between the most disparate of the information-centered fields described, such as intellectual capital management and command and control systems [42], where the ability to acquire more information about a particular object and its place in a system becomes a highly generalizable feature. Some systems may have more need for explanatory information than others, such as in IT-related knowledge management and decision support systems, where addressing why a particular event may be occurring is essential to addressing the event itself [124], but all may benefit from providing access to a deeper explanation about resources or components (e.g., how was this data produced?) [125,126].

Past work on narratology and the management of narrative information fits explanatory and descriptive information to patterns and formats which can help the brain parse or construct a story in the absence of traditional storytelling structure [65,111,113,115]. These methods, such as knowledge graphs, can be used in conjunction with situational goal-orientation in order to reveal those elements of incoming information which matter most [113,115], thereby reducing the information load on the user:

*“When a reader summarizes a story, vast amounts of information in memory are selectively ignored in order to produce a distilled version of [a] narrative. This process of simplification relies on a global structuring of memory that allows search procedures to concentrate on central elements of the story while ignoring peripheral details.”*

[115]

Framing the provision of descriptive and explanatory features under the domain of narrative frameworks and ontology may allow for new avenues to handle challenges posed by information systems which need to be context aware (e.g., role, goal-orientation, and mission awareness) in order to avoid triggering scope creep (continuous or uncontrolled growth in a project's scope), unintended access to

resources, and/or overwhelming (or underwhelming) the user with information [127–129].

## Facilitating Exploration

Exploration of an information environment with high structural complexity and a large volume of resources is similar to any other kind of complex work in that it leaves teams “susceptible to scope creep because new opportunities, interesting ideas, undiscovered alternatives, and a wealth of other information emerges as the project progresses” [130], resulting in the fundamental explore-exploit dilemma [131–133]. In most information-centered fields and the systems they design and provide, the user’s ability to explore beyond their known unknowns and forage for unexpected information in novel locations is an obvious feature, even in records or archive management where the usual use-case is mundane access and retrieval of documents [98].

The ability to traverse beyond known unknowns unfortunately comes with a number of consistent challenges. Chief among them is the fact that each exploratory step constitutes both a context shift and expansion, accompanied by the risk of fatigue and scope or mission creep [134]. Further, both risk and success in exploration are difficult to measure, which is why explore-exploit maintains its position as a fundamental dilemma [131,134]. Narrative approaches such as the use of thematic maps [135], narrative archetypes [136], and the ability to review side-by-side comparisons of narratives about similar or the same events [137] have been proposed as approaches to remedy these challenges, as they may help frame what should be explored or what is missing from current analyses, thereby calibrating and improving precision in exploration. Of particular interest are tools which help the user construct a narrative about their own exploration beyond a simple search-history. Narrative construction tools could help the user form timelines and annotations about their “expedition” which enable the rapid recollection of the location of information, the selection of appropriate tools for the job, and facilitate the integration of their findings [137].

## Compression: Visualization, Structure, Collation, Curation, and Interaction

All information-centered disciplines, either implicitly or explicitly, abstractly or concretely, have to contend with the need to compress information by merit of their need to communicate it. As the volume of relevant and necessary information increases, “the trade-off between ‘relevance’ and ‘intelligibility’ becomes akin to Heisenberg’s Uncertainty Principle: as one becomes more precise, the other becomes dangerously less so” [138], especially under time pressure [109]. The ability to balance this tradeoff between relevance and intelligibility is essential for

facilitating exploration and situational awareness. Information systems make use of a number of compression mechanisms available for reducing cognitive load in order to allow for intelligibility of the information environment while still including as much relevant information as possible:

**Visualization.** Though auditory cues can be of value [139] and some users may be more verbally focused than others [140], human beings primarily forage for information through vision [141]. Proper visualizations can facilitate or even enable the communication of enormous amounts of information that would otherwise be intelligible [75,142]. Designing systems that are visually informative about complex information, while also accessible to users with visual limitations, remains a significant challenge across areas [51,143]. Visualization does not necessarily refer exclusively to graphics and charts, though the strategic placement of text without multimedia content can facilitate more rapid parsing and stronger retention [137,139]. Text can also be strategically placed with multimedia content in order to trigger effects such as the temporal contiguity effect (better information transfer when relevant visualizations are presented simultaneously with narration) or the spatial contiguity effect (better information transfer when descriptions are placed near corresponding parts of graphics) [139]. Humans are also strongly predisposed to look for and interpret symbols and our use of sophisticated symbolic representation goes back to prehistory [144–147]. In fact, people are so strongly predisposed toward searching for symbols that we will often see symbols where there are none [148]. This predisposition can be used to compress large amounts of information into symbol sets which can be decoded rapidly by trained users in order to direct their attention or help generate situational awareness [107,139].

**Structure.** As described elsewhere, providing pattern and structure to content reduces cognitive load and improves the use of working memory, and the strategic composition and arrangement of content can allow even traditionally very dry or technical information, such as project documentation, to tell a story [32,77,149]. Further, when these patterns of content structure are in common use by users, they allow for deeper compression over time—memory studies on chess players and research on artificial intelligence has indicated that this

pattern-based inference may actually be synonymous with what we know as expertise [150–152].

**Collation and Curation.** As volume and structural complexity of information increases, the need for collation and curation (or filtering) of information becomes increasingly necessary. Collations do not have to be simple lists of content and curations do not necessarily correspond to interactive search and retrieval. Rather, collations can be treated as part of a more abstract process of intermediation—where curation and collation can result in their own information products, such as ensembles and clusters, or new reports which take what might otherwise be an unintelligible list of disconnected content and create narratives and counter-narratives which are easier to parse [48,153].

**Interaction.** When visualization, structure, collation, or curation cannot be applied without sacrificing necessary details or nuance, information systems can make use of interactivity. Interactive elements might include real time user-driven rearrangement of view, restructuring based on focus or purpose, or linking and relationship views, all of which can allow users to make use of visualization, structure, collation, and curation more flexible or convenient across many more dimensions than they could otherwise [154,155].

## Enabling Case Management and Providing Prescriptive Information

Case management is a key feature of many knowledge management, decision support, security management, intelligence management, relationship management, and, of course, case management systems. In medicine and human services, the care and services provided to vulnerable people are managed as to increase efficiency and reduce the likelihood of information and opportunities slipping through the cracks, warning signs going unnoticed, and basic procedures, or prescribed process, not being followed due to factors such as large caseloads or interorganizational information sharing [156,157]. These principles are arguably the same across the many disparate areas that require case management, such as security and law enforcement [158,159], counter-terrorism [5], customer service and outreach [160,161], law [162], and intelligence [48]. The typical case management system user could be described as either an individual whose job is to develop a plausible story using available information and requests for information (e.g., “Who is the most likely suspect given the information available?”, “Which precedents can we use to

structure a legal defense?”), or an individual whose job is to rapidly manage context shifts, develop or understand a story in order to fulfill their role, and figure out what to do next in some larger process while guided by prescriptive information (e.g., “Should this customer be given a refund?”, “What should I be asking this suspect given what other officers have already discovered?”).

As the structural complexity and volume of information increases and more parties become involved in the management of a particular “case”, the potential for error also increases. Basic procedures or prescribed tasks may go unfollowed, very obvious or critical information may be uncommunicated, unused, or lost, and further, the conversion of available information into a coherent narrative can be impossible [5]. For example, the failure to apprehend the serial killer Paul Bernardo was blamed on the lack of case management systems to help investigators collaboratively develop narrative [159]. Post-mortems on the investigation indicated that the organizations involved had the necessary information, but simply failed to connect that information in a coherent way fast enough [159]. Also alarming was the arrest of Brandon Mayfield, a lawyer from Oregon, on suspicion of his involvement in the 2004 Madrid bombing. His fingerprints were matched in an international, automated information fusion system, but the facts that he had never before traveled to Madrid, that he was arrested in Oregon and not Spain, and that the fingerprint system required additional checks after a match all failed to become immediately relevant to investigators during the multi-organization collaboration [5]. In yet another chilling case, a man mistaken for another individual with an outstanding warrant was arrested, placed in a mental hospital, and forced to take psychiatric drugs—“the more [the man] vocalized his innocence” by asserting he was not who they thought he was, “the more he was declared delusional and psychotic by [the hospital’s] staff and doctors and heavily medicated” [163]. After nearly 3 years, a hospital psychiatrist decided to consider the possibility something had gone wrong and was able to confirm the mistaken identity with “a few Google searches and phone calls” [163]. This case is of particular interest because of how easily this might have been avoided had proper case management procedures and tools been available or used. A simple comparison of photographs, fingerprints, arrest records, and the story they told would have made his release obvious at any stage—as it was publicly available knowledge that the individual he was mistaken for was already incarcerated in Alaska at the time of his arrest [163]. Such cases may seem extreme; however, as data-driven policing and legal sentencing become more common, situations of mistaken identity and inappropriate communication of narrative confidence have the potential to influence the lives of many.

Narrative approaches have been recommended in the past to remedy the types of problems described above, such as the use of timelines and storyboards and the fitting of information to narrative structure and pattern to make information more parsable by, easily communicated to, and easily extracted from teams [137,164].

Narrative structure has also been recommended for use in problems of task-transfer, project documentation, and rapid onboarding, in which knowledge and case management systems are often implemented [32,77,165]. Case management in task-transfer contexts is especially important to consider in high reliability activities, such as in passing on all necessary information to understand what is happening and why in command and control [165] and mental health care settings [157,164].

## Synthesizing Intelligence

Across all of the disciplines mentioned and the systems they intend to design and implement, there is, by merit of their interests in the various features noted above, an accompanying interest in using those features to collect, process, analyze, and synthesize information in order to create new information products. While this process may be best formalized by intelligence production [153], the myriad data and information fusion methodologies for taking raw data and other information and synthesizing them into viable intelligence could be considered a member of this category as well. Intelligence has been argued elsewhere, extensively, to be a primarily narrative process in which quantitative measures should play a moderating or bounding role, but not defining one [153,166]. Narrative and narrative-related sensemaking approaches have been recommended in the past on this basis in order to improve intelligence practice and systems [167,168].

## Concluding Comments

While these common categories of NIM features are often discussed in the literature within information-centered fields, there are likely other features of importance that are rarely given attention. This may be in part because these features exist in ad hoc solutions in the field (unknown unknowns), have yet to be generalized (known unknowns), or have been studied and generalized in some other field (unknown knowns).

## NIM in Various Domains

In the following sections we explore the past, present, and future of Narrative Information Management (NIM) in various domains. These sections were sampled based upon the experience of the co-authors, and by no means are exhaustive in terms of breadth (across disciplines) or depth (within a discipline). The sections serve to (1) raise awareness of the commonalities of some challenges faced by different fields, (2) explore both theoretical and practical insights about the implementation and design of NIM features, (3) provide opportunities to discover and generalize NIM features, and (4) begin the process of working towards NIM as a unifying framework.



## Personal Finance

Narrative information management in finance can be divided into personal finance and institutional finance. Globally, affordances vary in both sectors. This overview will discuss the narrative pertaining largely to personal finance in the United States (although it may be applicable elsewhere). The individual financial narrative begins at birth. Even in the wealthiest countries in the world, there is a chasm that divides those who are able to consider what to do with their money and those who don't have ample funds to cover an emergency. The cost of poverty is very real, and can be compounded by various disparities (e.g., social, medical, educational, likelihood of experiencing trauma). It is important to recognize that attitudes and knowledge about money start to develop at a young age, vary across generations, and that intergenerational wealth has an impact on the personal finance narrative. Financial psychology is also shaped by genetic and biochemical factors, particularly the aspect pertaining to risk tolerance and power [169,170].

The variation in financial psychology makes it difficult to establish a single purpose that is achieved through processing relevant information. The standard K-12 curriculum does not include finance, therefore, the motivation to find meaningful financial information may come from life experiences, such as debt accumulation, or the desire to sequester financial resources. There is a limited time frame in which to accomplish any financial goal, leading to a temporal pressure. Investors must choose how to decide (what amounts, what investments to make), but also when to decide [171]. Furthermore, because financial resources are (relatively) finite, there is also competitive pressure. Common starting points for those who weren't exposed to extracurricular financial education in their early life include books by Suze Orman, Dave Ramsey, and Robert Kiyosaki. However, one substantial and important subject has been omitted from all of their books: detailed information about investing [172].

Investments maximize the accumulation of financial resources over time. However, searching for the right investments can lead to a deluge of information. This makes financial literacy difficult to achieve for the everyday investor. In fact, due to the increasing complexity of the economy, even experts struggle with defining financial literacy [173]. People with excess capital primarily invest in traditional investments (stocks and bonds). Some investors also include nontraditional investments such as art, real estate, foreign currencies, and cryptocurrencies or non-fungible tokens, among others. A mix of investments is frequently chosen based on investor demographics, including age, gender, portfolio value, interests, style of portfolio management, and risk tolerance [174]. Furthermore, real estate and stocks have intra-asset investment scales ranging from macro to micro. For real estate, macro scales include Real Estate Investment Trusts (REITs) and online lending platforms, whereas micro scales include rental properties and house flipping. For the stock



market, macro-level investing is done in index funds or exchange traded funds (ETFs), and micro-level investing can be individual stock purchases, financial derivatives, or partial stock shares. Informational burdens can prevent individuals from making wise investment decisions, hence the relevance of NIM for understanding real-world behavior.

Many investors choose to outsource their financial decisions to a credible third party. Outsourcing can be done through financial advisors or even using online robo-investing platforms. Cognitive offloading through a third party reduces the decision space from high dimensionality (such as which house or which stock to buy, when do I buy it, etc.) to low dimensionality, consisting of, perhaps, only choosing a financial advisor or platform and the amount of money to invest. Moreover, crowdsourcing the reviews of financial advisors and investing platforms relieves the cognitive burden of even these basic investment choices. There is a great degree of trust that comes into play when putting money away, which has resulted in professional certifications and related duties (e.g., fiduciary duty) that reduce the cognitive burden on the investor and consequences for certified financial fiduciaries who don't act in their clients' best interests [175]. Regardless, choosing an accredited third party can be much simpler than trying to search through the glut of information that is available about investing, much of which is promoted by those with a vested interest. Furthermore, investment prices are swayed by the weekly economic statistics as well as other news pertinent to individual stocks, and it can be difficult for individuals to track this information as they navigate their own investment path. Individuals who take this route will confront many of the challenges from a NIM perspective, such as information overload, incorrect or misleading information, and the need for effective action-oriented sensemaking (buying and selling) amidst uncertainty. However, for those who decide to take investing into their own hands, informative resources are available.

Resources for investors are available on even the most basic investing mobile platforms. Platforms such as Robinhood include the price of the stocks over the last five years, stick charts, market capitalization, earnings per share, price/earnings (P/E) ratio and dividend yield. Higher level data is available on free platforms that retail investors can use, such as Thinkorswim, which contains more than 400,000 economic indicators as well as sentiment analysis tools that can be used to evaluate stocks [176]. Critical information that has the potential to give users an edge in investing is concentrated in the Bloomberg Terminal, which costs around \$2,500/month for access [177]. This is what quantitative analysts use in professional trading. If you want to evaluate a particular company's stock, the terminal has all of the financial statements, a compilation of analyst research on the company, and a network of their biggest suppliers and customers that can be pinned to a world map, among many other features. Perhaps the most important feature of the Bloomberg Terminal is access to the Enterprise IB chat. This feature facilitates communication

among brokers and portfolio managers, and is where many off-exchange trades happen. Off-exchange trades can be for over-the-counter (OTC) securities, which are unlisted stocks, or for publicly traded stocks. Publicly traded stocks that are sold off of the exchange are referred to as dark pools. These trades are usually for a large amount of stock, at a price that isn't always the listing price of the stock. Both OTC and dark pool trades are prevalent in the cryptocurrency market as well, as cryptocurrency is starting to resemble more traditional asset classes [178]. Moreover, while cryptocurrency is not currently regulated by the SEC, top federal officials have called for guidelines on cryptocurrency governance due to the potential risk for investors [179].

The intersection of personal and institutional financial narratives is a tightrope walk, largely because it is illegal to leverage critical, uniquely held information about stocks for financial gain (a practice known as insider trading). Regulation Fair Disclosure was enacted in 2000 to limit the practice of selective disclosure, where companies provide material information to analysts and institutional investors in advance of public disclosure [180]. Essentially this regulation ensures that the institutional financial narrative is consistent. In 2013, the Securities and Exchange Commission (SEC) verified that social media was an appropriate non-exclusionary channel by which material information could be disclosed [180]. The SEC is charged with regulating instances of market manipulation, which is the intentional manipulation of security prices. Individuals working in business-financial news, technology news, and media news have restrictions on owning securities that extend to their family members [181]. This prevents overt manipulation of security prices by news outlets. However, social media provides potential rallying points for individuals to potentially participate in pump-and-dump schemes or other nefarious market-related actions.

Situational awareness is frequently co-constructed in emergent online investing communities. The diversity of user opinions in these spaces usually prohibits the development of a team consensus; however, there are some strong opinions that are widely held by the majority of users. For example, in the Reddit platform r/wallstreetbets, the consensus narrative asserts that you should never bet against Tesla (TSLA). Many users have, and continue to do so, and when they have lost lots of money, they will publicly seek absolution from "Papa Elon," referring to the iconic Tesla CEO, Elon Musk. The price history of Tesla stock has been drastically divergent from their actual earnings. Reasons for this discrepancy could include the cult of personality that has developed around Elon Musk, or the herd mentality of investing communities [182]. The influence of the Tesla CEO is so profound that the SEC has mandated that Tesla pre-screens all of his tweets to prevent manipulation of the stock price [183]. He has also been accused of manipulating the cryptocurrency market [184]. Seeking explanations for the influence of Elon Musk points to the

mechanisms people use to model and monitor the financial markets, such as the subreddit of r/wallstreetbets and FinTwit (Financial Twitter).

Investors turn to online financial communities on Reddit and Discord, or follow influential investors on YouTube, Twitch, or Twitter for many reasons. They could be seeking to confirm their own biases regarding the fitness of their portfolio, or trying to select their next investment. Online communities also serve as a way of monitoring information. A tweet from Elon Musk could serve as a buy or sell signal for cryptocurrency or TSLA (or even ETSY), because historically the prices can skyrocket or plummet depending on what he says. Investing communities also serve as a way to analyze sentiment about the current market and the herd mentality. These communities have largely superseded mass media news outlets for younger investors. However, the price of stocks will still increase when financial news personalities, such as Jim Cramer, plug stocks on their prime-time shows.

The management of narrative information related to financial decision-making amidst uncertainty plays out continually – every time an investing firm makes a trade, or a retail investor interacts with modern financial affordances. Amidst the barrage of technical information (e.g., charts, data, disclosures) and ongoing context (e.g., online chatter, memes, intuition about sector), investors seek to make wise decisions about which actions to take. As the discussion above reflects, there is significant fragmentation of platforms, markets, and perspectives related to finance, with the implication that there are inadequate frameworks for narrative sensemaking, especially for retail investors. This gap in sensemaking capacity can result in decisions that are sub-optimal in terms of value, risk, or cognitive burden. Further research into financial sensemaking specifically, and the role of narrative in decision-making more broadly, might find interesting applications and implications in the financial systems of the future.

## **Ancestry Research**

Amateur ancestry and genealogical research have been steadily growing in popularity over the last decade and this growth has been accompanied by the development of a wide variety of tools to facilitate the process [185–187]. The COVID-19 lockdowns starting early 2020 greatly increased this growth, drawing millions of more people to engage in and contribute to private and collaborative research activity in the interest of understanding who they are in the context of their family, national and cultural heritage, and their genetics [188]. These individuals are not simply searching for existing information, but actively performing research guided by investigatory processes and questions. The motivations and methodology of amateur ancestry researchers are often identical to those of academic historians, and amateurs grapple with similar information load as professionals, even if they do

so to inform the development of a personal and familial narrative rather than to contribute to a historical commons [189]. Further, there is often a dialectic and informal collaboration between academic historians and amateurs, as amateurs have different “rules of engagement” with sources, can take larger risks, and can forage for information “in fields where historians have seldom toiled” [190]. In this section we explore some of the past, present, and upcoming challenges of the field of ancestry research, with a focus on how Narrative Information Management (NIM) concepts are woven into the process.

There are tens of billions of digitized historical artifacts available for use to these researchers through available tools such as those offered by Ancestry or MyHeritage [191,192]. While only a minute fraction of these documents and images may be of use to any particular researcher within the scope of their family tree, this small fraction may amount to tens of thousands of documents, causing users to encounter information overload [193]. Among these documents are newspapers, letters, census records, church records, financial documents, wills, and many other formalized and non-formalized documents; some are in different languages, and some are written using shorthand, long forgotten slang, and other forms and styles of writing which are no longer common in modern times [194–196]. The collection and processing of these documents is done by a mix of professionals and users. The growing market for genealogy products has meant that companies are incentivized to broker access to document repositories and to hire experts to provide and curate archival materials and suites of frontend and backend tools to analyze them [185,189,191,194,196–200]. While the bulk of the archival material is supplied by these experts and document repositories, users also continue the development of annotations on available documents and forage for resources to add to collections to support their research, filling in the gaps within professionally developed archives [189,195,201]. The combination of professional and user-sourced objects and metadata means that there is an unfathomable amount of potentially relevant material for any individual researcher to engage with [193].

The development of resources to assist with research methodology and tradecraft has always been ubiquitous with the amateur genealogy community [194], but with the introduction of these large digital repositories, knowledge management, case management, and information fusion systems have become necessary in order to keep up with the information flow and avoid redundancy even in basic research activity [193,195]. Members of the online amateur genealogy community have taken to suggesting young or novice family history researchers to avoid structured research activity at first, instead recommending that they engage initially in “unstructured, exploratory activity” on these systems to familiarize themselves with the information environment before fully committing to semi-formalized work-flows [195]. These kinds of recommendations are not unfounded as introduction to the tool suites, dashboards, and document repositories is daunting enough that most

new researchers fall into a common pattern during onboarding which focuses on off-platform collection (e.g., physical photo-albums and documents physically accessible to the user, taking physical notes before uploading to the platform) [195]. These kinds of common patterns within this community have been modeled as a series of stages with separations of concern, scope, and expectations which are similar to other sensemaking frameworks, such as the intelligence production cycle [195,202–204]. Unlike other sensemaking frameworks [109,202,203,205–208], these stages are generally represented as a linear process with key transition points being marked not just by progress in the research but in the capability and skill of the researcher, with the earliest stage representing the aforementioned pattern of onboarding [195,204].

This onboarding pattern typically begins with gathering information from within the family, off of the platform. This activity consists of collecting and uploading anecdotes, documents, physical artifacts, and photographs [195,204]. Following this, in a phase denoted “learn the process”, researchers begin collecting itinerary-driven resources on how to handle information gathering, attending events, connecting with the staff of organizations who can answer questions or help them retrieve documents, and engaging in a trial-and-error approach of learning by doing [195]. The next phase is considered a key inflection point, referred to as “breaking in”, at which point researchers finally become comfortable enough to begin searching census data [195]. Given that census collections do not contain “browsing” materials—use of census data indicates a transition in terms of comfortability with the tools as well as a transition from exploration to exploitation as users begin to use data collections to fill gaps in developing historical narratives rather than simply exploring other narrative material, such as old newspaper articles or family photographs [195].

Once users have begun the process of making use of external document repositories, tool-suites, such as those found within the ancestry.com or myheritage.com genealogy platforms, assist them with exploring and exploiting relevant materials [195]. These tools and the community education resources on their use are necessary for success given that some of the services available to amateur genealogists are adding millions of new documents per day [209]. In the case of ancestry.com, visual hints will be placed on relatives in the user’s family tree which have information that is similar to objects in one of 32,000 external databases, such as dates of birth or mentions of surname—these hints allow the user to access links and context about these objects and are sorted based on likelihood of relevance [209]. If a user reviews an object via a hint and marks it as related to the relative the original hint was attached to, this will create an ensemble of “secondary hints”, which are other objects which may now be considered potentially related (e.g. an individual is noted in one document with an administrative identification number, the individual in that document is accepted to be the same as the one in the user’s

family tree, so all documents which are associated with that administrative identification number now become secondary hints for the user to review) [209]. Hints are accessible in a variety of ways based on workflow and objectives, for example, a user can review all hints, to see if there are recent relevant updates to review in aggregate, or see hints related to particular individuals based on a number of filters if they're in the process of a scoped investigation [209].

For researchers in this space, it's not enough to simply associate a resource with an individual. The goal for many of these users isn't to simply trace a family line but to construct narratives which provide context both for their ancestors' experiences and their own place in history [188,189,194]. Much like academic historians, the narratives have to be constructed of ensembles of facts sourced from various historical documents and accounts—however, unlike academic historians, amateur genealogists have specialized tools that facilitate the rapid and collaborative construction of these narratives. Where academic historians are left with tool recommendation lists which are often either barren or limited to simple citation managers, collection and archive search managers, and ad hoc tools designed for other fields [210–214], tools available to amateur genealogists allow for case management workflows rarely found outside of legal case management tools, which are intended to construct well-cited narratives built to stand up against scrutiny [215–217].

The use of “narrative scenarios” for describing typical research itineraries as a basis for the design of adaptive, personalized, task-focused access to multimedia, multilingual cultural heritage knowledge bases has transitioned from theory to practical, accessible tool-sets to assist in case management [198,200]. For example, when amateur genealogists attempt to research ancestors who took part in migrations, the accompanying name changes, lost records, sudden transitions, and separation from loved ones means that their more common research methods are no longer adequate [218]. While many services use an entity-focused approach, allowing for many names (or referents) for any given object, increasing the likelihood of finding an opportunity to merge common ancestors found by distant relatives that may have found those ancestors via other paths, it may require a great deal of luck to make these connections [218,219]. To continue, researchers would traditionally have to either rely on this luck or shift from the use of document archives and qualitative analysis to the use of bioinformatics and statistical analysis [220,221].

Amateur genealogy software providers have now integrated new tool sets, built on genomic “identity-by-descent” mapping methodology, which place users themselves in multiple ensembles, called “communities” [200,220,221]. These ensembles are constructed of members which share ancestors which likely hailed from common populations, groups which either “traveled to the same place around the same time or from the same place around the same time” [200], helping users rapidly develop



narrative about their ancestors which informs where to look for more information and, more importantly, who to collaborate with in order to fill knowledge gaps [200,221]. This formalization of a “narrative history” through the use of such tools has been argued to “allow for a group of individuals to be conceived as if they were united... for past and present individuals to be conceived of as one united group embarking on the same quest” [221–223]. Tool suites such as these help a community of practice that may not have had the benefit of STEM education connect with and make use of knowledge from communities of practice that use advanced tooling that would otherwise be inaccessible [221]. Further, this kind of connection creates incentives for the use and development of semiotic, visualization, and rhetorical techniques to construct micro-narratives that make the work of specialized communities accessible “without requiring command of an exclusive body of knowledge” [75,221].

NIM tool development in the amateur genealogy domain could benefit from incorporating design principles from other spaces with similar tooling requirements. For example, in terms of interoperability and information exchange between entities, which is often discussed in relation to geospatial intelligence, open-source intelligence, and the crowdsourcing of research and situational awareness resources [17,51], the amateur genealogy community currently has a one-way relationship with the expert communities that manage document repositories and provide them with tools—missing an opportunity to harness this massive collective effort of millions of hours a year in the research, linking, and annotation of historical documents [189]. Between competition over attribution [201,224,225], perverse incentives and social pressure associated with finding direct relations to famous or historically significant figures [189], limited consequences for incorporating poorly sourced facts or creating logical inconsistency [226], and the potential for errors resulting from these factors to propagate through the system, these user-managed knowledge bases are likely a negative resource for actual historians as aggregation would be too risky [189,199]. If user-generated knowledge bases were structured correctly with consideration for governance and trust signaling, taking account of the incentives generated by the desire to develop and present aesthetic and pleasing personal and familial narratives, then the data could be of more use not only to historical analysis and aggregation—but also for other purposes [51]. For example, data from AncestryDNA customers was filtered and cleaned for use in COVID-19 research but could have had much more impact had the system been built with protocols for information exchanges [227]. Further, the exchange of information between these communities could provide valuable feedback from more technically advanced, as the tooling they provide to the amateur genealogy community comes with great risk of being misused and misinterpreted [199,221].

Domains with similar tooling requirements could also benefit from considering NIM design impacts in the amateur genealogy space. For example, regular exploration of

a knowledge base is essential to its maintenance [48], and there appears to be a tendency in general toward exploratory browsing over searching in general throughout most of the amateur genealogy research process, which may be linked to the focus on intrinsic incentives for activity [195,228]. The intrinsic incentives associated with outcomes is associated with increased technological adoption among demographics traditionally left behind as well as patterns of behavior which lead to advanced learning, information use, and information foraging [229]. The value of this exploration is amplified by the fact that the popular tool-suites help users identify where others are missing information they might have, and vice versa, through linking and hints [195].

In terms of research facilitation and production, the ability to programmatically generate scoped and formatted research reports, charts and graphics, and even whole books prevents researchers from feeling punished for intentionally or unintentionally maintaining a separation of concern between the research itself and the presentation and dissemination (the development of research “products”) [230,231]. This conceptual separation of concern between analysis and dissemination is considered essential in high-reliability research and analysis communities and features which enable it would be beneficial to any domain concerned with or requiring NIM tooling [232–236]. Finally, enabling these research facilitation and production features are user experience (UX) design features that allow for the scoping of the user’s information environment based on relevance, relationships, and degrees of separation between the object in primary focus or center of gravity for attention (e.g., a relative in focus) and other objects with which that object has a relationship which prevents information overload [193,230]. The underlying, universal entity identifiers that allow for these features also allow users to rapidly develop surfaces of agreement even where they do not agree on all facts or interpretations associated with content (e.g., we can agree that this is a photo, that this is a photo of this person, and that it was added by this user, but do not agree it was taken at this location) [191,218]. Similar to many other areas of ancestry research and amateur genealogy relevant to NIM, there is an apparent need to consider the incentives of the user and the potential damage that those incentives may bring to the knowledge base. If there was one insight to draw from this area, it would be that the failure to consider consensus, governance, and trust mechanisms in contributions will inevitably lead to a tragedy of the commons—in the case of ancestry research, this tragedy is expressed in the unusability of what could otherwise be a mountain of valuable historical data, robbing millions of their opportunity to contribute meaningfully to the corpus of historical knowledge.



## Hybrid Cloud Infrastructure Security

The modern economy is supported by a vast array of layered and interconnected information systems, which enable the internet and various intranets, and generate dozens of zetabytes of novel data per year [35,36]. At all layers, from users accessing social media platforms to data centers processing underlying workloads, there is a persistent, complicated, and complex set of challenges associated with hosting servers that resolve website traffic and provide secure access to data. These challenges are generally associated with resolving who and what should be able to access particular digital resources and under what conditions identities should be allowed to interact by reading, writing, deleting, changing permissions, or other actions on said resources. Users, administrators, and machines engage in facilitated interaction with cloud infrastructure through credential, entitlement, password, and permission management systems, each of which are types of trust management systems designed to handle the aforementioned challenges behind the scenes and strike a balance between fundamental tradeoffs, such as the tension between security and convenience [237]. For example, password and permission management systems facilitate the management and safekeeping of a burgeoning list of access credentials and permissions for users of information systems and online platforms [238,239]. Trust management is becoming increasingly difficult—especially with the introduction of hybrid cloud computing. We will explore the current state and future possibilities of narrative information management approaches as they relate to the world of security for hybrid cloud infrastructure.

First, a primer on definitions is necessary for this discussion. A data center is an interacting network of computers across one or more physical locations, which handle computational or information processing workloads [240,241]. These workloads might be maintaining and developing web services, executing large-scale data management [240], offering compute power for research and data analysis tasks [242], managing data access, or enabling business continuity through disasters or cyber attacks [240,243]. Data centers can be on-site or externally-located, and they can be either owned or rented [244]. There are three terms commonly used to describe the nature of an organization's cloud infrastructure choices: private cloud, public cloud, and hybrid cloud (see Table 1).

---

**Private Cloud.** The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

**Public Cloud.** The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

**\* Hybrid Cloud.** The cloud infrastructure is a composition of two or more distinct cloud infrastructures that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

*\* Hybrid cloud can be seen as an overarching trend in industrial computing toward mixing and matching different private and public cloud options when deciding the infrastructure composition for a given organization.*

---

*Table 1. Types of Cloud Infrastructure.*

In all types of cloud infrastructure, computational resources and user privileges must be balanced and managed to keep development projects running efficiently, while also detecting and remediating technical and security issues in real time under pressure [245]. The number of issues that may arise is difficult to comprehend. Some estimates have suggested that, just in terms of security events, “analysts [can] be expected to handle only about 0.00001% of overall event volume”. One analysis of a mid-sized enterprise platform revealed that, based on an average of 40 million log entries per day, 40,000 analysts would be needed to address all security events without triage [245,246]. Among these types of cloud infrastructure, hybrid cloud may contend with the most complicated and complex set of challenges, due to the scale and dynamic nature of the access required by various types of users and systems [247–249]. Hybrid cloud solutions are utilized despite all of these challenges because of the numerous advantages they provide, particularly in terms of flexibility and antifragility. For example, hybrid cloud infrastructure provides a customizability and specialization that permits a better fit between workload, platform, and users—allowing teams to choose the platforms and authorization systems best suited for their particular workloads and team dynamics initially and over time. Further, hybrid cloud solutions enable grouping by type of workload, thereby improving efficiency and the ability to maintain function under increased or fluctuating demand. Given these advantages, and the number of organizations now offering services in this domain, hybrid cloud infrastructure may become dominant.

The influence of trust management systems in modern cloud infrastructure is pervasive. As the modern world moves toward a reliance on hybrid cloud infrastructure, the control, ownership, brokerage, and regulation of information, information privileges, and the information infrastructure itself is becoming a very high leverage point—financially, geopolitically, and ethically [250–254]. On the horizon, citizenship, voting, and other core rights may be facilitated digitally. In fact, the digital facilitation of banking, taxation, access to electricity, and other core functions is already becoming commonplace. Therefore, effective management of credentials, permissions, entitlements, and trust may become one of the most important problem spaces of our time. The fundamental aspects of life in modern democracies that are currently being managed and manipulated digitally beg the

question: what happens when adversaries successfully disrupt or compromise these systems? How do user-specific narratives of personal experience and action feedback into the computationally-aided design of trust management systems? How do these massive systems remain resilient when feedback loops and low-reliability nodes might interact to form complex threat surfaces [255], resulting in endogenous failure modes? Such targeted interventions and intrinsic failure modes in these complex cyber-physical systems might be subtle or unnoticed initially, with devastating repercussions.

Novel types of hybrid cloud infrastructure and trust management systems are now being explored in various areas, such as the digitization of Department of Defense and civilian supply chains [256], intellectual property [257], 3D manufacturing [258,259], and bioinformatics [33]. These explorations in disparate areas bring new interconnected risks, and raise questions of how different types of organizations should respond to threats and anomalies, both alone and in concert [260]. Compromised hybrid cloud infrastructure results in security events of varying type and severity. While some security events can be limited in scope, other events can prove costly, and even fatal, to individuals, governments, and businesses in terms of loss or discovery of identity, irreversible loss or inappropriate access of data, or denial of service at critical moments (such as voting intervals for a government, holiday shopping period for an online store, or loss of trust due to exposure of personal data). Additionally, unauthorized access can have network effects leading to further inter-organizational risks and threat surfaces, and are happening more frequently to both small and large operations alike [261]. Wargames and red-team events are currently used to help security professionals and stakeholders better understand and classify external threat actors and types of target organizations. This understanding can be compressed into categories for simple communication, helping to teach security professionals and students about common patterns and risks [262–264], sharpen team capabilities and resilience [265], and develop scenarios for emergent or unexpected events. While there is often an emphasis on threat actors, security threats can also be caused by misconfigured bots and human error, in isolation or in interaction.

The complex dynamics of human-machine interfaces (the basis by which human organizations interface with hybrid cloud infrastructure) results in another fundamental challenge in cloud computing security. As mentioned previously, analysts, developers, administrators, and users are all under time pressure to perform their duties using hybrid cloud infrastructure, engaging in a fundamental tradeoff between security and efficiency, sometimes resulting in the provision of permissions beyond what was needed. When admins fail to account for these overprivileging events and fail to take actions to minimize ongoing risk, these errors accumulate, leading to a phenomenon referred to as “privilege creep” [266]. Hybrid cloud administrators are thus tasked not only with identifying individual errors at a

moment in time, but also with identifying cases of missed or unhandled error accumulation over time. They must then remove unnecessary privileges in so-called “remediation events.” Unfortunately, these realistic and fundamental challenges run the risk of being ignored or underestimated in the academic literature, due to disconnects between theory and practice, the speed at which new security threats emerge, and the assumption that negative externalities borne of human-machine interface dynamics are linear and might simply be engineered away [267].

Hybrid cloud admins are usually assisted in the identification and tracking of privilege creep in their data centers and practically minimize it over time by using a framework called the “Principle of Least Privilege” (POLP). Examples of successful applications of POLP include issuing temporary access tokens for identities in a data center, right-sizing roles for particular categories of hybrid cloud workers, and limiting access to high risk resources or actions that aren’t often used by that identity. Generally, POLP can help reduce the informational complexity of the narratives used by hybrid cloud admins when planning beneficial actions to lower risk over time. Similar to POLP, the Confidentiality-Integrity-Accessibility (CIA) triad is commonly used to simplify the assessment of threats to data center resources, where risk is examined in terms of potential for the theft or exposure of sensitive information (confidentiality), the corruption or malicious altering of information (integrity), or the removal of access to critical resources at a critical time (accessibility) [268]. In cloud settings, actors don’t need to be intentionally-malicious to represent a threat; they may instead represent misconfigured automated users or service accounts (bots), or simply human users making mistakes, cutting corners to save time, or acting in destructive interference with others unknowingly [17].

In the face of such fundamental uncertainty, hybrid cloud managers adopt frameworks like POLP and the CIA triad as a practical means of rapidly developing a narrative from which to derive prescriptive information and explore risk minimization in data center operations. While these mental models are imperfect, they do offer a dimensionality reduction in information- and relationship-rich environments such as hybrid cloud infrastructure. This use of narrative to provide situational awareness makes it easier to form and communicate with stakeholders, avoid analysis paralysis, and take beneficial action. With this approach, effective hybrid cloud management occurs over time, with small actions of limited scope that make the environment iteratively more manageable and secure with each admin engagement [1]. Software that provides auditing and case management, streaming anomaly detection, as well as visualization of current state and projection of future state, enable both batch and streaming remediation as evidence of unusual and risky behavior accumulates past a certain threshold. In addition, information fusion methodology (e.g., automatic collation of data from multiple systems) is sometimes applied to weave non-privilege related events into a story of potential risks, such as equipment reported as lost or the misuse of software or hardware [269], thereby

facilitating NIM in hybrid cloud systems. The value of information fusion systems increases as interorganizational credential management adds new layers of complexity. For example, the need for multiple organizations to share in governance and management of trust in providing access to common information and resources (e.g., computing power for biomedical image processing [270–272]), roles, tasks, and job assignments). Indeed, the operations of cloud computing infrastructure present a dizzying and evolving complex threat surface [32].

The field of hybrid cloud infrastructure security is still in its infancy, and it is unclear which technical solutions will remain stable given the presence of the fundamental, adversarial, co-evolutionary relationship between potential threat-actors and security professionals. Compounding the challenge of problem definition and solution development in the field of trust management, the number of relevant threat surfaces is increasing rapidly. As field devices (e.g., remote sensors, tablet devices in industry) are increasingly placed into use, exposing critical information systems to new complex threat surfaces, such as those created by requirements for use under sporadic connectivity, leave these systems more porous than ever before [273]. Further, credentials aren't just for people using technology, but also for autonomous objects such as IoT (internet of things) devices—as of 2010, it was estimated that there were already twice as many IoT devices than there were human beings [274], each of which represents a threat surface and new degrees of agency which may require new technical solutions. However, it appears that the approaches and frameworks noted here that are relevant to the management of narrative information, such as POLP, CIA triad, and information fusion are relatively immutable in the face of technical changes in the space. In other words, while the hardware, datasets, and software pipelines that compose data center and related trust management systems might be undergoing constant evolution over time, the centrality of narrative-based heuristics for actionable risk remediation frameworks may remain fundamental.

Due to the instabilities inherent in these early stages of trust management system development in hybrid cloud infrastructure, there is ample opportunity for the field of hybrid cloud trust management to both benefit from and contribute to narrative approaches and frameworks. With the right levels of generalization, transfer of models and tool suites between domains could be expedited. For example, the narrative models and tool suites which help inform scientists about the state of immune systems, homeostasis, and other elements of biological health could be converted to inform administrators about analogous features within hybrid cloud infrastructure, thereby helping to communicate and calculate risk more effectively [275]. Further, the use of models transferred from other fields may come with the benefit of established and tested collection and processing methodology in other fields such as crowd-sourcing and pattern analysis. A deeper dive into the specific types of narrative information (e.g., prescriptive, predictive) used in hybrid cloud

management systems is recommended, and it should be noted that Trust Management exists well beyond hybrid cloud infrastructure. Many of the problems and solutions discussed here could generalize well beyond this domain.

## Translational Neuroscience

Neuroscience is the scientific study of the nervous system. It is a multidisciplinary field that combines approaches from genetics, molecular biology, physiology, psychology, medicine, and many more. Translational research is the realm that connects basic research (performed on isolated systems in the lab) with clinical research (including diagnostics, treatment, and management of human diseases). Translational neuroscience research benefits greatly from the use of mammalian animal models such as mice and non-human primates to mimic and treat disease states in experimental ways, before attempting human trials. As a paradigmatic case of the challenges inherent in applying basic neuroscientific research insights, and example of Narrative Information Management “in the wild”, we focus on the area of neurodegenerative brain disease. Treating brain disease has its own set of challenges—mainly that changes in human behavior and cognitive skills often don't have a clear connection to the pathophysiology or systems studied in the lab. In this section, we provide some perspective on Narrative Information Management in the field of Translational Neuroscience, using Alzheimer's Disease as a case study.

One of the first challenges of medicine and biomedical research is to describe the disease in the population and identify the cause. Patient case studies and postmortem tissue analysis provide the first glimpse at the connection between behavior and pathophysiology. Alzheimer's Disease (AD) is an irreversible and progressive brain disorder that affects 6.2 million people in the USA [276]. It is the most common form of dementia, presenting clinically with memory loss and cognitive decline. Only 5% of cases can be linked directly to genetic mutations, for all other cases (called sporadic AD), the main risk factor is age; AD incidence doubles every five years after 65 [277]. Neurochemically, AD is characterized by the presence of amyloid plaques, neurofibrillary tangles (NFTs) and loss of synapses in the brain [278]. AD pathology is complex—it may present with all or some of these pathologies: amyloid plaques, NFTs, inflammation, oxidative damage, iron deregulation, blood-brain barrier dysfunction, and alpha-synuclein toxicity [279]. The relationship between these pathologies remains unclear, as observational studies cannot differentiate between “cause, consequence, compensation or confound” [280]. Clinicians are limited in their diagnostics for patients, because many of these symptoms do not have biomarkers, and the diagnosis of AD can only be confirmed post-mortem. The NIM challenge for clinicians and scientists remains: what causes AD? What is the “story” that connects disparate empirical results across decades and domains? Is there a causal link between the common symptoms? For



now, the approach has been unidirectional in the sense that molecular changes are hypothesized to lead to changes in patient outcomes, and each of the molecular pathologies have been explored in relative isolation.

One shared process of NIM or sensemaking among scientists and clinicians is that experiments are designed to explore hypotheses. Following an established hypothesis, scientists design the experiments to support or reject. The design of the experiments depends on the perceived relevance of the proposed hypothesis and extent of support from funding agencies (e.g. the US National Institute of Health). To mimic AD neuropathology, scientists often make use of cell cultures and mouse models, where the neurotoxic proteins can be added externally in cultures or genetically encoded to accumulate in the brain of the mouse. Mice have a shorter lifespan, different brain structure, and different behaviors than humans; therefore, direct extrapolation from mouse studies to human biology is hardly straightforward. One caveat is that mice lack the core protein components involved in the plaques and NFTs, which are hallmarks of AD pathology. Mice can only develop these protein aggregates with human neurotoxic proteins [281]. Another critical interpretation issue is whether or not it is possible to measure small, slow changes in the cognitive performance of mice, as typically measured in humans. Animal studies commonly measure changes in spatial memory, but often ignore neuropsychiatric axes, like anxiety [282]. The question remains—how can we model this disease in a useful way that allows for mechanistic exploration of the pathology? Can we treat the behavioral symptoms of memory loss by removing the underlying pathology? In a genetic mouse model of AD, yes, but in patients—no. Alarmingly, the same drug that removed plaques and improved memory in mice actually led to cognitive decline in patients, which continued even after the trial [282]. Among the proposed solutions are biomedical efforts to create mouse models with multiple pathologies [283] and connect the symptoms mechanistically. Thankfully, these findings are published in peer reviewed journals and are accessible to the research community. In navigating the wealth of publications, scientists are often taught to consider each publication as a story, such that specific findings are easier to remember in the context of the whole story. Due to the daunting amount of published literature and plausible research avenues, scientists and funding agencies are faced with a narrative challenge: which studies should be funded, which hypotheses should be explored? Such questions are often pondered by individuals, agencies, labs, and researchers, but such efforts are rarely connected back to the broader literature on narrative sensemaking.

Beyond the direct reach of academics, NIM plays an important role in research, strategy, and decision-making in industrial and pharmaceutical sectors. The actions of these large entities bear strongly on clinicians, who eventually deploy the solutions/therapies that stem from neuroscientific research. Pharmaceutical companies access the public knowledge of animal and clinical studies, but also

create their own private research depots. As such, pharmaceutical companies navigate the complex processes of scientific development, FDA regulation, patenting, and marketing. Pharmaceutical companies work with clinicians and researchers to develop large scale clinical trials. Clinical trials require an interface between patients and the public. As of 2007, clinical trial data is compiled at the NIH clinical trial database, although timely reporting is not enforced (clinicaltrials.gov) [284]. In Phase III clinical trials, the drug is given to a patient for the first time and tested for efficacy. Therefore, designing these clinical trials is a multifaceted challenge, as researchers try to recruit the right number and type of patients, as well as determining the time of treatment and appropriate measures [284]. Collecting, storing, and analyzing such quantities for sensitive health information calls for NIM solutions. Recent advances for improving experimental design include Bayesian modeling for determining appropriate endpoints, classifying patients based on medical history, and novel detection of AD biomarkers [278].

The last mile for applied neuroscientific research is in the NIM of patients, especially in their interactions with clinicians. Patients and their families learn about potential treatments and manage disease in patients, based on information they integrate into personal narratives. All of this starts with access to medical care and proper diagnosis of health conditions. Outside of the doctor's office, patients receive a highly profitable stream of direct-to-consumer advertising (DTCA) from pharmaceutical companies, such that patients can learn about new drugs and request them from their doctor. A common side effect of DTCA is the increasing demand for new and costly treatments in lieu of existing low-cost options [285]. Another way that patients learn about therapies is through social media and scientific communication. Unfortunately, the headlines may give false hope, and animal research gets more media coverage if they don't include "mice" in the title [286]. The recent controversy around the FDA approval (and reversal) of the drug Aduhelm (which targets plaques) has done a lot to shift the narrative around accepted hypotheses for AD [287], which now include targeting NFTs, light/sound therapy and immune cell stimulation. Clinical trials on lifestyle changes such as exercise have shown that regular physical exercise prevents age-related brain atrophy and helps with neuropsychiatric symptoms of AD [288], however research on public health interventions can be misrepresented greatly [289].

Health NIM exists at multiple nested scales, and while AD is one such case, it's becoming clear that everyone is participating in the management of health narratives on some level. Researchers, clinicians and the public need new tools and training for making appropriate decisions about health policies. In the scope of treating human disease, translational research is positioned between basic and clinical research, and therefore experiences the burden of NIM challenges: managing information gaps, exploring the informational environment, and synthesizing diverse sets of information. Future studies in the NIM of health could examine how



public policy influences the narratives of individuals. Particularly for individuals dealing with long term health issues, NIM tools may help alleviate the mental, psychological, and logistical burden of decision making.

## Genomics

Genomics is an area of theory and application where biological datasets are analyzed to address a variety of questions related to human health, government policy, agriculture, industry-led research, environmental monitoring programs, and more. “Genetics” refers to the broader study of trait development and inheritance in biological systems, while “Genomics” usually refers to the modern (post-2000) high-throughput technologies used to measure biological molecules such as DNA, RNA, protein, and metabolites.

A failure of NIM for genomics at the institutional level could look like inadequate or grievous policy deployment, based upon improper assessment of biological information or risk (e.g., a false-positive or false-negative decision to institute a regional lockdown based upon the perceived risk of a virus identified only from genomic sequences). In contrast, for individuals a failure in genomic NIM could have life-altering consequences regarding the perceived meaning of genomic information obtained from medical experts or personalized genome sequencing services. Socially, when NIM in genomics falters, it can lead to increased distrust in the scientific endeavor and an acceleration of the epistemic crisis in the knowledge commons – especially as genomics technologies such as human personalized medicine and viral sequencing become prevalent. This may be linked to the nature of our own genomes, in that it is linked to our shared identities as well as personal uniqueness.

In this section we provide a few views on NIM in 2021 within the field of Genomics. This section is not a broad review of the wide topic of Narrative Genomics [290–292], rather it is a selection of enduring and recent features of genomics in the context of NIM and cognitive security. Genomics presents domain-specific and transdisciplinary teams with a set of constraints and opportunities, some of which are unique to genomics and other aspects are shared broadly across fields:

**Underlying system complexity.** Genomics data, while sometimes vast in terms of computational size [33], are only the tip of the iceberg in terms of the complexity of the actual biological system (e.g., the inner workings of cells and tissues). Even though genomic technologies provide high-resolution maps for humans to navigate biological systems from the cellular to the ecosystem scales, the underlying territory is

vastly more intricate and nuanced. Biological systems consist of many kinds of interacting molecular components (proteins, lipids, nucleotides, carbohydrates); the overwhelming majority of which are involved in numerous relationships and thus, may not have a clear function when considered in natural contexts. As higher levels of organization in biological systems (e.g., social) are in dynamic feedback with lower levels of organization (e.g., cellular), it can be unrealistic or impossible to disentangle the effects of interactions among layers [293,294].

**Sheer scale of data.** Biological datasets have exploded in size recently, as the costs of genomics experiments drop and their throughput increases. Since the 1980's, the total amount of genomic data has been increasing roughly exponentially [33,295,296]. This access to genomic data is providing new opportunities for genomics researchers, technology developers, and medical practitioners. However, for researchers looking to investigate these data sets, even with relatively straightforward questions, a new level of computational skill is required. Even best-in-class information, such as gene expression profiling at the single cell scale, are very partial representations of living systems, and require extensive computational analysis in order to derive insight.

**Social relevance and sensitivity.** Genomic data play significant roles in individual and collective narratives around various topics, including the legality of discrimination (as per The Genetic Information Nondiscrimination Act of 2008 [297,298]), the nature of ethnic and sexual identities [299–301], and broader discussions around the relationship between inheritance systems (genetic, epigenetic, and cultural) [302,303]. As genomic editing technologies like CRISPR/Cas9 become increasingly accessible to laboratories around the world, contemporary narratives around human genome modification are of historical importance [304]. Also of note here is the recent deployment of almost real-time genomics analysis in response and policy planning around the emergence and spread of the SARS-CoV-2 virus, responsible for the COVID-19 disease.

**Personal Identifiability.** The data generated by genomics experiments are essentially personal – they can be used to

identify relationships among living and dead people. Genomic information can be extremely informative or even conclusive regarding various questions related to forensics, law, heredity, and medical diagnoses. Biological and genomic data can be extremely sensitive in terms of personal privacy, to the point of being able to identify individuals who have not even submitted their own genomes for analysis (as in the recent case of the "Golden State Killer" who was triangulated using a combination of detective work and DNA evidence [305]). Dealing with large datasets of potentially-identifiable or health-related information, genomic or otherwise, comes with new challenges.

Genomics is a technical area that recently is experiencing wide public participation in the analysis and interpretation of data. This expansion of social accessibility in the genomics process can be attributed to multiple factors, including the increasing prevalence of direct-to-consumer genomics tests, and the growing role of genetic data in driving individual health decisions and public biosecurity policy. Those who work directly with genomics data might fall into a few categories, each with different pressures, incentives, affordances, and narrative contexts:

**Academic Researchers.** Academic researchers are more likely to be working on non-human data, more likely to be working on basic or theoretical questions, and may have knowledge of the field but remain unaware of state-of-the-art tools used by computer scientists for secure cloud computation at large scale. Academic researchers face the pressures of science as a career (e.g., pressure to publish and their working environment).

**Industry Researchers.** In industry and government, researchers face a different set of affordances and pressures than academic researchers. These researchers may variously be working on human, microbial, livestock, or agricultural genomics data, often with a more direct focus on applications. Applied genomics research in industry occurs under direct or indirect business pressures, as the results of the analysis are financialized in a way that is distinct from other research domains. Government researchers may use genomic data in a range of settings, of particular interest is the consideration of public health implications for viral variants. As the SARS-CoV-2 pandemic shows, genomics data support governmental decisions in real-time, meaning that increased emphasis is placed on reliable bioinformatic pipelines, clear visualization of

essential data features, and contextualization of genomic data so that it is informative for non-experts.

**Medical Analysts.** Medical analysts are more likely to be working with human (or veterinary) topics and data; thus, they are under pressures related to efficacy, timeliness, and data privacy. Medical decision-making occurs in the context of transdisciplinary teams, where genomic data plays an increasingly large role as the price of acquiring personalized genomic information drops. Genetic counselors, specifically, are the contact point between the technical details of genomic data and interpersonal communications with patients, most of whom are not familiar with the intricacies of genetic medicine [306–308].

**Non-institutional Researchers.** Individuals outside academia, industry, and medicine are also beginning to gain access to genomic data – for example through the use of personal genomics services, or public databases containing viral sequencing data. Developing communities that use genomic data and tools include citizen scientists, biohackers, and data-driven journalists. Many of the tools useful for genomics are open-source and utilize free public databases. However, non-institutional researchers may face computational constraints, gaps in their knowledge of genomics, or be unfamiliar with norms around communication of results. Not every citizen can be expected to have the knowledge required to perform bioinformatic analyses or write genomics papers – but when common topics of public discussion include nuanced and “science-informed” discussions, shared understandings are essential.

Genomics as a field stands at the intersection of biology, identity, data, and policy. Practitioners of genomics come from a wide range of backgrounds, and increasingly genomics data is playing a real-time role in decision-making. Some of these developments have been unfolding for decades, such as the continued trends of decreasing costs of sequencing and increasing capacity for genome editing. Other changes in the deployment of genomics have specifically arisen in response to the pandemic spread of the SARS-CoV-2 virus and subsequent global response. It is imperative that analysis and communication of technical findings be made rigorous and accessible, especially where genomics is playing a directly narrative role in the public eye, for example related to viral variants, genetically-modified agriculture, and disease-associated human alleles. Further research and collaboration can seek

to understand the interface between the ever-expanding frontier of genomic technologies, and one of the essential features of human cognition: effective narrative sensemaking amidst uncertainty.

## Discussion

Our initial search for commonality within information-centered fields, such as knowledge management, yielded a broad set of useful features common to Narrative Information Management (NIM) systems. In the interest of discovering other NIM-related features, which are perhaps understudied or obscure, we explored an eclectic selection of fields sampled from the experience of the coauthors.

Here we review our initial insights about common NIM features and introduce 4 additional features of NIM that were revealed upon deep, field-specific consideration: (1) facilitating communication, (2) handling of errors and inconsistency, (3) management of trust signals, and (4) social systems engineering and education. These features were illuminated while contemplating the challenges, requirements, and ad hoc solutions related to the management of narrative within the domains of personal finance, ancestry research, hybrid cloud infrastructure security, neuroscience, and genomics.

### **Managing Information Gaps**

The need to manage information gaps was central in all fields considered, indicating a degree of overlap between various NIM features. In the case of ancestry research, it was not only essential, but the defining element in the domain—with a variety of ad hoc and platform-provided methods for identifying and resolving these gaps. Well-designed schemas and structures are used to help direct the attention of ancestry researchers to missing pieces within the knowledge base. In personal finance, externalization was a key solution to handling information gaps, both through community message boards and financial professionals. However, this externalization is accompanied by problems of its own—as the choices in who to trust is in itself a difficult challenge which has led in some cases to herd mentality and cult of personality. A key insight is that the presence of investing communities further complicates the space as the members are not just consumers but also components of the information economy. Where ancestry research and personal finance provided insights regarding implementation, the domains of genomics, neuroscience, and

trust management in hybrid cloud illuminated the need for information systems that facilitate handling the sheer volume and complexity of the gaps, as well as systems that highlight and acknowledge areas that cannot be disambiguated. For transdisciplinary challenges involving multiple domains (e.g., a genomics researcher investigating the structure of a viral protein in order to make recommendations related to medical policy), information gaps may need to be bridged both within and among areas of expertise.

### **Facilitating Situational Awareness**

Maintaining situational awareness was of obvious importance to hybrid cloud infrastructure security, where the need to monitor for security threats and vulnerabilities is constant, yet exploring these various domains indicated it's still vital in other areas, albeit in less pressing ways. Researchers in genomics, neuroscience, and in the sciences in general need to keep up to date on the never-ending stream of new literature, as do regulatory and funding agencies. In personal finance, situational awareness has some of the same aspects of time-sensitivity and risk-deterrence as those found in hybrid cloud infrastructure but with the added interest of spotting potential opportunities. This use of situational awareness for directing attention toward opportunities was more codified in ancestry research, where platforms are context aware and help bound scope to reduce cognitive load while prescribing actions. In order to make situational awareness achievable despite the high volume and complexity of information, personal finance and ancestry research were shown to primarily make use of streaming dashboard visualization and symbolic compression, whereas the fields of genomics, neuroscience, and hybrid cloud infrastructure security appeared to make more use of information fusion and modeling. An insight drawn from this distinction might be that both situational awareness and the processes by which it is achieved must be fit to the community. In other words: no single system would be of equal value to communities facing different kinds of informational challenges under different conditions even where those systems might benefit from common mechanisms.

## **Providing Descriptive and Explanatory Information**

The ability for individuals to dig into particular components and objects of the information environment to find description and explanation was of obvious value in all fields, to varying degrees. In particular, IT administrators and those attempting to understand the market are faced with near constant changes regarding which objects are of interest day to day, or even minute to minute, making capabilities associated with accessing and committing information to working memory far more pressing than capabilities associated with storing it. In making sense of very complex systems, the use of mental models, schemas, and codification of patterns of expectation appeared to be of great value to all fields.

## **Facilitating Exploration**

The ability to assist in the exploration of new information was emphasized in ancestry research and personal finance, where the untrained and self-educated are not provided with the same kinds of guides to the informational terrain as would be found in the sciences. In ancestry research, less focused exploration serves as a basis for helping to maintain the knowledge base, and, in the case of purposeful exploration, providing tools to help scope the needs and boundaries of exploration is potentially more important than providing curations of resources. As information volume expands, curation is simply not enough and recommendation systems need to be tuned to project and mission context, not just personalized to the individual's past interests and searches. In personal finance, exploration serves as a function of situational awareness—and here we acknowledge the need for methodology and tool transfer between domains, as those attempting to make sense of the market have an immediate, pressing need, yet do not have the kinds of tools available to ancestry researchers. This is seen in ancestry research as well, where it was noted that not even historical researchers have access to the kinds of tools of their amateur counterparts.

## **Compression**

Compression of information through visualization, structure, collation, curation, and interaction mechanisms was of particular interest as it was so often embedded as a basis for performing other functions. While some fields emphasized



certain mechanisms of compression more heavily than others, all were still relevant. The insight drawn from all fields in this case, is that this may be the most fundamental aspect of NIM—which is fitting, given that narrative itself can be considered an information compression mechanism.

### **Case Management and Prescriptive Information**

Case management functions were only emphasized in ancestry research and in hybrid cloud infrastructure security. However, the need to string together disparate events encoded in myriad forms, which may have otherwise been considered unrelated, was apparent in all fields. The insight drawn here, as has been drawn from other categories, is that there is a need for more tool and methodology transfer between fields. Case management methodology is highly generalizable, as discussed when introducing NIM features, and those working in genomics and neuroscience or those trying to make sense of the market or their finances could have large reductions in cognitive overload should tooling be made available. The importance of trust and the value of structure and codification of patterns in prescriptive information, or information regarding what the user should do or look to next, are seen in ancestry research, through its use of data schema and platform structure, and personal finance, through its use of externalization. From hybrid cloud infrastructure security, a key insight was the importance of prescriptive information in terms of scale—professionals in the space have to contend with such a high volume of events, that externalizing to some level of automation to prescribe or suggest action and to triage and prioritize tasks is not just valuable but inescapable. Finally, in neuroscience and genomics, prescriptive information was generally found in the processes by which individuals perform the work—however the communities informed by the sciences, such as patients, clinicians, and policy officials, suggest a need for cross-community prescriptive information, rather than a focus on provision of prescriptive information within the field itself.

### **Synthesizing Intelligence**

The need to synthesize extant information into new information products, similar to compression of information, appeared to be fundamental across the domains to varying degrees. There

is a clear need to improve information sharing between research-oriented and application-oriented areas within a given field to ensure more comprehensive and useful synthesis. In addition, all areas, as discussed when considering insights about case management, had demonstrable need for information fusion capabilities in the interest of developing new information products from myriad sources. Further, insights could be drawn from neuroscience and personal finance, pursuit of what is relevant to funding agencies and personal investments may affect the resulting syntheses, respectively. In terms of potential solutions, ancestry research was an arguably surprising place to have found such advanced mechanisms for rapidly and automatically producing coherent documentation, reports, and even entire context-specific books about particular research projects—this automatic rendering of content could be invaluable to researchers in other domains.

### **Facilitating Communication**

The facilitation of communication both within and between communities and users is the first of the features not included in the initial list. Much of the knowledge management and adjacent literature initially surveyed appears to assume, often for good reason, that the users of a particular, managed knowledge base will be a part of the same organization or profession. However, as shown in all sections, this will not necessarily be the case in practice. For example, in neuroscience and genomics, there is a complex interplay between scientists, researchers, governments, regulatory agencies, funding agencies, patients and concerned citizens, caregivers and counselors, and even ancestry researchers, as they share an abstract information commons without tools for managing the asymmetries in training, interests, and information access. A key insight can be drawn from both neuroscience and hybrid cloud infrastructure security, where there appeared to be a difficulty communicating between the application-oriented and theory-oriented aspects within those fields, as was noted in the discussion of intelligence synthesis. Facilitating communications within and between communities and users can enable both dialectics and interfaces for cross-community NIM.

## **Handling of Errors and Inconsistencies**

The importance of addressing error and inconsistency was not addressed as a primary concern within the literature initially surveyed, except where it concerned fraud in archive and records management. In the fields sampled, however, handling of errors and direction of attention toward inconsistency appeared to be of notable importance. In trust management in hybrid cloud architecture, detection, preventing, and handling of error and inconsistency in terms of permissions was a defining characteristic. In ancestry research, the lack of methods to contend with error accumulation in crowd-submitted annotations means the enormous corpus assembled is arguably useless to historical researchers. Moreover, inconsistency in details such as birth dates on two documents may suggest either differences in identity or bureaucratic errors and changes. Neuroscience-centered inconsistency, such as the differences between expected effects in human and animal trials, isn't always about correction, but instead about direction of attention toward information gaps and acknowledgement of complexity. This same insight can also be drawn from hybrid cloud infrastructure security where inconsistent behavior or expectations about use of equipment can signal vulnerabilities.

## **Management of Trust Signals**

An unforeseen addition to the list of NIM features was trust management, or more specifically, the management of trust signals. Our initial expectation was that trust management would be an area that would benefit from NIM, as opposed to an area which would be an explicit feature of it. As shown in numerous sections, contributions may contain errors, inconsistency, or be influenced by perverse incentives. Information quality in any knowledge base should then be expected to be somewhat unstable, and as such, there is a need to manage signals associated with the veracity and quality of information—lest all information become questionable, preventing users of the knowledge base from forming coherent narratives.

## **Social Systems Engineering**

As a final discovered feature, possibly the defining, fundamental characteristic of NIM systems is the treatment of

users as components of the knowledge base—not just consumers. In hybrid cloud infrastructure security, ancestry research, and personal finance, users are up against various tradeoffs while contributing to and interacting with aspects of information systems such as information quality and security, as well as aspects that run counter to the maintenance of the information commons, such as convenience, time, efficiency, and event reputation. In personal finance and ancestry research, where non-professionals make up a large portion of the interactions and contributions to their respective information commons, the risks, such as corruption of the corpus or the creation of feedback loops of negative interactions with the world outside the commons, are even higher. However, the benefits to narrative sensemaking which come from community involvement in the commons outweigh these risks, and there is a rich, social systems engineering literature to draw from in mitigating them. Investigating other domains may be of further value, as personal finance revealed the importance of role and duty assignment and judicial function, through the use of fiduciaries to moderate contributions to the financial information commons.

In this paper we proposed Narrative Information Management (NIM) as a term to describe the common set of system features that facilitate narrative sensemaking. In the interest of clarity, we define the term here as follows:

**Narrative Information Management:** The design, use, study, and implementation of aspects and features of processes and systems which manage information in order to facilitate narrative sensemaking.

With increasing fragmentation and information overload in the very domains which intend to address these challenges, we propose the term in the interest of helping to unify research interests and connect those research interests to requirements, challenges, and ad hoc solutions in the field. We do so with the caution which should accompany any introduction of new terminology, and with consideration for its economy (does it compress and communicate well for its size?), precision (does it refer to one idea only?), stability of definition (will this still mean the same thing a year from now?), and other aspects [309]. Whereas past introductions of similar terminology in the information sciences have generally divided or generated new fields [42], NIM may instead be of most use if considered as an analog to complexity theory centered in the information sciences, existing as a nexus or bridge between many disciplines purposed with facilitating the discovery and codification of

regularities, generalizations, and methodologies of global use. In the spirit of this usage, we conclude by offering recommendations for continuing work on NIM.

## **Recommendations**

- Continue the search for additional general NIM features through exploration of the challenges, requirements, and ad hoc solutions in various applied disciplines.
- Focus on development of common interfaces, common theory, and common data structures that help tools and communities communicate, rather than on singular, common tools. As evidenced by the exploration of the sampled fields in this paper, each community has their own unique needs, and no single platform should be expected to meet all of them.
- Developing education and curriculum around NIM and sensemaking in the interest of developing shared language and improving accessibility and communication of research on meta-sensemaking.
- Encourage interdisciplinary collaboration in research on information systems and their use in the interest of generating useful bridges and synthesis between fields.

# Contribution Statements

**Administration and Facilitation:** R.J. Cordes

**Initial Conceptualization:** R.J. Cordes, Daniel A. Friedman

**Analysis of Information-Centered Fields:** R.J. Cordes

## Section Authorship:

**Personal Finance:** V. Bleu Knight

**Ancestry Research:** R.J. Cordes

**Hybrid Cloud Infrastructure Security:** Shaun Applegate-Swanson, R.J. Cordes

**Translational Neuroscience:** Alexandra Mikhailova

**Genomics:** Daniel A. Friedman

**Editing and Revision:** All authors made substantial contributions to editing and revisions across all sections.

# Funding and Acknowledgements

R.J. Cordes is funded by the NSF Convergence Accelerator Trust and Authenticity in Communication Systems Program (NSF 21-572), under award ID #49100421C0036 and is supported in research efforts through a Nonresident Fellowship with the Atlantic Council on appointment to the GeoTech Center.

Daniel A. Friedman is funded by the NSF program Postdoctoral Research Fellowships in Biology (NSF 20-077), under award ID #2010290.

Thank you to Sam Young for contributions to early discussions on NIM and Satellite Management Systems.

## Chapter II

---

# Digital Rhetorical Ecosystem Analysis

## Sensemaking of Digital Memetic Discourse

Mridula Mascarenhas, R.J. Cordes,  
& Daniel A. Friedman

### Abstract

---

There are many areas of research defined by their interest in information dynamics related to facilitating organizational sensemaking, such as knowledge management, information management, and library science, and many more areas of research, disciplines, and even hobbies which are facing information-related challenges. While all may be concerned with very similar challenges, lack of information exchange and common ontology between these areas may be causing silos, missed opportunities, and potentially even friction among areas. In this paper, we address the need for synthesis and exchange of knowledge, tools, and approaches among various fields by proposing Narrative Information Management (NIM) as a unifying term and framework for the fundamental features and challenges of facilitating collective sensemaking. Through this framework, we offer an initial common set of features of impactful information systems found in literature on information-focused disciplines, such as knowledge management, and explore what insights and ad-hoc solutions may be found in an eclectic set of fields facing information challenges, including personal finance, ancestry research, hybrid cloud infrastructure security, translational neuroscience, and genomics. Finally, we offer recommendations for future research.

---

Digital Rhetorical Ecosystem Analysis was originally published in the COGSEC 2021 volume "Narrative Information Ecosystems: Conflict and Trust on the Endless Frontier".



# Part I

## A Rhetorical Approach to Understanding the Impact of Image Memes

We are in the throes of a widespread epistemic crisis that is damaging individual and collective sensemaking function and capacity ([1,2]). The crisis, articulated as “a state of affairs in which partisans disagree not simply on policy, but on facts themselves” [3], is attributed to a set of conditions including a “combination of political polarization, declining trust in media institutions, and asymmetric media ecosystems” ([3], para. 1). Concern about fake news, alternative facts, and misinformation has been escalating. Despite legitimate concerns about the degradation of public information due to the infusion of spurious content, we argue that viewing the information crisis as a competition between truth and falsity obscures the nature of the digital information crisis we are facing and, worse still, hampers efforts to restore trust and rework social consensus, which are essential for collective social action. Rather than approach the digital information problem as a battle between true and fake information, we urge attention to the rhetorical conditions and processes that contribute to eroding trust in established channels of information, and mainstream institutions and publics.

Framing the crisis as a battle between true and fake information has not proved effective in regaining the trust of those disaffected by mainstream channels of information. A simplistic true/fake dichotomy ignores the rhetorical conditions that have allowed competing narratives to displace mainstream ones. The hyper-complexity of digital information ecosystems is one such condition that makes achieving consensus on facticity and truth highly challenging [4], a condition that has, indeed, been exploited by malevolent actors. Nevertheless, addressing our epistemic crisis requires more than targeting and neutralizing sources of misinformation. We advocate a framework that combines rhetorical analysis with an ecosystem approach to trace the ebb and flow of narratives across digital publics. A rhetorical approach to understanding the information crisis focuses on message features that target audience vulnerabilities. An ecosystem approach goes beyond analysis of specific messages and audiences to highlight complex and long-term message-audience interactions, which can illuminate the changing web of narratives that influence public beliefs, opinions, and actions. Accordingly, we recommend addressing the epistemic crisis by developing a fine-grained understanding of the rhetorical forms and processes through which information circulates in the digital

public sphere and introducing rhetorical intervention as needed, rather than focusing exclusively on source control.

Contemporary digital information ecosystems create particular burdens on individual and collective capacities for reliable sensemaking and robust public discourse. The increased volume and diversity of information on the Internet create unprecedented cognitive complexity, and challenge clarity and social agreement on issues of public concern [5]. The default mode of online engagement—rapid surfing through endless streams of information, rather than focused deep immersion in selective limited information—further curtails information-processing capacity. Platform affordances and constraints, such as limited expressivity in communication (e.g., being encouraged to use a “like” reaction button in lieu of natural language elaboration on a post), the ability to rapidly scroll on digital screens, and the glut of emotionally charged material can also encourage peripheral rather than central processing of information [6–8].

Digital infrastructures also shape digital artifacts. The rhetorical features of these artifacts further encourage superficial engagement with online information. In our paper, we focus on one particular online artifact form—the image meme—that has played a crucial, yet understudied role, in destabilizing former epistemic foundations and traditional sources for public sensemaking. As we demonstrate below, the image meme has evolved into a ubiquitous unit of public discourse. Moreover, image memes function consistently as quasi-arguments in digital public spheres.

The word “meme” has gathered a great deal of semantic elasticity at this point [9,10], stretching from a general “unit of culture” to the specific genre and form of the image-macro [11,12]. We adopt a narrow definition of the image meme that allows us to capture and trace its role in public sensemaking. While the image macro refers to “captioned images that typically consist of a picture and a witty message or a catchphrase” [13], we use the term “image meme,” instead, because many specimens that draw from the image macro genre are devoid of text. In those cases, a juxtaposition of images within the meme compensates for its lack of textual elements. In image memes, configuration of the images themselves create meaning by making or implying arguments. We define the image meme by two features—form and function. The form of the image meme is established by the rectangular box frame which circumscribes one or more rhetorical elements, demarcating the meme as a discrete communication unit on platforms like Facebook, Instagram, and Twitter. While image memes perform a variety of rhetorical functions [14,15], we restrict our attention to image memes that play a particular rhetorical role—i.e., they participate in public argumentation by advancing claims [9]. In sum, the rhetorical artifact at the center of our paper is the ubiquitous rectangular box that is deployed to make a claim about a public issue.

The image meme has proved remarkably effective as a currency for public discourse, especially on Facebook and Instagram [16]. In particular, image memes have become integral to the destabilizing projects of the digital radical. They have been deployed strenuously in efforts to challenge and disrupt official and institutional discourses. The rhetorical dominance of image memes can be attributed to their ability to function argumentatively and, thereby, persuasively in the public sphere, constituting radical communities of discourse that are engaged in decoding, sharing, and amplifying their contents [17].

## What does a rhetorical approach to the study of memes entail?

Aristotle defined rhetoric as “the ability to see what is possibly persuasive in every given case” [18]. Rhetorical study emphasizes the *how* of persuasion. Therefore, a rhetorical approach to addressing our epistemic crisis moves us past solutions like banning digital sources of information or playing fact-check whack-a-mole with spurious message content, to focus on the persuasiveness of the message medium. While rhetorical critics are invested in analyzing message content, they are also invested in analyzing message form. The digital artifact at the center of our paper, the image meme, is a powerful example of the persuasiveness of rhetorical form. Repetition of form contributes to the crystallization of a rhetorical genre [19]. The widespread and increasing deployment of the image meme in digital public spaces has elevated the image meme into a rhetorical genre, one that is capable of charging a large scope of content with persuasive appeal.

Image memes have immense rhetorical power to shape online and offline sensemaking and action. During the 2016 United States election, Internet memes “enabled users to rapidly take a stand on and react to developing political events in real time; they provided alternative parallel discourses to mainstream media viewpoints; and they enabled mobilizing voters outside of official political discourses” [20]. The rhetorical power of multimedia memes has strengthened since 2016 [21,22]. Therefore, we argue for treating these artifacts as serious agents that shape public narrative and action.

A rhetorical approach to analyzing image memes can advance our understanding of their persuasive influence beyond the current practices of syntactic tagging of memes, for example by text recognition [23]. A rhetorical approach fills in the gaps endemic to tagging practices by enriching analysis of image memes with rich semantic information embedded in the parsimonious combination of the meme components. Symbolic cues in the memes not only advance logical claims but also encode ambiguous yet intense emotional charge that could spur public action. Interpreting cues within the meme against contextual knowledge surrounding the

meme is vital for the process of rhetorical analysis, and, as we will discuss later, computational analysis of digital discourse using a rhetorical approach.

A rhetorical approach encourages attention to the ways in which memes galvanize specific audiences to change their thoughts and actions. Image memes have constitutive potential; that is, they simultaneously call into being (constitute) audience groups while influencing audience thinking and possibly action—a process which rhetoricians call interpellation [24]. This constitutive potential is contained in the argument potential of the meme—its ability to advance claims, provide/imply evidence, and rely heavily on a discursive community to supply the necessary warrants (assumptions) to complete the argument [17]. The capacity of image memes to compel audience participation in semantic decoding contributes to the persuasive appeal of memes because the act of figuring out the meme's claim constructs the experience of truth-seeking, and consequently a sense of shared in-group identity, for the audience. Having successfully completed the decoding effort, audiences are interpellated as truth-seekers which enhances their investment in the meme's claim.

Another rhetorical feature of image memes that makes them conducive to interpellating audiences as truth seekers is that image memes are often free-floating. They seem to appear out of nowhere and do not typically disclose their sources unlike other digital content. As such, image memes represent an epistemic break. They gain credibility not because they arise from authoritative sources but precisely because they claim no source. The rejection of source credibility makes image memes a very powerful parallel discourse to more formal media channels and, in many cases, a direct challenge to information, claims, or narratives that emerge from publicly-vetted sources. When interpellated audiences decode and share image memes and engage in discourse about memes on forum threads, they build credibility for the meme in the absence of authoritative source credibility.

Therefore, tracking image memes (the claims they advance and the audiences they interpellate) in digital public spheres has become essential. Robust and far-reaching alternative and counter narratives circulate through social media platforms displacing mainstream narratives and flow under the radar of traditional mechanisms for capturing public belief and opinion. These online parallel currents of public discourse grew on social media platforms in relative obscurity between 2016 and 2020. The 2020 pandemic year, however, surfaced the proliferation of underground narratives when they started to manifest as widespread overt resistance to official COVID-19 narratives and policies, among large noticeable sections of the public. Towards the end of 2020, the galvanization of digital memetic energy around the visible public agitation against the 2020 US election results, culminating in the events at the United States Capitol on January 6, 2021, initially caught public officials and mainstream media off guard but subsequently drew

further attention to the robust discursive spaces in which competing narratives have been spawning and flourishing. Competing narratives have had and continue to have global impacts, as digital public spheres transcend the national boundaries of mainstream and official media channels. As researchers and organizations, interested in improving the immunity of digital public spheres to misinformation, invest in understanding the emergence of competing narratives, we urge attention not simply to the content of the narratives but, equally, to understanding of how those narratives are constructed through the circulation of digital artifacts, such as image memes. The philosopher Bruno Latour has noted that “whether or not a statement is believed depends far less on its veracity than on the conditions of its ‘construction’—that is, who is making it, to whom it’s being addressed and from which institutions it emerges and is made visible.” [25] To Latour’s list, we add the importance of attending to the rhetorical form in which the statement is packaged, i.e. the form of the image meme. Understanding the rhetorical form and function of image memes is crucial for any effort to observe, model, and respond to memetically-driven narratives.

## Rhetorical Anatomy of an Image-Meme

Although digital image memes can be used to circulate official narratives online, they have more successfully been deployed disruptively, across the political spectrum. Their truncated or compressed form is well-suited to inject targeted challenges to mainstream claims. The parsimonious form of the image meme provides a great deal of capacity for semantic encoding to advance persuasive claims while diminishing burdens of proof and elaboration that other rhetorical artifacts, like news articles, require. Various image meme formats such as text-only, image-only, screenshot, and image-text juxtaposition can all create polysemic affordances [26]; that is, the possibility of extracting multiple and multi-layered interpretations within a range of meanings. The strategic ambiguity inherent in memetic artifacts allows for rich semantic encoding. At the same time, the structural features of the memetic form (i.e., the containment of its content in a box, and the text/image syntax) strategically constrain meaning-making by setting up the key elements of an argument and cutting off counter-arguments. Below, in Figure 1 we illustrate the construction of an argument contained in one sample image-text meme.

Figure 1 constructs an argument with the simple juxtaposition of two lines of text above and below a stock photo. The choice of the photo combined with the double textual framing relies on the contextual knowledge of discursive communities to decode the argument. While the explicit memetic content is sparse, its signifying layers are rich, thus allowing the meme to argue a clear and persuasive claim.

The primary claim distilled from this image-text meme is that the official narratives about the origins of the SARS-CoV-2 virus, and the official masking policies to combat the virus, are not to be trusted. The rhetorical power of the meme draws from its strategy of juxtaposing two official narratives that appear to be mutually exclusive—that is, if the virus is virulent enough to escape the strict safety protocols of a world-class laboratory, then it can definitely penetrate the ordinary masks that the public has been asked to wear to stem the spread of the virus. The meme simultaneously alleges dissonance in official claims and expresses a snide disdain for those who accept the official narratives and are oblivious to the dissonance. The meme carries content designed to appeal to audiences' logical reasoning as well as to activate an emotional charge in the audience. The logic and emotion evoked by the meme are abetted by the meme's use of the "Condescending Wonka" image deployed memetically since 2011 to convey patronizing sarcasm [27].

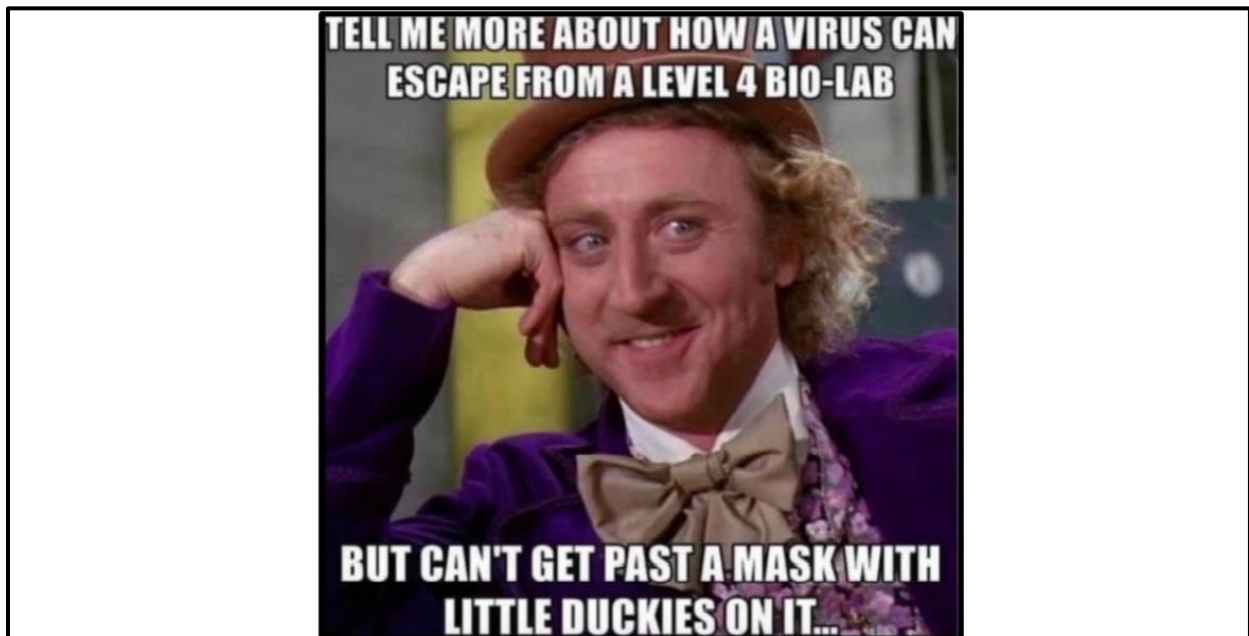


Figure 1. Rhetorical analysis Example 1. A "Condescending Wonka" image meme, with top text reading "Tell me more about how a virus can escape from a level 4 bio-lab", and bottom text reading "But can't get past a mask with little duckies on it..."

The two lines of text interspersed with the image interpellate an audience into the persona of Condescending Wonka, questioning with disdain, not only the official COVID-19 narratives but also the intelligence of those who have not yet figured out the contradiction. The meme positions the audience that agrees with its claim on one side against lying officials and people that trust official narratives on the other. The rhetorical deftness of this particular image text meme lies in its ability to swoop an audience, in the course of a single engagement with the meme, into both the line of reasoning set up by the meme and into an interpellated audience identity. That is, even as a viewer might be encountering the meme's reasoning for the first time,



having followed the reasoning and accepted it, the viewer comes to embody the persona of the one questioning the official narrative and condemning the naiveté of those who don't. The semantic decoding effort demanded by the meme works to enhance the credibility of the meme's claim by interpellating audiences as truth-discoverers. By advancing claims, memes not only shape public beliefs but also constitute powerful rhetorical audiences, knitting together discursive communities that share memes and bond over decoding and accepting memetic claims.

Furthermore, the boundedness of the image meme above (i.e., its containment within the rectangular box frame) and the parsimony of the rhetorical elements within the meme inhibit central processing and encourage peripheral processing of the meme's claim. The particular rhetorical form of the meme thwarts further questioning into possible reasons why the two supposedly contradictory claims may, in fact, not contradict each other. The success of the meme's argument relies on its ability to evoke the assumption that the initial event of the virus's escape signals its inability to be contained in any way. The possibility that the initial spread was virulent because the virus encountered an unsuspecting maskless population is elided by the memetic structure. Likewise the claim that masks only mitigate but do not necessarily prevent infection, entirely, is also obscured by the certainty evoked in the meme's juxtaposition of claims. Memes often simultaneously function as assertive yet weak arguments. Their weakness lies in the fact that their parsimonious form limits elaboration. However, this form feature is also responsible for obscuring the weakness of memes. The limited information, visually bounded by the meme's rectangular box, seals a particular conclusion while deflecting attention from warrants (assumptions) that could challenge the meme's claims.



Figure 2. Rhetorical analysis Example 2. The image foreground has hands that are using a pencil to write in a small book. The image background is blurred and appears to show a person on the left. The top text of the image reads: "So is 'Antifa' in the room with us right now, Karen?".

In the second example (Figure 2), we see intertextuality of memetic discourse at work because of the ways in which the image meme deploys another previously established meme, namely the Karen persona. This image meme attacks the claims that Antifa are responsible for some or most instances of violent unrest in the United States, for example during 2020. The primary claim available for decoding by an interpellated audience is that right wing hysteria both deludes and fuels itself by using Antifa as a bogeyman. The claim and inherent interpellation of a left-wing audience are achieved through multiple semiotic layers encoded in the meme's rhetorical choices.

Whether the memetic content is somber or lighthearted, explicit or implicit, memes are overwhelmingly deployed in the digital public sphere to assert and persuade through claim-making. The foundational intertextuality of memetic discourse demands that any study of memes as public sensemaking needs to go beyond rhetorical analysis of individual memes and consider how memes interact with and draw from each other to constitute, sustain, or destroy claims, and thereby narrative patterns, in response to unfolding events over time. Therefore, applying an ecosystem framework becomes essential to understanding how memes produce public sensemaking. Our next section details the rich potential in leveraging the ecosystem as a metaphor for studying the production and circulation of memes.

Ultimately, we coalesce a rhetorical analysis of memes and a digital ecosystem framework into our proposed Supervisory Control and Data Acquisition (SCADA) model for meme analysis. The SCADA focuses on identifying the key claim(s) embedded in image memes and the connections between memetic claims in order to trace the emergence, proliferation, and demise of public narratives on issues of public concern. The proposed SCADA system would provide a rich, real-time monitoring and analysis of narrative formation and propagation that circumvents limitations imposed by syntax and natural language-focused approaches. Further, open access to such a system would provide a counterbalance to both coordinated narrative influence campaigns and organic perturbations in memetic ecosystems, and provide more reliable analytic foundations for considering interventions to quell their effects.



## Part II

# Ecological Extensions of Rhetorical Analysis: Trends and Theory

Ecological metaphors for socio-technical systems have been applied productively to describe the physical and information aspects of the global operating environment, and recently notions of narrative, digital, and rhetorical ecologies are also gaining in popularity (Figure 3) [1,28–30]. Ecological or ecosystem metaphors for digital systems are applied as an integrative framework in different systems such as large-scale data analytics [31], “app ecosystems” [32] corporate strategy [33], and interactive role-playing games [34]. Across these diverse fields, ecosystem metaphors can encourage holistic analysis and connect abstract concepts to tangible systems and accessible experiences.

The idea and terminology of a “digital ecosystem” has been used since at least the 1980s, and has seen exponentially increasing use since the early 2000s (Figure 3B). A search using Google Books Ngram viewer revealed the recent growth of research interest in applying the ecosystem metaphor to online discourse (Figure 3A). While there is new interest in “digital ecosystems” as a term, as well as “narrative ecosystem” perspectives, the term “rhetorical ecosystem” is entirely absent from the literature corpus (Figure 3B).

Multiple previous works have applied the ecosystem metaphor to address questions related to digital discourse and memes. For example, empirical work on various popular websites has deployed the ecosystem metaphor to study the dynamics of the “meme ecosystem”. These studies have analyzed copyable plain text memes, sometimes referred to as “copypasta”, [35] as well as shareable image memes [36]. In these studies, the text and/or image data are downloaded en masse from publicly-accessible platforms. The ecosystem metaphor stands in the background referring more to the broad scope of data collection, rather than in the foreground as an appeal to see the data emerging from an ecosystem (e.g., analyzing the data in terms of interaction types among agents in an ecosystem).

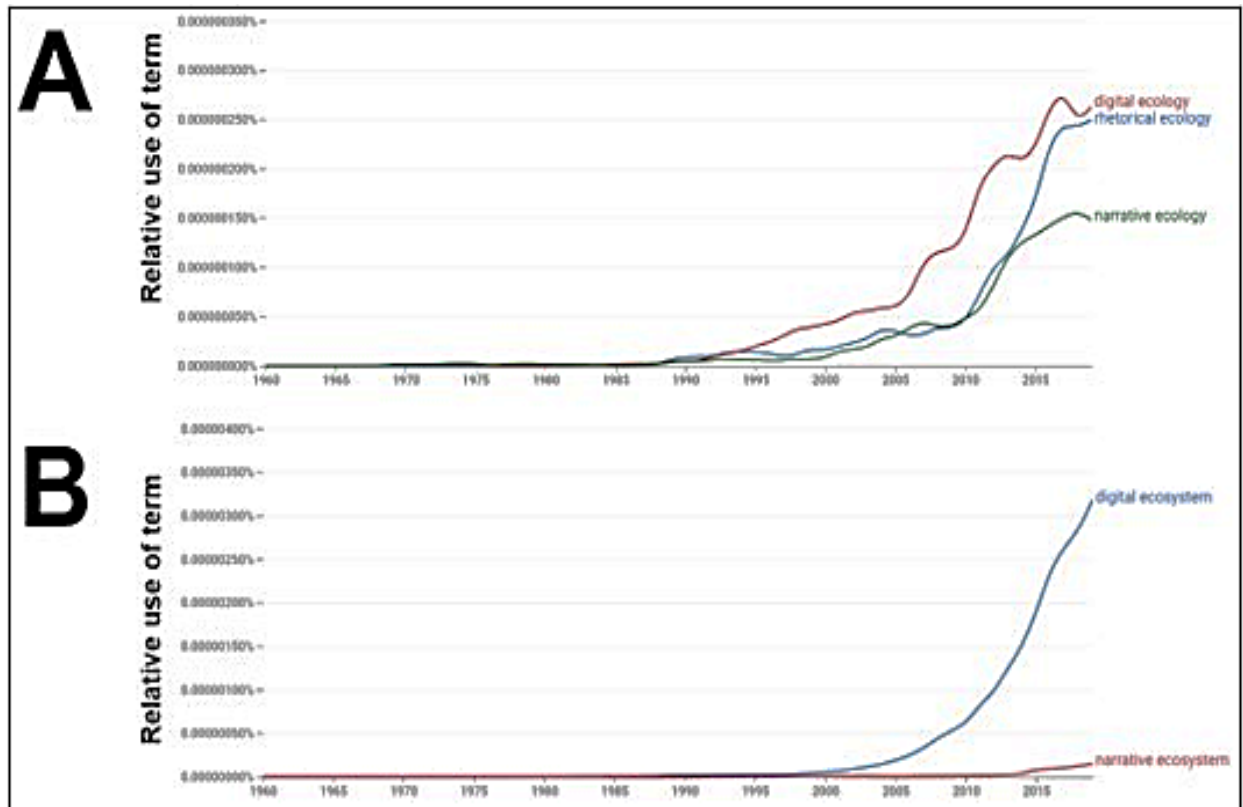


Figure 3. Trends in the usage of keywords in the Google Books Ngram search engine. Search terms used were (digital/rhetorical/narrative) + (ecology/ecosystem). A) Google Books Ngram search for “rhetorical ecology” (green), “digital ecology” (blue), and “narrative ecology” (red), from 1960-2019. B) Google Books Ngram search for “rhetorical ecosystem” (green), “digital ecosystem” (blue), and “narrative ecosystem” (red), from 1960-2019.

This suggests that the ecological metaphor applied to rhetoric (especially online rhetoric) has been conceptual and qualitative, drawing on conceptual similarities with ecology but not formulating ecosystem models or deploying recent developments in ecological toolkits. Thus we worked from the assumption that pragmatic implications for high-throughput rhetorical analysis of online discourse might be found in ecology, if the connections could be drawn out more clearly.

## Part III

# The Digital Rhetorical Ecosystem

### Three-Tier (DRE3) Model: Mappings, Applications, and Implications

For research into socio-technical systems and digital discourse, the field of ecology provides much more than qualitative metaphors. Others have offered a variety of fundamental points of contact between ecology and rhetoric, noting that both fields explore how systems exhibit multiscale patterns of organization arising from interactions among many subunits [37]. Both rhetoric and ecology study how information is communicated through time, and how agents interact with or modify their context. In the case of rhetoric, this is through the production, perception, and interactions with artifacts and social entities, and in the case of ecology, this is the phenomena of niche modifications or stigmergy [38]). Here we extend the interface between rhetoric and ecology to argue that the mapping between these two domains can find productive application in the monitoring and design of digital ecosystems. The specific implications of ecosystem metaphors for digital discourse are explored in the following section.

“Rhetorical ecology” is an established term (Figure 3A) that refers to the context-dependent rhetorical implications of texts as they are deployed in changing spatio-temporal contexts. The concept of “rhetorical ecologies” has been used to describe the level of modeling and abstraction that generalizes above any given rhetorical situation or element [39]. The ecological framework surfaces relationships between texts. For example, in ecology, the concept of a predator-prey relationship refers broadly to a type of behavioral interaction between two species, where one species consumes the other. Understanding that two species are in a predator-prey relationship helps make sense of an otherwise-disconnected set of questions and observations in the world, for example the daily activities of both species and their bodily morphology. In the case of rhetoric, we can also imagine predator-prey type relationships—for example two digital communities connected because one systematically follows and attacks the other, through memes. Additionally, online ecosystems may present totally new kinds of relationships among interacting agents; so any framework for rhetorical ecosystems should be able to infer novel types of relationships without being limited to the archetypes present in wild ecosystems (e.g., predator-prey as above, symbiosis, mutualism, parasitism). We hypothesize

that with appropriate ecological-rhetorical mappings in hand, new sets of frameworks and tools developed to study ecosystems could become rapidly useful for analysis of online discourse.

Here we introduce the Digital Rhetorical Ecosystem three-tier (DRE3) model (Figure 4) which expands previous work on the ecosystem metaphor for online systems and builds towards system design implications for analysis of memetic discourse.

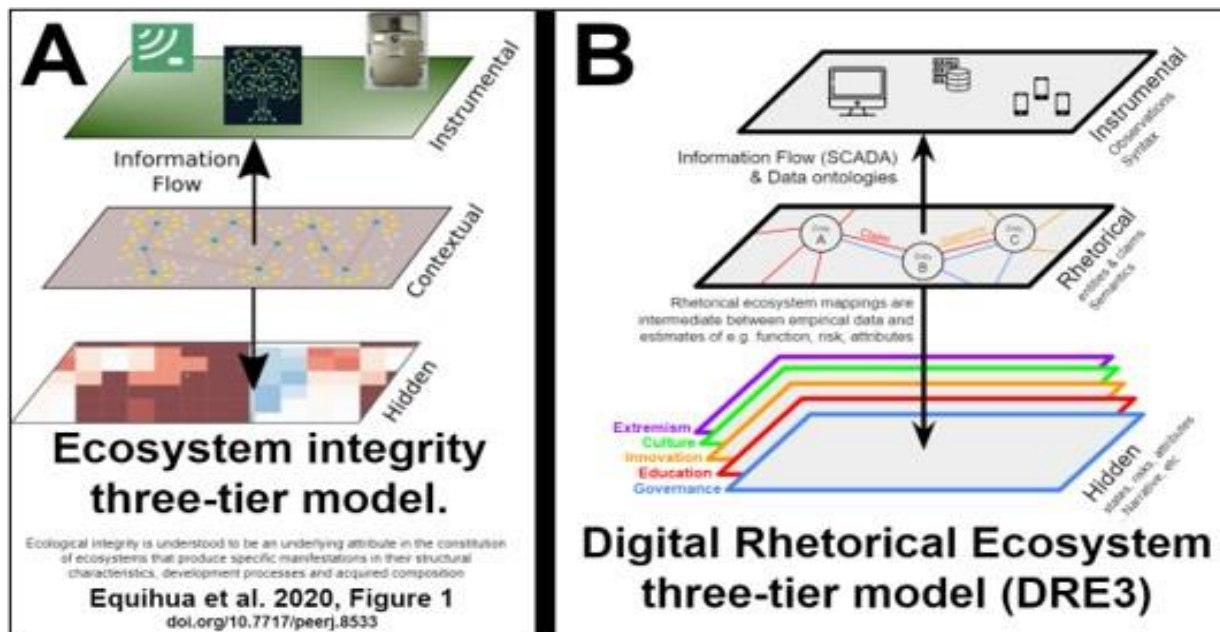


Figure 4. Ecosystem integrity model & the Rhetorical Ecosystem three-tier (DRE3) model. A) Figure 1 reproduced from Equihua et al. 2020 [40]. B) Digital Rhetorical Ecosystem Three-tier model.

The DRE3 model was inspired by the three-tier model of ecosystem integrity (3TEI) developed by Equihua et al. 2020 [40] (Figure 4A). In their 3TEI, the topmost tier is the Instrumental tier, reflecting measurements from the world, for example by sensors or cameras. The middle tier of the 3TEI is the Contextual level, reflecting the network of interacting agents in the niche that give rise to the observed information at the Instrumental tier. The bottom tier in the 3TEI are the Hidden variables of the ecosystem, such as risk of fire or capacity for agriculture. These variables are not directly observable through the use of any kind of physical instrument—hence statistical tools must be used to infer these states from the Contextual states that are in turn estimated from the empirical data at the Instrumental tier.

For the DRE3 model applied to digital ecosystems (Figure 4), we translate each of the tiers from the 3TEI into corresponding domains related to online discourse. The Instrumental tier of the DRE3 reflects the empirical observations of digital activity, for example rhetorical artifacts such as image memes, as well as metadata and other platform information (e.g., traffic logs, user ratings or responses to content). The

middle tier of the DRE3 is the Rhetorical tier. This Rhetorical tier reflects the networks of entities, claims, and warrants evoked by artifacts at the Instrumental tier. The bottom tier in the DRE3 reflects the multiple possible Hidden layers which might be significant targets of analysis, for example the risk of extremism, production of subcultures, degree of innovation, quality of public information, trust in government, and process of governance.

Importantly, the information in the Instrumental tier is mediated and augmented by a Rhetorical tier in the process of Hidden State inference. The direct mapping from rhetorical artifacts to hidden state inferences can be challenging and noisy (e.g., in the case of hashtags or syntax-driven analyses used to identify conspiracy theories [41]), or essentially impossible (in the case of image and multimedia artifacts). A better approach to high-throughput analysis of multi-media digital discourse is needed. We suggest that the introduction of a rhetorical layer (consisting of entities, claims, and warrants) in between the instrumental and hidden layers is a useful direction to pursue.

## Ecology: Key Concepts and Mappings

This section applies the DRE3 model in the context of the modern global information environment. Like insights gleaned from regional ecosystems [42], analyses of rhetorical ecosystems ideally should be use-oriented, in close-to-real-time, and able to be represented differently for different stakeholders. Contemporary and future analysis of online discourse will involve the use of heterogeneous data to detect, monitor, and perturb discourse. This requires a significant amount of actionable and estimative intelligence regarding the real-time state of online discourse, especially if the goal is to ameliorate the aforementioned epistemic crisis and increase the capacity to understand and respond to the use of image memes in online discourse.

In this work we do not present any formalisms or explore all possible ecosystem-rhetoric connections, but rather focus on deriving implications for rhetorical analysis and online system design by focusing on three key areas of ecological theory and application:

- Multiscale perspective on ecosystems
- Ecosystem antifragility
- Ecosystem services

For each of these three ecological topics, we 1) define the term, 2) clarify the mapping from ecology to rhetoric, 3) consider which concepts might transfer from ecology to rhetoric, and 4) provide a preliminary investigation of the implication of these mappings in terms of systems design.

## Multiscale Perspective on Ecosystems

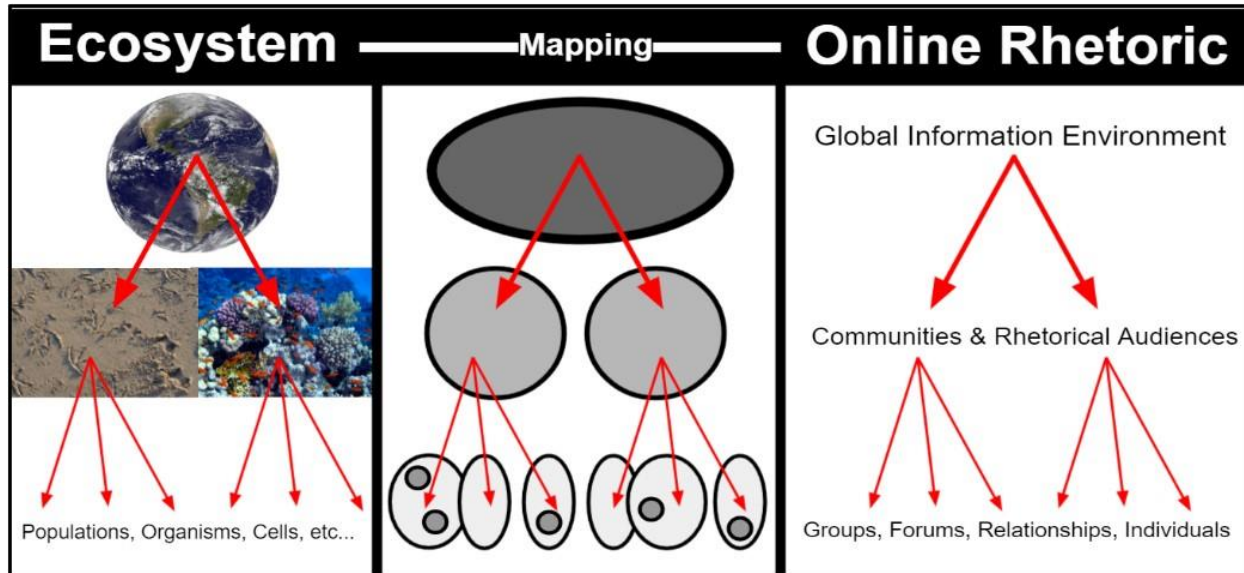


Figure 5. Representation of the multiscale perspective on Ecosystems. At left, ecological modeling of the world can proceed via decomposition into disparate ecosystems. At right, online rhetoric occurs within the global information environments, via increasingly-fragmented platforms, channels, and chats. The common mapping, in the middle, is the notion of overlapping and nested systems.

### What is the multiscale perspective on ecosystems?

- Modern ecological frameworks are built around the idea that biological systems present as nested scales of organization [43]. At each scale of organization such as cell, organism, and population, the system consists of interacting agents of various types [44,45]. System subunits can interact in non-linear ways, and the integrated function of the ecosystem as a whole can be considered as cognitive in its own right in that the system can learn, integrate information, display persistent memory, and act in an anticipatory fashion [46].



### **What is the mapping from the multiscale perspective on ecosystems to online digital discourse?**

- Today's digital landscapes consist of human and non-human agents, interacting with each other and with textual artifacts, as if they were on rhetorical landscapes. Ecosystems and landscapes are rich and generative metaphors that help capture the many ways in which agents of various types and in various roles interact massively in parallel. These distributed rhetorical interactions contribute to information integration, collective decision making, memory, education, and anticipation across the digital public sphere. Rhetorical ecosystems exhibit structure and regularities across multiple scales of analysis, for example the individual, relationship, group, and community. Thus digital rhetorical ecologies can be considered as an integrated multiscale cognitive system.
- The case of an image meme posted on a social media platform can be seen as a niche modifying action of mobile agents, with the intention of signaling to similar or dissimilar agents, resulting in functional consequences for the further evolution of the biosemiotics of the niche. These stigmergic processes in nature, such as an ant depositing pheromone, or large mammal making territorial markings [47,48], are essential for ecosystem function. Digital platforms present affordances for niche modifications, whether extremely limited (e.g., only a "like" button"), or more extensive (e.g., a Wiki model where content can be edited, or even a platform where the code and affordances can be modified by users). The availability and incentives for using different kind of digital affordances will be user-, platform-, and context-specific. This corresponds to ecosystem contexts where contextual niche modification processes play out over rapid behavioral timescales versus slower evolutionary timescales.

### **Which key ideas and tools from the multiscale perspective of biological ecosystems transfer to digital discourse spaces?**

- Ecosystems around the world vary in fundamental ways but still can be modeled with common frameworks. Similarly, in the case of online discourse, we are interested in the similarities and differences across languages, platforms, and settings. The multiscale perspective in ecology highlights how interacting agents and situations can generate emergent patterns that are stable (or metastable/oscillatory) within acceptable attractor states,



rather than causing cascading failures [49,50]. In ecology, even antagonistic interactions such as predator-prey may be stabilizing at the macro scale. In the case of online rhetoric, we might map individual-level interactions to behavioral ecology, and group-level dynamics to macroecological outcomes. For example, a pairwise relationship might be unstable or antagonistic among two users of an online platform (behavioral ecological scale) yet be a part of a stable broader online community of users (macroecological scale).

- The idea of niche modification from ecology translates to the kinds of changes that agents make to their information niche. In the case of online communication, this is known as digital stigmergy [51,52]. Just as the behavior of individual animals is nested within (and in feedback with) surrounding ecosystem dynamics, rhetorical agents are actively exploring and modifying their informational niche.
- Various ecological toolkits exist to infer agent states and actions across spatial-temporal scales and use these inferences to understand how agent behavior is in feedback with broader trends. These toolkits include software packages and approaches related to movement tracking, multi-scale network analysis [53], system simulation [54], and characterization of the relationship between animal behavior and the animal's niche [55–57]. In the case of online discourse, agents are moving across informational landscapes, updating their models of the world, interacting with other agents, and increasing or decreasing their likelihood of engaging in different kinds of action. In both ecological and rhetorical settings, one may be interested in modeling how interaction among agents influence individual and collective behavior, as a function of context in the niche.

## Ecosystem Antifragility

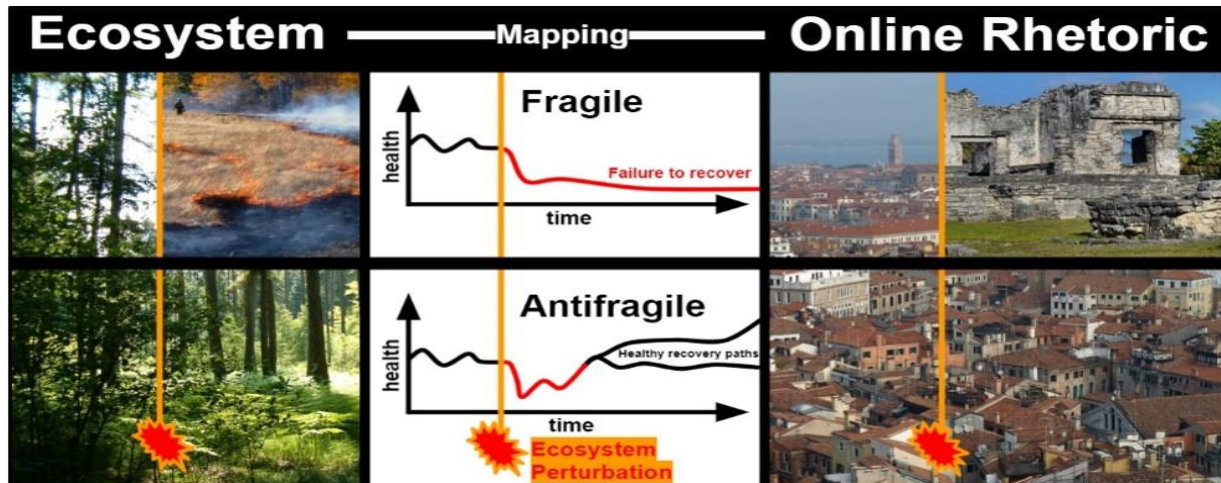


Figure 6. Representation of the concept of Ecosystem antifragility. At left, a forest experiences a perturbation such as a fire event. This event may either lead to devastation of the forest (top), or result in a forest that either burns completely and/or grows back stronger (bottom). At right, using a city as an analogy for the online rhetorical commons, a perturbation event can result in a destroyed commons (top), or a strengthened and vibrant community (bottom). The common mapping, in the middle, connecting biological ecosystem antifragility to digital ecosystems is that complex systems can undergo various recovery or response dynamics in response to perturbations, broadly classified as fragile (failure-prone, top) or antifragile (resilient and regenerative, bottom). For digital discourse platforms, fragility would refer to the inability to adapt or recover function following technological or rhetorical perturbation.

### What is ecosystem antifragility?

- Ecosystem antifragility refers to the vibrancy, stability, and dynamic variability of a system. Recently, Equihua et al [40] have used various approaches from Complexity science to describe ecosystem antifragility as “beyond resilience and integrity”. Their working definition is that an “ecosystem is antifragile if it benefits from environmental variability” [40]. Antifragility is similar to the notion of resilience, which captures how a system resists change or returns to functional capacity after a perturbation [58]. However, antifragile systems are those that actively grow or increase in capacity after stressors, as opposed to merely returning to previous operating modes.

### **What is the mapping from ecosystem antifragility to online digital discourse?**

- **Health.** The stability and flourishing of the rhetorical commons is a primary goal for participatory communities and societies. This is akin to the concept of ecosystem health: even where different regions or seasons may have distinctly different healthy modes, maintenance of ecosystem vitality may be an overarching regional goal. While humans have long relied on qualitative or felt measures of ecological health, quantitative data collection allows for entirely new measurable notions of health only enabled by instrumentation and modeling [59–61]. We highlight the need to develop statistical indicators for the health and vitality of digital ecosystems so that policy for and management of digital commons spaces can be driven by shared empirical understanding rather than the potentially discordant experience of individuals.
- **Resilience.** The resilience of a rhetorical ecology might be defined in terms of the system's maintain function during a crisis, informational update, or structural change. The resilience metaphor draws attention not just to the regular or functional operating modes of rhetorical ecosystems, but also to the emergency and recovery modes available to these systems. Ecosystem resilience is critical when humans have a vital dependence on continued ecosystem function, as in the case for agriculture [62]. Increasingly, online communications are a lifeline, and thus also need to be managed carefully with uninterrupted service and content integrity in mind. Disruption of internet services can occur through physical damage to infrastructure, as well as software intrusions (e.g., ransomware, denial of service attacks). Even when hardware and software are running according to performance standards, breakdowns of sensemaking (e.g., due to spam, targeted disinformation) can lead to perturbations on digital platforms and breakdowns in their typical functioning.

### **Which key ideas and tools from antifragility perspectives of biological ecosystems transfer to digital discourse spaces?**

- Ecological antifragility has several kinds of ideas and tools to offer to the domain of rhetoric. Equihua et al. [40] characterize antifragile systems as those that benefit from variability, which provides a valuable parallel for measuring the health of the rhetorical commons by its type and extent of diversity (here of rhetorical

claims and perspectives, rather than, for example, a species number). That the variability of rhetorical claims can be a source of collective vitality provides a helpful starting point for viewing online discourse and dissuades approaches that promote total consensus as a goal, or reflexive suppression of alternative viewpoints.

- Some approaches towards ecosystem antifragility feature participatory roles for ecosystem inhabitants, for example local cleanup events, long-running citizen science projects related to birdwatching [63] and regional ecosystem biodiversity events like a BioBlitz (“an event that focuses on finding and identifying as many species as possible in a specific area over a short period of time” [64]). In the context of digital ecosystems, these kinds of local programs for ecosystem improvement can scale to include large numbers of participants, for a Wiki editathon, for example [65,66]. Coordinated efforts to “fix trails” in digital ecosystems could contribute to antifragility by providing a scalable approach for reducing risks from cascading or complex failure modes related to out-of-date information, fragile network structures, or incapacity to deal with anomalous system usage.
- Quantitative tools also exist to help stakeholders measure and model ecosystem antifragility from a Complexity perspective [67]. Dynamic models allow for simulation and analysis of various kinds of systems and their stability in different situations [68,69]. In the context of ecosystem health, these kinds of analysis ask how it might be possible to build stable networks rather than network structures. An exclusive focus on network structures might lead to fragility of network function when edges are lost or nodes change. Modeling ecosystem health as a phenomenon arising from interacting networks, offers new and potentially more-effective ways of thinking about how multiple ecosystem stressors interact [70]. Network models also can be expanded to include “games on graphs” models, which use the tools of game theory to explore how strategies interact on landscapes and how information propagates through groups [71,72]. In the context of digital ecosystems these kinds of models could provide descriptive, prescriptive, and proscriptive information on the general function and well-being of digital platforms.

## Ecosystem Services

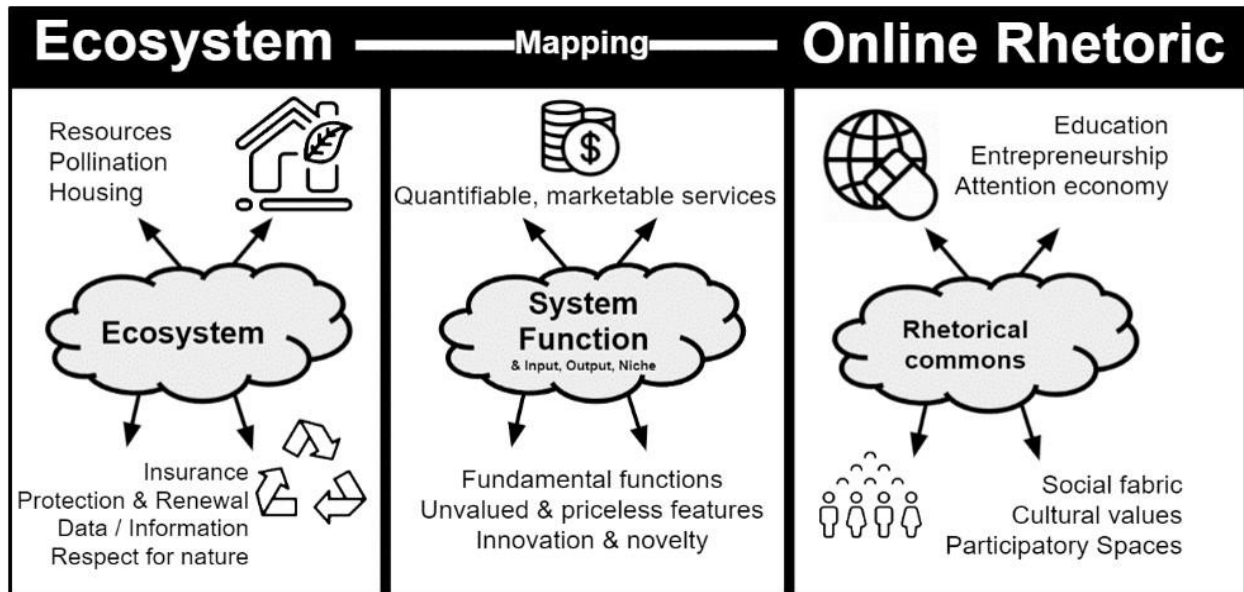


Figure 7. Representation of the concept of Ecosystem services. At left, physical ecosystem services such as natural resources and pollination are enacted by various actors within ecosystems. At right, online rhetorical commons can be considered to enact or emit services such as education and innovation. The common mapping, in the middle, is that value and valuable outcomes are generated through the function of the target system. Putting quantitative value on “intangible” outcomes can be challenging. Seeing online outcomes as analogous to ecosystem services is not a solution in and of itself, but rather a framework for approaching system management and design.

### What are ecosystem services?

- Ecosystem services are the functions that ecosystems provide which are useful for humans directly or incidentally, for example the provision of food, erosion control, composting of decaying matter, recreational spaces, or generation of natural resources, [73]. As is the case with ecosystem antifragility and health, many types and measures of ecosystem services exist.

### What is the mapping from ecosystem services to online digital discourse?

- If we imagine rhetorical ecosystems to encompass the biotic and abiotic aspects of the system that contribute to its function and regulation, “rhetorical ecosystem services” could include a broad range of outcomes, including education, communication,

innovation, and development of cultural norms and practices. Just as high-level biological ecosystem services, like the production of food, arise from direct interactions among many kinds of actors (e.g., plant, pollinator, microbes), and might be influenced by indirect factors as well (e.g., noise/light pollution, presence of predators), rhetorical ecosystem services emerge from the direct and indirect interactions of many actors and artifacts in the space. Understanding these influences can support modeling and management of the valuable outputs of a rhetorical ecosystem.

- We can consider image memes as a special case of ecosystem services, in that image memes are valued or relevant products of an underlying ecological process. The image meme format reflects the intersection of digital content production affordances, and the rhetorical cross-pollination occurring online. The services that image memes provide in the rhetorical ecosystem can include advertising, information sharing, governance, entertainment, persuasion, and more—essentially any functional outcome of the deployment of image memes that can be tracked and valued.
- Other studies have investigated the dynamics by which images memes originate and diffuse through time among communities [36]. This is akin to a source-sink analysis common in ecology: source locations are net exporters (of image memes on digital platforms) while sink locations are net importers (on digital platforms reflecting image meme consumption) [74]. This source-sink analysis of image memes can link the dynamics of memetic spread to their function for different audiences, and thus shed more light on the causes, context, and consequences of particular image memes for the rhetorical commons.

### **Which key ideas and tools from ecosystem services transfer to digital discourse spaces?**

- Conservation & management of ecosystem services is an area of practice with a long history of analyzing the intersection of human individuals, human groups, and the rest of the biotic and abiotic surroundings. Some of the legal, mathematical, scientific, and game theoretic approaches to ecosystem services might transfer usefully to cases of online rhetoric. For example, when considering the design or regulation of digital platforms, various areas of law and policy interact, for example finance, business, and privacy. Framing digital platforms (and the functions they



perform) as ecological commons introduces precedent for addressing legal dimensions of individual/public/private ownership, and processes for dispute resolution related to common resources [75].

- Ecosystem antifragility (discussed above) plays directly into the stability and accessibility of vital and valuable services [76]. Healthy rhetorical ecosystems will display variability in productivity through time. However, an ecosystem at high risk of catastrophic failure cannot be considered as valuable as a dependable ecosystem (e.g., a forest at risk of destructive fire presents higher uncertainty about its future productivity). The relationship between ecosystem health and productivity provides an economic motivation for policies that balance multiple contrasting requirements, by thinking about system function through time.

## Implications

We argue that insights from modern Ecology can help scaffold the future of computational systems for rhetorical analysis. Ecological perspectives can retain the semiotic insights from rhetoric analysis while tracing meanings and their interactions within a quantitative framework [37]. At this time, manual rhetorical analysis requires trained experts who identify how artifacts produce meanings for different audiences, or, in the case of image memes, how memes generate claims. This process of rhetorical analysis is analogous to a natural historian observing a species operating skillfully in their niche, in that a specific occurrence (observation of a bird, or a digital text) is modeled in terms of its relationship to the context and niche (whether biological or rhetorical). Computational frameworks for rhetoric provide a set of ideas and tools that, if properly designed, could help accelerate rhetorical claim analysis. This type of “next-generation natural history” [77] for rhetorical ecosystems would integrate well with existing computational frameworks, apply well to the multimedia setting, and also work toward grounding analysis of digital discourse in rhetorical principles. Functionally, Ecology is the bridge that would allow rhetorical information to play a more central role in the computationally-aided analysis of contextualized digital discourse. We suggest that, in addition to the quantitative tools it provides (such as network analysis, sparse sampling, agent-based modeling, meta-community dynamics), Ecology can supplement rhetorical analysis by foregrounding concepts like ecosystem health, biodiversity, anti-fragility, and more. Below are some possible implications arising out of the application of the Ecological perspective to online rhetorical commons (by no means comprehensive).



- Create and adapt within the niche. Online platform and systems designers can ask what services they are providing to stakeholders and the broader ecosystem (defined as the entities, audiences, and cyberphysical systems constituting the stakeholders and zone of influence of the target platform). Platforms provide and interact with the rhetorical commons, and thus services of value are being provided or modified by them. As digital platforms require inputs from the broader ecosystem in terms of energy, attention, and other resources, platforms must be anticipatory and responsive to changes in their operating ecosystem.
- Trace artifacts and claims to understand function. The DRE3 model of digital discourse has the capacity of creating clustering, detecting thresholds, or permitting inference at the level of rhetorical claims, an extension of approaches built solely on syntactic inputs (e.g., hashtags, keywords) or lexical semantics (e.g., natural language processing, sentiment analysis). We need to integrate artifacts and claims (beyond, or instead of tracking individuals) for effective sensemaking of digital discourse. Thinking of claims in terms of functional patterns in the ecosystem, platform designers could analyze the relative fitness and spreading/mutation/co-occurrence dynamics of memetic claims, across communities, languages, media formats, and platforms.
- Consider dynamics, not just snapshots. Some of the dynamical systems and network analysis tools developed for ecosystem management could generate models that may transfer directly to online datasets. Similar kinds of observations can be made in the ecological as well as digital situation (e.g., about the movement or communications among agents through a space described as a network) and similar kinds of questions might be asked (e.g., which initial conditions and patterns of relationships might result in stable vs. unstable regimes). For example, migration can occur among geographical distances as well as among digital communities on social media. Complementary tools and perspectives for the analysis of migrations might be found across research on patterns of ecological and digital migrations [78,79].
- Design for multiscale interactions. Online platform design could take the multiscale perspective directly into account, for example by making certain peer-to-peer interaction mechanisms transparent, so that agents at various scales (e.g., individuals,

groups, communities) are aware of how user-level affordances influence the niche and system as a whole. Top-down (e.g., platform-dictated) and bottom-up (e.g., user-generated) signaling mechanisms could be clearly marked (or if not marked, could be annotated as such by analytics platforms).

- Fit generative models (of rhetoric) that can deal with sparse as well as complete data. The task of ecosystem characterization is to go from sparse and heterogeneous observations (for example ambient conditions and bird sightings through time), to a useful and communicable model. This task of ecosystem characterization, depending on the scope of the analysis and desired level of detail, may require multiple kinds of models to be specified: the cellular, organismal, social, community, and ecosystem. For online discourse, integrating the multiple scales at which decisions are made (human internet user, community, networks of networks), ecologically-informed models might provide a principled path for modeling various phenomena of interest.
- Think about the ecosystem's leverage points and failure modes when designing an intervention. Ecosystem modification efforts are famously non-linear—careless interventions may be ineffectual or even have deleterious effects (as in the case of using broad-spectrum toxins in an attempt to eradicate the fire ant in the Southern USA [80]). For social discourse, influence operations used to be evaluated in terms of a direct rhetoric source, such as centralized media. Now the operating landscape is much more akin to a complex ecosystem, contextualizing diverse social strategies as types of social ecosystem modification [81]. Modifications of the rhetorical ecosystem through various means (e.g., algorithmic distortion, misleading information) might have behavioral consequences rippling out far beyond the locus of direct action, akin to the introduction of a new species to an ecosystem. The relative efficacy and risk of different ecological interventions is variable across different regions. Proactive, long-term interventions such as restoring native habitat are often at odds with short-term interventions like intentional introduction of novel predators (as in the case of the cane toad in Australia [82]) or application of broadly-acting chemicals. Ecosystem interventions are irreversible, and often have non-linear consequences for different kinds of actors and audiences [83,84].

- Consider humans in the design of platforms, as well as non-human and computational actors. Taking a human-centric perspective on ecosystem function would be incomplete or even fallacious, depending on the region and goals of ecological modeling. Similarly, today for online discourse, given the prevalence and influence of purely-computational agents or computationally-augmented humans, it is essential that platforms be designed for use by human and non-human agents. Already a significant fraction of internet activity is carried out by purely computational agents or networks (e.g., chatbots and automated accounts). While the exact amount of human and computer activity likely varies among destinations, already in 2016 it was estimated that certain types of internet activity might be majority non-human [85,86]. The multiscale cognitive perspective on ecosystems provides a framework for modeling rhetorical ecosystems consisting of only human actors, only computational actors, and any conceivable composition in-between [87]. Already falling within this scope are existing tools that distinguish the activity of human vs. bot actors online in games, forums, and other platforms [88,89].
- Frame healthy and antifragile rhetorical ecosystems as a common pursuit. Promoting antifragility is a broad social goal that can apply across systems and scales. Ecosystem health as a concept helps humanize otherwise-unrelated environmental phenomena and might be able to play a similar role in making online rhetoric more tangible. Exact specifications of “health” for the digital commons may differ, just as they do for ecosystems. Analyzing the health of a given ecosystem might require the consideration of the abundance, composition, diversity, function, and tolerance of various kinds of life forms in the system (such as microbes, invertebrates, plants, etc.) [60]. And even in this case, individuals may still disagree on the health of a given ecosystem, if for example they diverge on the optimal usage of the region (e.g., for development vs. recreation vs. agriculture). When designing platforms for digital discourse, it would be valuable to consider how differences in opinion about “what is healthy” among users could be harnessed and channeled, rather than lead to system failure.
- Use rhetorical measures as a diagnostic when modeling digital discourse by framing the resulting artifacts and functions in terms of ecosystem services. Failure of rhetorical ecosystem services could occur from an adversarial or unhealthy dynamic,

such as an inability to communicate leading to breakdown of trust among otherwise-cooperative individuals. To thwart, or recover from, such failures, it could be helpful to search for analogous situations in ecology. For example, ecosystem services could be threatened by the introduction of an invasive new species, a toxic chemical, habitat fragmentation, light/sound pollution, or loss of biodiversity [90,91]. In the case of rhetorical ecosystems, being able to connect failures of services to past ecosystem interventions or modifications (influx of new users, introduction of toxic rhetoric, alteration of platform affordances, etc.) could provide a useful lens for protecting the valuable outcomes of digital discourse.

## Part IV

# The Digital Rhetorical Ecosystem Three-Tier Model Example Usage

The Digital Rhetorical Ecosystem three-tier (DRE3) model (Figure 4) integrates enriched rhetorical analysis of multimedia discourse with ecological theory and modern computational analytics pipelines. In this section, we present examples of rhetorical analysis using the DRE3 model. Specifically, we describe three analytic phases in the context of “boutique meme analysis” using two examples. At the end of the section, we provide a bridge between the traditional methodology of rhetoric and the types of computational representations that are useful for modern digital sensemaking systems.

There is a lack of usable platforms for computational rhetorical analysis, although several prescient calls have been made for such frameworks and tools [92–94]]. Partially, this gap exists due to the challenge of accurately and effectively scaling expert rhetorical analysis. While multiple complicated sub-tasks are required for rhetorical analysis, digital tools exist today to carry out some similar functions (such as face-, voice- and text-recognizing algorithms, and natural language processing). We suggest that modern software algorithms are adequate to perform many of the sub-tasks required for the rhetorical analyses of image memes, and that crowd-sourced annotations (via participatory research, or micro-task platforms) could be used to support algorithms where the software alone are as yet insufficient. Already in the case of digital discursive ecosystems today, some fraction of users contribute their time and energy to improving discourse, for example by providing context or reporting behavioral violations. Approaches for online platforms that combine gamified participation with behind-the-scenes machine learning have been successful in advancing research in biochemistry and a variety of other fields. These crowd sourced projects can take a variety of forms, and can be designed to operate directly on the engaging digital platforms that people already use [95].

Here we present what a case-by-case rhetorical analysis of image memes might look like, within a framework that is ultimately designed to scale up to high-throughput ecosystemic annotation, while retaining the semantic richness afforded by case-by-case rhetorical analysis. These analyses are performed in three phases:

**Phase 1. Entity Identification.** The first phase of analyzing the rhetorical function of a meme entails recognizing visual entities embedded in the meme. Entities can be of different types and are interchangeable across memes.

**Phase 2. Rhetorical Analysis.** The second phase of decoding the function of a meme entails identifying its semantic and consequently persuasive potential. This phase begins with tracing relationships between the entities implied by their arrangement within the meme. The relationships will typically synthesize into an implied (or stated, if the meme includes text) claim, sometimes accompanied by evidence included in the meme. The claim often rests on implied warrants (assumptions) supplied by the viewer who is aware of the rhetorical context that the meme invokes.

**Phase 3. Hidden State Identification.** The third phase of decoding the function of a meme is hidden state identification. The exact nature of the hidden state inference will be situational and depend on what the analyst is attempting to reduce their uncertainty about; for example, the extent to which the image meme in context is consistent with social values, providing specific valuable services, or eliciting violence. What distinguishes the various possible hidden state inferences from rhetorical inferences in Phase 2, is that hidden states are deeper than specific claims about entities, and reflect underlying attributes of the rhetorical ecosystem that gives rise to and are strengthened by such claims.

Two examples below (Figure 8 and Figure 9) represent the qualitative application of the DRE3 model to shareable image memes. The rhetorical analyses below uncover preferred readings of these image memes [96] and are not exhaustive in terms of entity or claim identification. Memes, as identified earlier, are polysemic. They are able to generate multiple and varied interpretations. A rhetorical analysis cannot comprehensively decode all meaning possibilities embedded in an image meme. Nevertheless, by following the rhetorical use of symbolic content within the meme, attending to the discursive contexts in which a meme may be harvested (such as a Facebook post thread or a Twitter thread), crowdsourcing the claims advanced by memes, and determining interpretation consensus across trained rhetorical analysts, we can identify likely, core, or agreed-upon, in other words the preferred arguments that memes advance [96]. In this case, we define preference by what a meme was originally designed to argue or the meanings that are most easily accessible (obvious) to the target audience. Even though the meaning of a meme can be altered by its discursive context (i.e., a meme can be deployed ironically to undermine its own message), such a subversive reading of the meme relies on consensus about its dominant meaning. Therefore, despite inherent polysemy, we believe it is both possible and useful to identify the dominant argument(s) that are encoded in an image meme.

## Example I

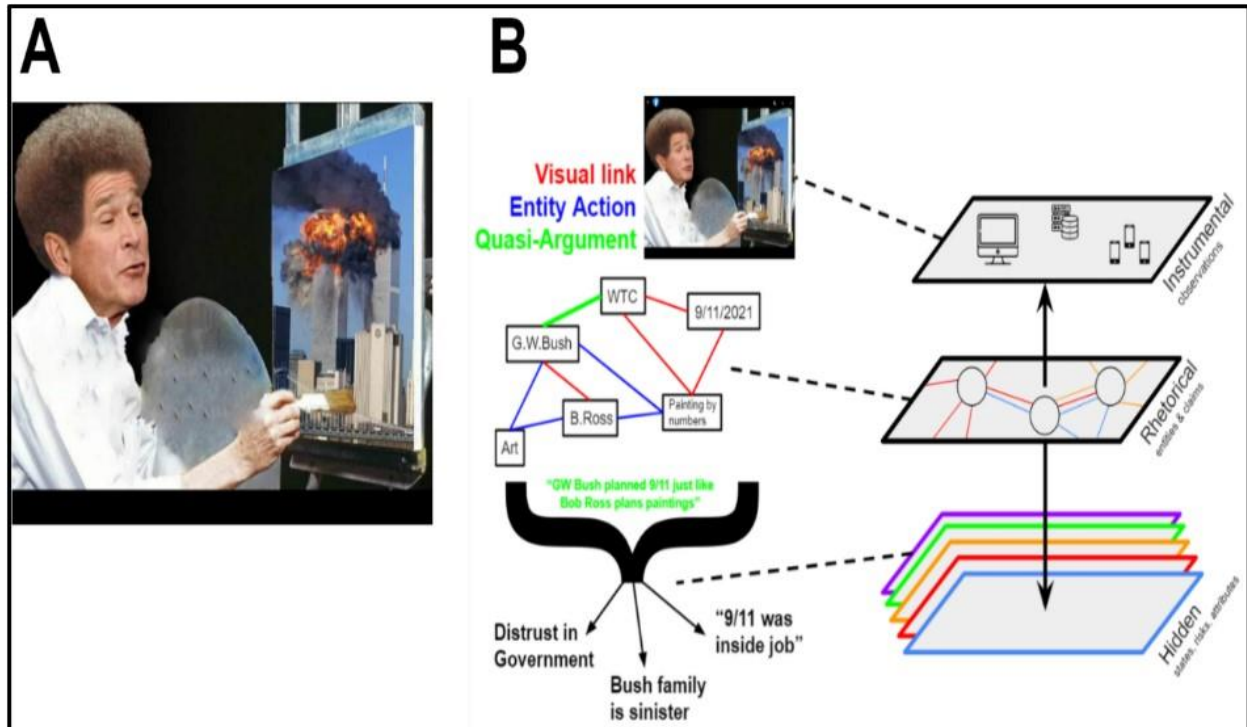


Figure 8. Illustration of the DRE3 model as applied to an image meme without text. A) a target image meme under analysis. B) Application of DRE3 model, breaking down the meme in terms of the Instrumental tier (what was observed), the Rhetorical tier (entities, warrants, claims), and the Hidden State tier (implications and use-specific inferences).

### Phase I. Entity Identification

In the above meme, the following entity categories are rhetorically significant:

**Persons:** Bob Ross, G.W. Bush

**Attributes:** Hair, shirt, hand of Bob Ross, Face of G.W. Bush

**Objects:** Twin Towers of the World Trade Center, Painting materials (palette, paintbrush, canvas, easel)



**Location:** New York City skyline

**Action/Relationship:** Individual painting on canvas

### **Phase II. Rhetorical Analysis**

In the above example, decoding the meme rhetorically by analyzing relationships between entities requires distinction between host images and parasitic images. The incorporation of the parasitic images to replace parts of the host images produces a parodic relationship between host and parasite entities. The insertion of G.W. Bush's face into the identifiable hair of the artist Bob Ross parodies the parasitic entity—Bush. The host image is the one that dominates the meme. An enculturated viewer recognizes the image as a still from the iconic Bob Ross televised painting class. Ross's hair, shirt, hand, palette, brush, and canvas on the easel are easily recognizable attributes/objects and constitute the majority of the image. The viewer is clear that it is G.W. Bush's face that is intruding within the Bob Ross image rather than reading the artist entity as the intruder. Having identified the host-parasite relationship, the viewer must now extract the semantic implications of this parody.

In deciding what the host-parasite parody means, the viewer recognizes that the visual juxtapositions in the meme are meant to paint former president G.W. Bush as an artist. The parasitic image that has taken over Ross's typical placid landscape scene on the canvas provides a stark contrast to what those familiar with Ross expect him to paint. The peaceful landscape of a Ross painting is replaced by a real scene of terror (the fall of the Twin Towers on 9/11) that is also highly recognizable because it has become widely circulated memetic content.

The face of G.W. Bush and the destruction of the World Trade Center towers in New York City are clearly linked in the rhetorical context available to the enculturated and interpellated viewer. The structuring of entities within the meme, however, superimposes an additional relationship that emerges out of the parodic analogy between G.W. Bush and Bob Ross. The parody is underscored with the use of an exaggerated expression on the face of G.W. Bush. This is the point at which

the viewer arrives at the claim embedded in the image structure of the meme. The claim could be articulated as follows: Like Bob Ross paints a landscape from imagination, G.W. Bush fabricated the 9/11 terror attacks. In this case, the memetic argument advances only a claim. The meme contains no evidence. Instead, the meme operates intertextually. To unpack the meme's claim, the viewer must be aware of multiple rhetorical contexts, such as the 9/11 truther movement that has sought to expose the terrorist attacks of 9/11 as a plan of the United States' own government, and the imputed role of the Bush family within the construct of a global cabal that controls worldwide events. In this way, the rhetorical analysis of memes leads us to identifying salient hidden states (e.g., social, political, and cultural beliefs/practices) that both influence and are shaped by memetic arguments.

### **Phase III. Hidden State Identification**

A rhetorical decoding of the Bob Ross-G.W. Bush meme both relies on and perpetuates claims about the Bush family, the G.W. Bush administration, the events of 9/11 and other global destructive events. Memetic argumentation analysis is ultimately useful to the extent to which it permits tracing evolving public beliefs and practices that could have real-world implications. We expect that, over time, the identification of rhetorical claims from varied memes will reveal patterns of connected beliefs that correspond to higher-order hidden states such as confidence in the government, or beliefs about the causes of past events. A hidden state in our framework refers to an implicit and volatile state of public belief, sentiment, or action. A belief that the United States government lies to its people is an example of a hidden state. This higher-order claim represents a public belief that produces a sentiment of distrust in the government. Tracing hidden state dynamics is useful because they can activate overt action in unrelated contexts, such as vaccine refusal because of a previously established distrust in government. Such a relationship between hidden states and public action can potentially be identified by tracing co-occurrence of memetic claims within networks.

## Example II

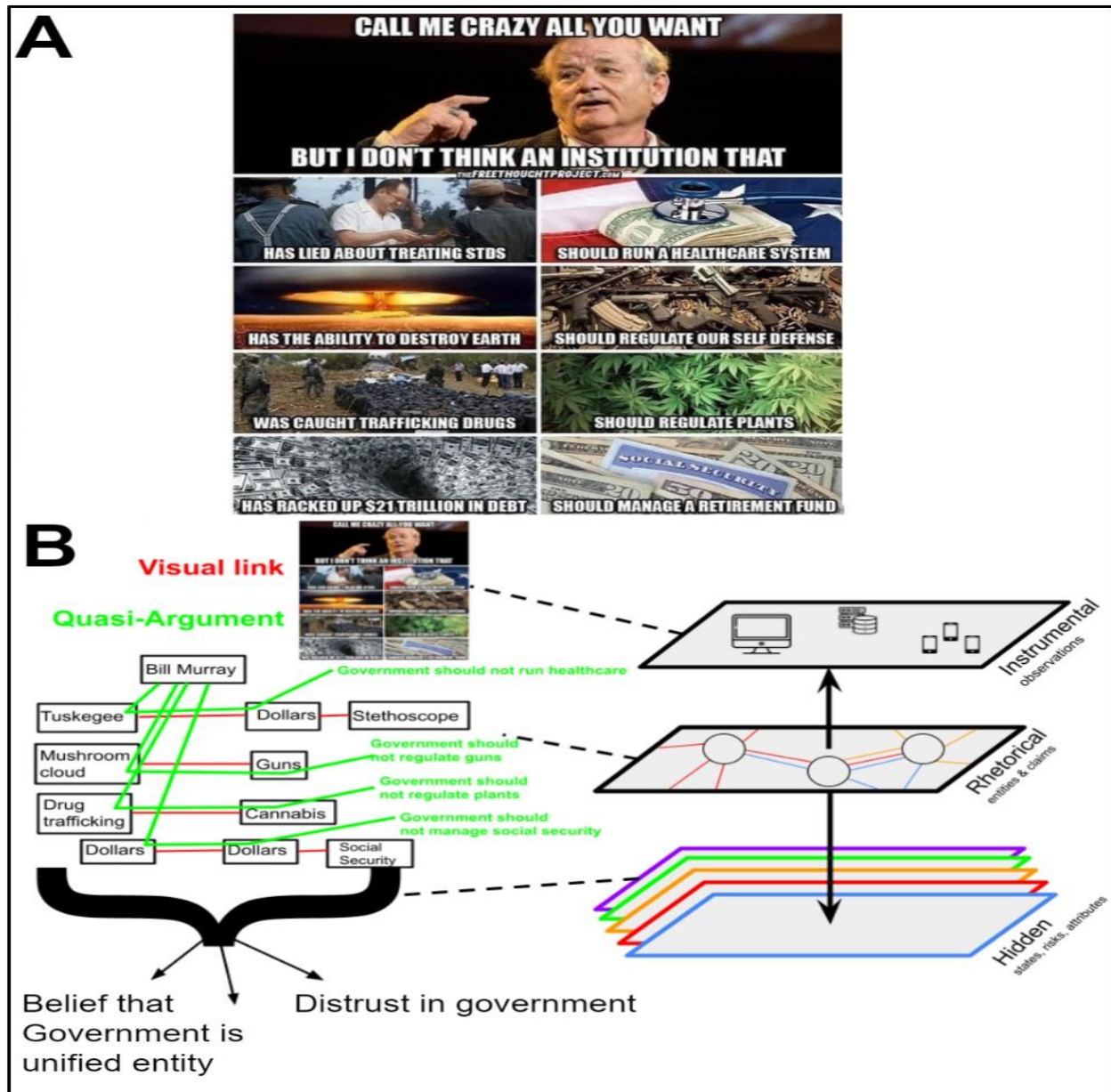


Figure 9. Example of the DRE3 model as applied to an image meme with text. A) a target image meme under analysis. B) Application of DRE3 model, breaking down the meme in terms of the Instrumental tier (what was observed), the Rhetorical tier (entities, warrants, claims), and the Hidden State tier (implications and use-specific inferences).

In this example, the higher-order claim that the United States government cannot be trusted is advanced by submitting lower-order arguments. The text-image pairing in this meme enacts argumentation differently than in Example 1. While the first example illustrates argument by analogy, this example supports its claims with visual evidence and follows an “if-then” pattern.

### Phase I. Entity Identification

In the above meme, the following entity categories are rhetorically significant:

**Persons:** Actor Bill Murray

**Scenes:** Tuskegee syphilis study, mushroom cloud, drug heist.

**Objects:** Dollar bills with a stethoscope, stock of guns, marijuana plants, vortex of dollar bills, dollar bills with social security card.

### Phase II. Rhetorical Analysis

The visual segmentation of the meme-box is crucial to how the argument is enacted. The visual sequencing relies on the viewer moving from the top to the bottom and from the left to the right. The top centered image features the actor Bill Murray. The text superimposed on this image invites the viewer into a dare with the person sharing the meme. The challenge “Call me crazy all you want” invokes the trope of the conspiracy theorist, a label typically branded on those who accuse the government of large-scale wrongdoing. The rest of the meme-box is set up to enact that challenge and rebut the conspiracy theorist label. Bill Murray, known for his antics that speak truth to power, functions as a symbol of interpellation for the conspiracy-minded, who are not taken seriously by the mainstream but are convinced of the truth to which they have awoken.

The lower order arguments are presented in claim-evidence pairs, each contained in smaller boxes in the left-hand column of the meme. Four claims about government malevolence are substantiated with images meant to provide evidence.

The first claim accuses the U.S. government of lying about medical treatments of STDs. The image over which the textual claim is superimposed features African Americans, a visual sign

meant to invoke the Tuskegee syphilis study that abused black American bodies in a deceptive government intervention. The image in fact is an iconic historical photograph of the study. But, even in the absence of audience knowledge about the provenance of the photograph, knowledge about the Tuskegee study itself is enough to decode the image as representing that particular instance of government dishonesty and failure.

The second claim accuses the government of the ability to destroy the planet and is substantiated with the paired image of a mushroom cloud that invokes the Hiroshima atomic bomb disaster.

The third claim accuses the government of trafficking in drugs. The textual claim is superimposed on an image meant to invoke the plane crash that exposed alleged CIA involvement in drug trafficking in Panama.

The fourth box in the left-hand column claims that the U.S. government has \$21 trillion in debt. Here the paired image simply shows a giant vortex of dollar bills illustrating the metaphor of “money down the drain”. The preceding images which pull from historical archives construct the credibility of the meme, leading the viewer to implicitly assume the facticity of the final allegation, even though the fourth argument departs from the claim-visual evidence pattern established by the previous three.

The visual segmentation and sequencing in the meme optimizes the constrained space of the meme-box to deliver a relatively complex argument with multiple claims and pieces of evidence. Each text-image pairing on the left works in conjunction with the text-image pairing on the right to both verbally and visually enact the if-then argument pattern. The boxes on the left provide evidence for the claims on the right. For example, the government’s dishonesty in the Tuskegee study is presented as evidence for the claim that a nationalized health care system cannot be trusted because of the ways in which it might abuse unsuspecting citizens. Likewise, its willingness to bring the planet to the brink of destruction by deploying nuclear weapons is provided as evidence that the government should not be allowed to regulate gun ownership. The strategic use of the meme-box to bound the argument is especially stark in this

sequence. While evidence of the government's disregard for human life can be leveraged to support curtailing the government's military power, the corresponding claim instead attacks gun regulation, implying that citizens need to be prepared to defend themselves against an untrustworthy government. However, the implication that guns are powerless in the face of nuclear destruction, which would undermine the meme argument, is suppressed by the visual alignment of evidence and claim side-by-side. This visual formatting contained within the meme box constrains the possibility of additional lines of reasoning even more powerfully than a similar argument made through other forms, such as orally in a speech or verbally in a news article. The visual demarcation of the meme box has the powerful potential to restrict reasoning to the elements displayed within the box. Because of how distinctly recognizable the meme-box has become and how unique it is in appearance compared to other visual modes of public discourse, the meme-box is able to separate itself from the rest of the landscape of public argumentation and create both discrete instances of argument unique to its own content and structure, as well as to interact within the ecosystem of related memetic arguments.

### **Phase III. Hidden State Identification**

The four boxes on the left in alignment with each of their counterparts on the right together advance the higher-order claims that the U.S. government is dangerous, unethical, and inept and its interventions should be substantially curtailed. This claim reifies the hidden-state sentiment of distrust in the government. It is important to note, also, how the argumentation enacted by the meme relies on some but not extensive contextual knowledge in the viewer. The parsimony of the symbols within the meme (restricted to a few words and images) relies on the audience's background knowledge and ability to supply warrants. For example, audience knowledge about the Tuskegee study and its targeting of African Americans is essential to reading the first image on the left-hand side as evidence for its paired textual claim. However, even minimal recognition of some elements is sufficient for the viewer to then accept the other image text pairings and submit to the lines of reasoning traced by the memetic elements. Likewise, the meme relies on an interpellated audience to supply the necessary

assumptions (warrants) to complete the arguments. For example, the leap from the government's moral failing in the Hiroshima bombing does not automatically lead to an argument against gun regulations, unless the viewer is already concerned about the erosion of Second Amendment rights and is thus primed to read the atomic bomb image as evidence that the government does not have its citizens' best interests at heart and would therefore regulate gun ownership to reduce the threat of self-defense from its citizens.

## Concluding Comments

The two examples elaborated above show the kinds of information about memetic claims and hidden states that can be inferred with a rhetorical approach. In the following section we integrate the insights from rhetoric and ecology to outline some considerations for the design of online discourse monitoring systems.



## Part V

# Toward High-Throughput Rhetorical Analysis (Meme SCADA)

The example applications of the DRE3 model in the prior section show the kinds of information about hidden underlying states inferable with a rhetorical approach, that are impossible using syntax-driven analysis such as keyword extraction or entity recognition alone. Digital discourse moves at a very fast pace. Rapid changes in digital discourse (e.g., during an unfolding political event) are likely the times when monitoring and analysis are most needed. Unfortunately, the DRE3 model, as applied above, is low-throughput. This problem is not unsolvable. The field of ecology offers a hopeful precedent, because it emerged from low-throughput observation of natural history, and later increased in scope and rigor through the application of quantitative frameworks and large-scale monitoring networks. We propose that rhetorical ecosystem analysis is capable of making a similar transition to a higher through-put research phase, in the case of some digital artifacts.

The value of developing capabilities for cataloging, indexing, searching, mapping, monitoring, and modeling digital discourse is also not limited to facilitating research. Just as better ecological understanding and monitoring has enabled forecasts, such as those related to algal blooms, disease, wildfires, and the potential risks of construction or development [97], better understanding and monitoring of digital discourse could forecast outbreaks of violence, acceptance of government initiatives, the spread of ideology, and the potential risks involved in narrative influence [98]. A wide variety of disciplines undoubtedly have interest in tools for modeling, mapping, and monitoring digital discourse, such as public relations, public health policy, and military information support operations (MISO) [98]. Many high reliability organizations, or organizations which must maintain low-failure rates or risk cascading failure [50], have expressed or demonstrated a need for tools which perform these functions [99–103]. While recent crisis events have drawn particular interest to the potential application of these functions in monitoring and modeling digital discourse about public health and political extremism, there has been a long-standing need for these functions in areas which are entirely apolitical, such as of multimodal content regarding interpretations of emergency situations like forest fires, floods, and earthquakes [104].

Transitioning from low-throughput to high-throughput, and from theory and research to forecasting and decision-making support, will only be accomplished by considering the related requirements of the outputs, of the processes and methods which lead to them, and of the tools and infrastructure which enable them. Here we explore and frame these requirements, consider methodology, and propose the structure of a monitoring system best categorized as a type of SCADA (Supervisory Control and Data Acquisition) system for digital discourse which incorporates the DRE3 model and modern computational techniques [105]. Addressing the use-case specific requirements of the many domains which might have interest in monitoring tools has been considered elsewhere [81]. Instead, the focus here will be on the requirements for more general sensemaking about public narratives generated by image memes.

## Narrative Intelligence

The general requirements for sensemaking common to all intelligent systems are the abilities to capture relevant data from the environment (sense), fit the data to expectations or adapt those expectations to fit the data (model), and use the expectations to consider or frame choices (policy) as a basis for informing action [87]. Various frameworks exist to convert these general requirements into formal processes and specific requirements for systems which facilitate sensemaking. These frameworks are often built for activities which require special consideration beyond the fundamental sense-model-policy framework, such as in militaries [106–108], teams [107–109], intimate relationships [110], machines and AI [111,112], and businesses [113]. Of the many sensemaking frameworks available, intelligence production may be the most appropriate for sensemaking related to digital discourse.

Intelligence production is an organizational sensemaking process which is intended to produce deliverables to inform policy that achieves or maintains the interests of an organization [114,115]. Formal intelligence production processes are particularly helpful for organizations that are large enough to make the natural emergence of synthetic intelligence or macrocognition unlikely or illusory, and for organizations which are interacting with systems of interest that are sufficiently complex to prevent existing synthetic intelligence from being able to manage available sense data appropriately [109,114,116–118]. The process of intelligence production was originally semi-formalized by the Roman military [118] and has been iteratively developed throughout history in response to situations where conditions complicating macrocognition arose or became exacerbated [114,119–123].

Intelligence production is a helpful way to frame the requirements of sensemaking in digital domains given that intelligence production was formalized to face similar challenges, such as voluminous collections across myriad surfaces, multimodal data

[124,125], deception and intentional disruption of data collections (counterintelligence) [126], and difficulty of detecting, monitoring, and interpreting counterpublic membership and activity [50,127–129]. Since intelligence production is usually performed by high reliability organizations [50] and faces the aforementioned challenges, it has been iteratively developed over time to maintain reliability and cope with imperfect data and uncertainty. While various specifications exist for particular use-cases, such as in business and commercial intelligence [113], generally intelligence production is modeled using 5 distinct stages: 1) planning and direction (requirements setting), 2) collection, 3) processing and evaluation, 4) production and analysis, and 5) dissemination [113,125,130,131]. These 5 stages provide opportunities for separations of concern between categories of function and process, as well as between personnel and access to information [131,132] to limit the possibility of “having either the facts or the conclusions warped by the inevitable and even proper prejudices” of those involved [133]. However, it should be noted that the steps formalized in the intelligence production model are not necessarily implemented in discrete phases, and that even where separate steps are intended, they still occur in parallel with blurs between processes [134,135].

Ecological and rhetorical metaphors and methodologies may offer unique and valuable approaches to monitoring and analyzing digital discourse, but no metaphor is a perfect mapping [136]. Here we apply the intelligence production framework to facilitate practical considerations for “mapping the gap” between ecology- and rhetoric- inspired methodology and the needs of a meme analysis pipeline at each stage.

## Planning and Direction

The first step of the intelligence production cycle is planning and requirements setting. This stage entails considering what kinds of intelligence products are needed and in what time frame, and translating these needs into technical and personnel requirements, scope, and expectations for the following steps [130–132]. In the case of a meme analysis pipeline, we suggest that the relevant products be broken into 5 broad categories:

**Data Sets.** While raw datasets do not constitute a formal intelligence product, the data collected and used for projections and other features are nonetheless a product which should be made available both internally and externally, similar to the provision of Twitter’s streaming API (application programming interface) and “Firehose” [137,138]. These releases are essential for 3 primary reasons. First, the analysis pipeline should never be considered entirely complete; data used and produced by various features should be available for both quality testing and

use in the development of new features. Second, datasets of content with semantic annotations could be invaluable for the development and training of AI. Finally, the capability to release data used allows for reproducibility and transparency in the case that outputs are considered partisan or questionable.

**Research Intelligence.** Research intelligence refers to information that may provide context or support for other intelligence products or help in further analysis or sensemaking, such as wikis, or “fact books” which might provide details about content and communities of interest in the context of digital discourse [114,139], field guides for providing education on common patterns and processes [98], exploratory search features for analysts and researchers, and research products such as academic articles or white papers.

**Estimative Intelligence.** Estimative intelligence refers to information regarding uncertain phenomena, such as the likelihood of an object impacting a particular hidden state, though some definitions place a larger emphasis on projection [140–143]. In the monitoring of digital discourse, helpful estimative intelligence might include metrics and projections regarding the state, rate of change, and impact, of beliefs, communities, patterns of activity, or content, informed by methodologies from ecology and rhetoric.

**Warning intelligence.** Warning intelligence refers to information about anomalous phenomena or rapid or unexpected changes to system state [139,144,145]. In the monitoring of digital discourse, useful examples of warning intelligence would include the detection of anomalous activity, the emergence of what may be coordinated, aggressive, and strategic activity associated with untracked or tracked objects or communities, notifications about other organized activity such as the censorship of content on a platform, or the presence of harassment, threats, or explicitly illegal activity.

**Actionable Intelligence.** Actionable intelligence suffers from a lack of consistent usage or a consistent definition, but generally refers to information which needs to be addressed urgently and informs or enables actions that might be or need to be taken [146]. In the monitoring of digital discourse, actionable intelligence would help inform interventions such as the

removal of content, inform design of content or messaging based on current trends, and guide sensemaking by providing new routes to consider when presented with ambiguity or structurally complex information.

## Collection

The second step is broadly referred to as “collection”. This term is sometimes used to refer to the entirety of the intelligence production cycle [133,147]. However, in the context of the production cycle and its processes, it refers to the conversion of requirements set during planning and direction into tangible targeting, selection, and instrumentation choices in order to collect data [125,130,148]. At this stage, the focus is on the collection of “raw intelligence”, or unanalyzed information, in accordance with requirements—as such, it is sometimes referred to as collation [132]. In the past, organization of raw intelligence was fairly disorganized [118–120,134,149]. But with the increase in volume, and the need to collect multimodal data from myriad surfaces, came a need for specialization not just in analysis but in the collection of raw intelligence as well, resulting in various formal categories of tradecraft, or types of intelligence collection and annotation methodologies [130,150].

There are a series of ethical and practical challenges to the development of collection requirements and procedure for image memes in the interest of developing an image meme analysis pipeline. A root problem, worth addressing first, is that even at the cutting edge of machine learning applications in analyzing memes, there are serious limitations imposed by the lack of existing annotated collections to use as training data [23]. Thus, the use of AI at this time for automated collections would likely be inappropriate given that even the ability to differentiate between an “image meme” and “just an image” is a difficult, semantic challenge—let alone the ability to analyze it. However, given the rate of change, complexity, and volume of image memes, collection would place too high a burden on researchers, experts, and analysts. Crowd-sourcing may therefore be the best avenue of approach. While crowd-sourcing approaches have come under criticism, recent successes indicate that more complex tasks may now be ready to be outsourced to crowds [95]. Choices in incentivization mechanisms and user experience design would need to be considered in depth elsewhere, but there is a rich history of crowd-sourcing data in ecology which could be of use in framing collection requirements. For example, millions of entries for bird sightings are generated by citizen bird watchers each month [151] and data from bird sighting submissions can be used by analysts for real-time monitoring of animal activity as well as for forecasting phenomena such as outbreaks of West Nile virus [152]. The frameworks used for crowd-sourcing in ecology may allow for a direct transfer to other domains, such as those which

provide data management principles [153] and offer methods for improving information quality or “Crowd IQ” [154].

Among the approaches developed in ecology and ecology-adjacent fields from learning-by-doing in crowd-sourcing, three stand out as both valuable and immediately applicable. First, based on crowd-sourced classification of plants and birds, quality of collections can be greatly improved simply by improving the quality and scope of the class structures (schema) and data standards the crowd will interact with [154]. Second, the study of crowd-sourcing fish classifications and remote-sensing in hydro-ecology has shown that quality can be improved over time by segmenting users by expertise and using these segmentations to provide different levels of responsibility [155,156]. Third, work on crowd-sourcing biomedical annotations has revealed that expert contributions can be used to train and tune user contributions, particularly to detect anomalies and unexpected deviations from patterns. Similarly, user contributions can be used to train and tune automated systems and be assisted and guided by them in performing contribution tasks (see figure 10) [95]. These approaches could be directly applied to “field” collections of image memes. Given that collections are occurring online, most relevant information, such as where the object was collected, the object’s file type, and reaction or “impact” data if it was collected from social media, could be automatically fit to pre-existing data standards with no need for experts involved in collections before being placed in a buffer for classification. The collected objects could then be used to train AI to determine what and what does not constitute a meme.

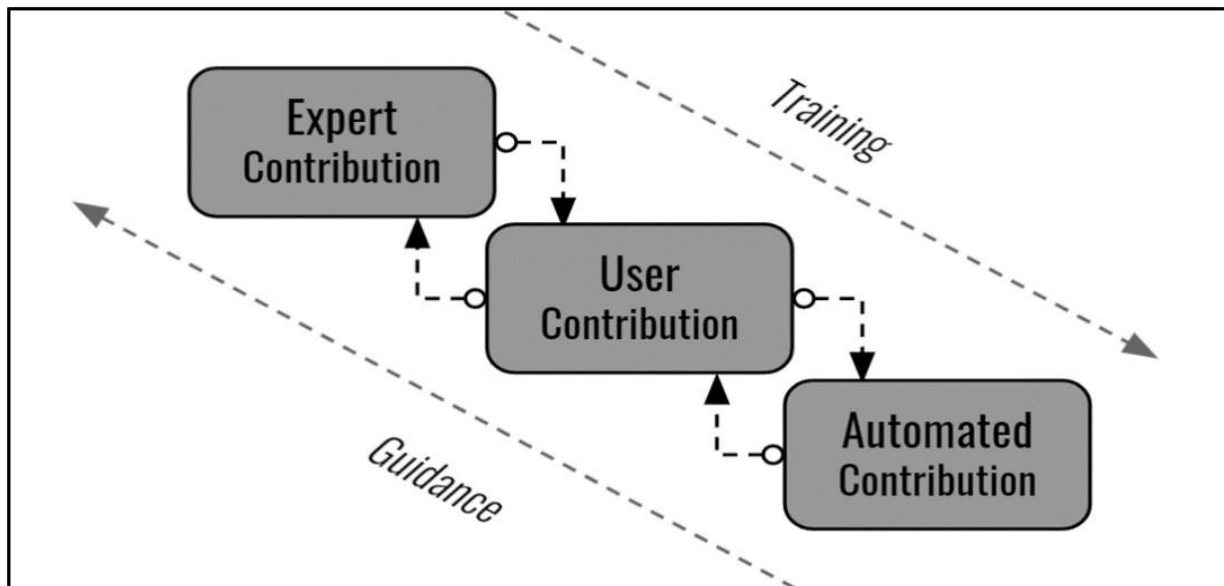


Figure 10. The flow of benefits offered between types of user contributions. Contributions by user segments with higher levels of competency in a task can be used as training data for those of a lower competency, while contributions from segments with lower levels of competency can be used to help provide guidance to those of a higher competency (e.g., suggested classifications).

While it might be reasonably assumed that data about the user who posted the collected object should be automatically parsed and collected as well, this may not be necessary. As noted elsewhere in this article, memes, particularly political memes, are often presented without attribution. Further, user data may need to be bypassed because creating or sharing political or even quasi-political memes or other content, especially within counter-publics where meme-activity is rich and of interest to researchers, is increasingly being accompanied by the expectation of potential consequences from peers [157], employers [158,159], and institutions [160,161], as well as by potential punitive consequences from media-sharing platforms [162–165] and governments [166,167]. The DRE3 model's focus on claims in memes informed by a rhetorical approach, and on relationships, placement, and change of that content informed by an ecological approach, as opposed to a focus on the identity of the poster, prevents misattribution or association inferred by posting history (e.g. a CDC official sharing an anti-vaccination meme for educational purposes), reduces the potential for harm by “outing” or “doxxing” internet users, especially in countries with higher potential for consequences for sharing political content, and reduces the potential for critical misuse of the analysis pipeline. For the purposes of understanding movement of memes specifically, the channel over which the meme travels is sufficient. If the collector of the meme in context with a particular platform constitutes a channel, then this channel can be considered a location—leaving no reason to deanonymize the collector and making the generation of an “identity” within the pipeline an opt-in exercise.

## Processing and Evaluation

The third step of the cycle is often referred to as processing and evaluation and refers to a pre-analysis stage in which data is cleaned, refined [148], and filtered [130] and the reliability and credibility of sources of the information are considered [132,134,168]. The raw intelligence assembled in the collection phase is now altered or reassembled for usability, “coded data is decrypted, foreign languages [are] translated, and photographic material [is] interpreted” [148]. The importance of processing and filtering cannot be overlooked. Without comparable measures, accessible reference information, or compression into usable formats, much of the data could essentially become meaningless [169]. When this processing is done in concert with proper scope and orientation introduced in the planning and direction phase, it also reduces the potential for endless abstraction by making the means and intentions of the process clear [87,170].

It is at this stage in an image meme analysis pipeline that experts would be needed to begin classifying objects and improving information quality as the pipeline begins to move beyond syntax and metadata toward semantic annotations. Even with the use of crowd-sourced and automated collections, the load would still be far too great for experts and trained analysts to handle alone. This being the case, the same



framework of training, guidance, and segmentation between the kinds of contributors described in the prior section would offer continued utility (see figure 10). Automated systems would be given responsibilities such as detecting quantitative features that are correlated with virality and longevity of the image meme, which can then be used to direct the attention of both experts and average users [23]. These systems would make use of data from the contributions of human users to train for more complex tasks. Expert users would have the primary responsibility of developing and detecting claim and argument patterns and applying these labels to content, which could then be used to train average users or even AI to do the same.

Claim identification presents the largest challenge to crowdsourcing the DRE3 model due to the subjectivity of the extraction whether it comes from rhetorical experts or average users. Image memes, as discussed in prior sections, tend to have an ambiguity which offers the poster plausible deniability about the assertion of claims. Therefore, simple automation of feature recognition cannot be relied on for extracting claims from images. However, this challenge may instead be seen as an opportunity. There are many viable methods for extracting and aggregating arguments from text [171–173], allowing for the substance of these common arguments in various phrasings to be aggregated and clustered. The remaining disparity between interpretations would not, and should not, be considered noise—but instead valuable data for producing metrics related to the subjectivity and complexity of the content and of diverse perspectives interpreting it. Average users would share responsibility for claim extraction, though their primary responsibility would be the extraction of relevant entities from the content.

Once experts have provided sufficient labeling of rhetorical pattern and structure, average users could be slowly trained. Segments of those users may even eventually be trusted with contributing rhetorical or other expert classifications, though the provision of greater responsibilities would likely require new tools or frameworks for managing trust in crowd-sourcing systems. Automated features however, would likely need to stay in a guidance role regarding most semantic analysis of image memes. Semantics on the internet are prone to rapid change and often require contextual knowledge. For example, triple brackets around an organization or person's name is now often considered an antisemitic symbol marking Jewish background or influence. But obviously, not all uses of triple brackets indicate this—and worse, prior to this association, the same triple brackets were used to indicate a “cyberhug”. This does not mean that automated features would be useless. For example, the ability to note that some typographical feature may mean something to specific audiences and to direct a user's attention to that symbol is a valuable guidance feature.

## Production and Analysis

The fourth stage of the cycle is referred to as production and analysis, wherein experts begin to produce the intelligence products requested, given the collected, processed, and evaluated information available and relevant to them [148]. At this stage in a meme analysis pipeline, data and content cataloged throughout the collection and processing stages can now be structured into data sets for developing, improving, and replicating automated features at all stages in the pipeline and for more specific exploratory analysis by experts. More importantly, it is also at this stage in the meme-analysis pipeline that rhetorical and ecological framing and techniques begin to have their most valuable contributions.

**Research Intelligence.** The content labels, entity extractions, and identified claims informed by rhetoric now have a role in enabling semantics-driven exploratory search. The bottom-up detection of patterns and topological motifs allow analysts to view single pieces of content as a part of memetic clusters, not just of other pieces of content, but of entities, claims, and subclaims expressed in that content, and of the hidden states that may be signaled by them. With the metrics and features which accompany the objects labeled within these memetic clusters, the analyst is able to monitor a semantic field, or rhetorical ecosystem, as described in previous sections, before analysis has even been performed. The data is now available to enable methods of analysis from ecology discussed elsewhere in this document. In addition, the content, patterns, and aforementioned ecological motifs can now be structured into coherent and navigable wikis, field guides, and fact books, modeled after the large, robust identification systems and guides found in ecology—helping improve methods and standards at all stages of the pipeline and increasing the likelihood of novel genres or features being detected.

**Estimative Intelligence.** The use of ecological frameworks and methods for identifying and communicating state features of content and claims, and considering the relationships between entities, memes, and claims, as discussed previously, could be of great value. The ability to classify and quantify state features implies the ability to consider potential for impact and spread, as well as the ability to measure rate. The provision of data regarding these changes to content and claims and related rates of changes may allow analysts to not only communicate current state, but also project future state of both claims and

associated hidden states. This information can be leveraged in order to generate reports regarding underlying ecosystem hidden state features and their potential for change.

**Warning Intelligence.** The ability to classify and quantify state features, and project future states, further implies the ability to use those projections in the production of warning intelligence or general alerts. First, with the presence of patterns of spread, rhetorical structure, and state changes, comes the ability to detect breaks from expected patterns, or anomaly signaling. These anomalies can be prioritized and reviewed in ex post analysis to reveal and catalog new patterns, allowing for indications of phenomena which urgently require attention, such as swarm-behavior in political happenings, communications, harassment, censorship events, or organized activity. In addition, the ability to simply index content paired with the ability to classify and quantify state features means an ability to tag or “track” content. Ecology already has robust methods for the tagging of animals, some of which are used to enable early warning and risk alert systems. Similar methods could help inform the translation of changes to state into relevant notifications and warnings [174].

**Actionable Intelligence.** State features and context provided by hidden state analysis could generate intelligence products to improve decision-making around digital discourse in a number of ways. First, design and timing of content could be informed by the hidden states behind the claims dominating the environments they are intended to be deployed in. Second, if certain activities presented in warning intelligence require action, state features and hidden states can inform interventions. Finally, organizations whose decisions are meant to be informed by the interests of their constituencies can learn, through the tracking of claims, what those interests are, to increase the relevance of, and avoid negative externalities in, content deployments.

## Dissemination

The final step of the cycle is the dissemination of intelligence products to stakeholders and decision-makers [102,104,113,119] and integration of those products into existing knowledge-bases for future use [96,119]. The various categories of individuals who would receive these intelligence products are often

broadly referred to as “consumers” or “users” [104]. These intelligence products are traditionally written or oral reports intended to be periodically disseminated [148]. However, an insight which may be gleaned from ecological and ecology-adjacent forecasting is that when threats tend to be fast-moving or ongoing, and cannot be solved, only managed, intelligence needs to be consistently available, updated in real-time, and automatically disseminated and tailored based on expected need or upon request [59,175]. While the release of both periodical and non-periodical publications, newsletters, and briefings would be of value, they could not be relied on as the only method of dissemination to stakeholders.

In addition to these static disseminations, intelligence products would have to be tailored and presented in several ways. First and foremost, would be automated and other on-demand reports, that could be made available when requested, on particular claims, clusters, or other queryable objects. The ability to have dissemination via notification would be significant as well, given that warning intelligence is, by its nature, emergent and non-periodic, and is therefore in need of a channel over which it can be provided to those to whom it would be most relevant. Further, who may need this warning intelligence can change greatly with context. For example, warning intelligence regarding purported foreign influence of memetic content would only become relevant to some users of pipeline outputs upon their viewing of that content. Thus, intelligence would also have to be made available upon encounter. On-encounter dissemination could also be useful in terms of actionable intelligence, to help facilitate interventions, or, in terms of estimative intelligence and research intelligence, to allow analysts to use the content in front of them to direct the exploratory search of the existing corpus in developing new intelligence products, or to allow contributors during the processing and evaluation phase to better understand how to perform classification. Finally, given the rate of change in digital discourse, the ability to watch intelligence update in real time becomes essential. This type of real-time analysis of large volumes of digital discourse would be useful for a range of individuals, for example, public health officials observing the dynamics of public sentiment and impact of government messaging [81].

## Toward a Meme SCADA

With these requirements in mind, there is one approach in particular which presents the affordances and flexibility necessary to handle all of the challenges posed by the production cycle discussed above: the use of dashboard-based SCADA (Supervisory, Control, and Data Acquisition) systems. SCADAs are used to supervise state, acquire data from remote sensors, and control operations in real time [176]. While SCADA systems were traditionally intended for use in industrial operations, approaches from this area of research and application have recently gained traction in ecology [177,178]. Framing image meme analysis pipeline as part of a SCADA infrastructure

is potentially the most practical approach for two primary reasons. First, SCADA infrastructure is built with real-time use in mind and designed to facilitate the production of dashboard-like presentations of multimodal data and hidden states which are often difficult to communicate. Second, SCADA infrastructure design methodologies assume the need to collect and aggregate data from myriad sensors, and help inform information fusion protocols needed to generate forecasts, estimates, and current state features in real-time. In the case of the meme-analysis pipeline, supervisory and data acquisition features would be most prominent, though control features might be expressed in the form of prioritizations for users performing classifications and collections (such as during political happenings or swarm-behavior events), and in the form of explicit direction of automated collections and classifications. Here we present the rough blueprint of a meme analysis pipeline built in the style of an ecological or industrial SCADA system, from the requirements and outputs discussed within the previous section (see Figures 11 and 12).

Figure 11 shows the process by which artifacts (image memes) are collected, processed, analyzed, and disseminated. It begins with automated and manual collections of artifacts being given standardized annotations related to the location, structure (data type), and impact of the item. Next, these yet-to-be-processed artifacts are placed into a buffer; experts, average users, and automated features select artifacts from this buffer to identify their (i) statistically or quantitatively derived attributes and classifications, (ii) featured entities, (iii) claims, and rhetorical structure. The artifacts are annotated with these classifications using rhetorical and format annotation standards before being placed into an indexed and queryable catalog. Automated features and experts can draw from this catalog to perform analyses offered through a dashboard system for dissemination and monitoring. In addition, developers could use the catalog for training and test data in the development of new automated features. Finally, experts can make requests through the dashboard for prioritizations on manual collections and could direct the prioritization of automated collections (e.g., on certain kinds of content or from specific communities). Figure 12 shows the various forms of analysis and products which should be made available both through the dashboard and otherwise.

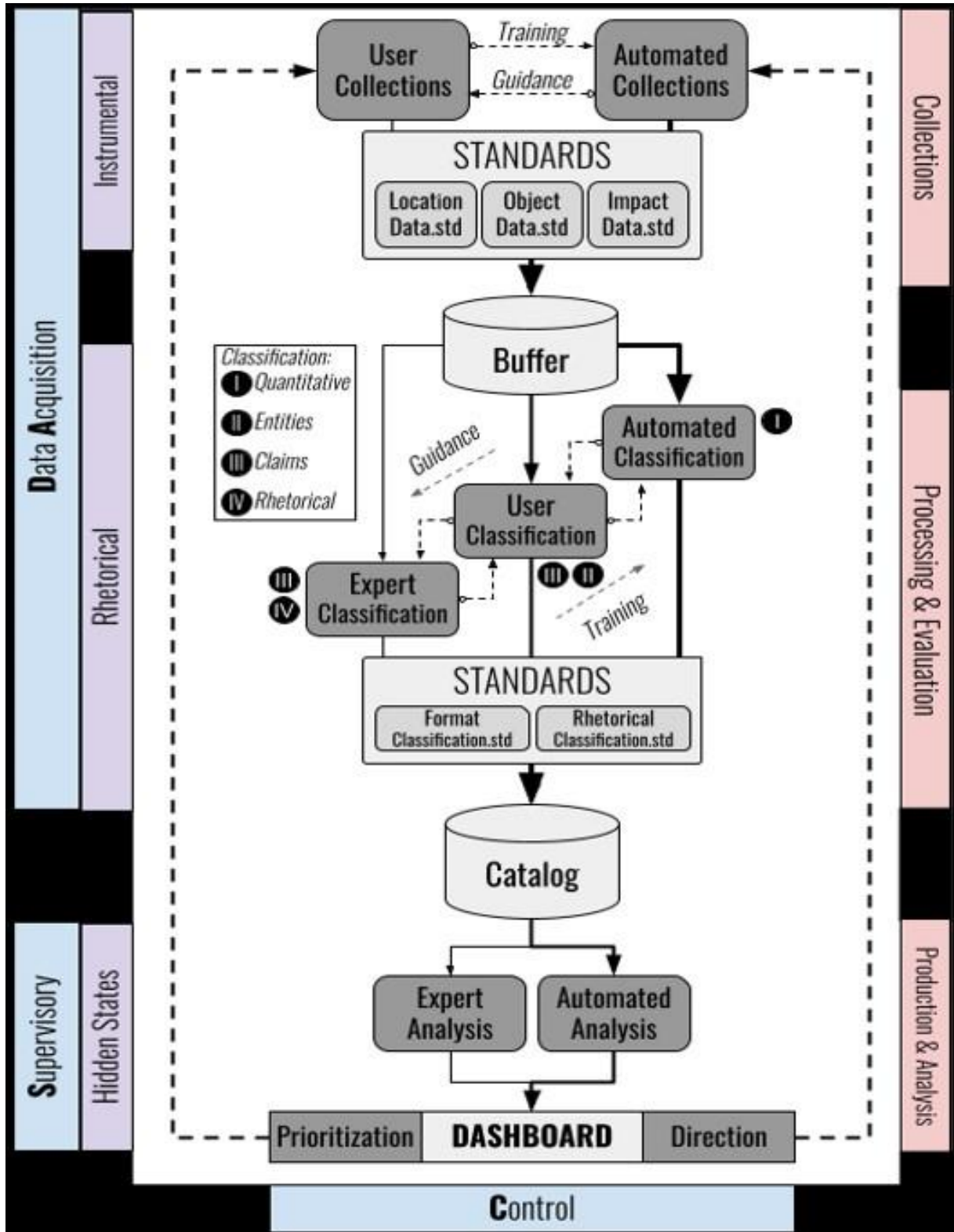


Figure 11. A rough blueprint of a meme-analysis pipeline. Color is used to indicate areas of the pipeline related to specific aspects of SCADA systems (blue), DRE3 analysis layers (purple), and intelligence analysis stages (red). The blueprint shows the various steps of content collection, processing, and analysis leading to the management of final intelligence products within a dashboard.



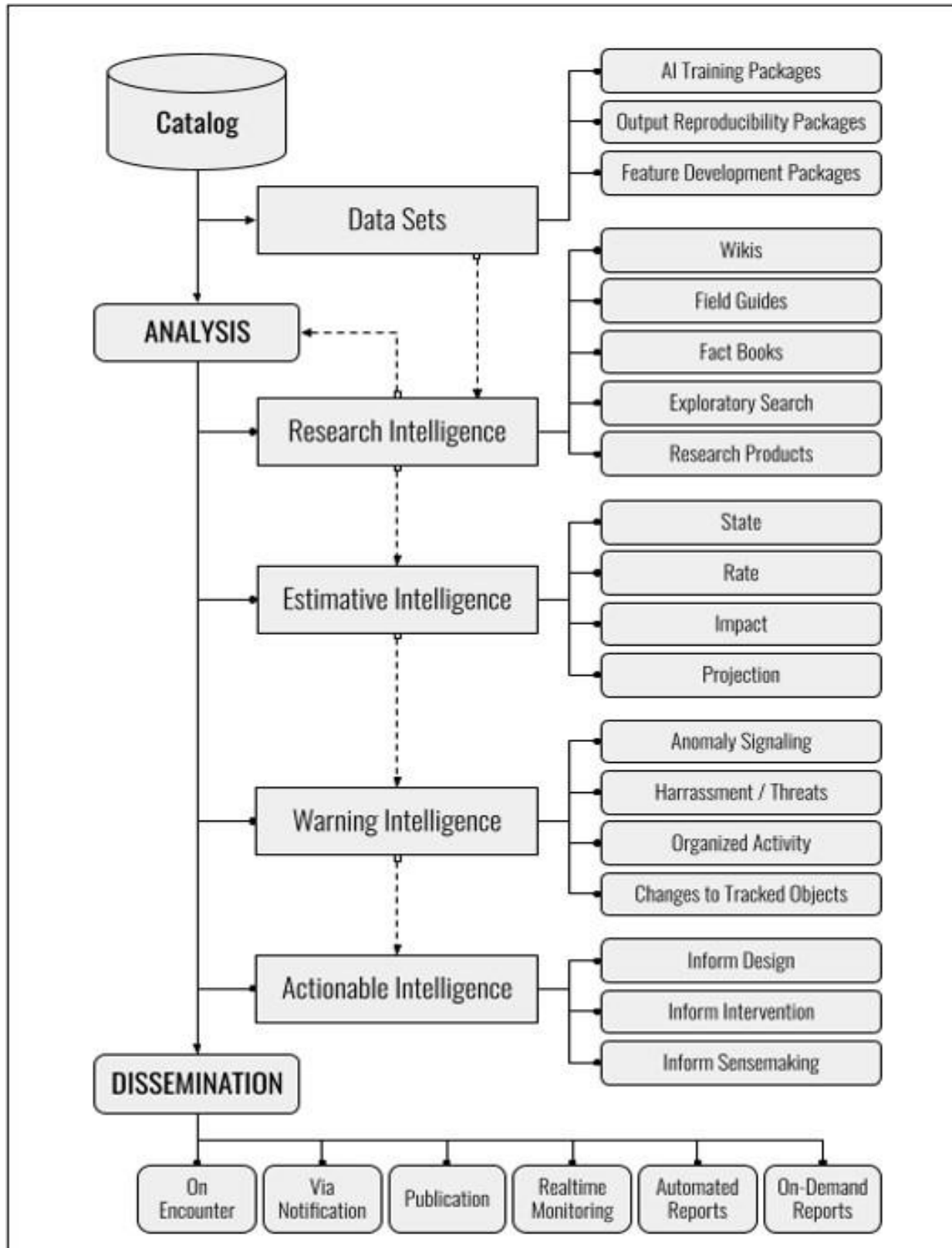


Figure 12. A map of desired outputs from a meme-analysis pipeline.



# Discussion

In this paper, we have reviewed the relevance of rhetorical and ecological approaches for analyzing multimedia digital discourse, such as shareable image memes. While rhetorical analysis captures the nuanced relationships between artifacts and audiences, ecological analysis captures the complex relationships among organisms and their niche. Others have explored similarities between the fields of ecology and rhetoric [37,179]. We have elaborated this connection through three key themes from modern ecology: the multilevel systems perspective, antifragility, and ecosystem services. These key themes integrated into the Digital Rhetorical Ecosystem three-tier (DRE3) model, providing a framework for incorporating rhetoric into computational pipelines for analyzing digital discourse, with ecological toolkits and frameworks as intermediaries. In addition to the transfer of concepts used in ecology into the digital discourse space and specific implications for SCADA design, here we conclude by exploring some broader implications.

We go so far as to hypothesize that a disruption or correction of narratives forged through memetic circulation needs to adopt the memetic form itself, sometimes known as a counter-meme [180]. We advocate re-deploying the memetic form to interrupt the credibility of a specific meme argument by illustrating why the claim advanced by the original meme does not rest soundly on the evidence or the warrants (assumptions) signaled explicitly or implicitly within the meme. Current efforts to fact-check memes address memes with a different genre of rebuttal discourse (e.g., the Facebook fact-check box that often links to news articles of official credibility). Digital audiences that have become vulnerable to the influence of memetic argument have also grown a staunch resistance to this particular form of fact-checking. Therefore, we argue that any attempt to neutralize memetically constructed narratives needs to understand the rhetorical power encoded within the memetic form and to use that form strategically to restructure public discourse. We urge, however, that counter-memetic efforts acknowledge the conditions of cognitive complexity endemic to digital knowledge environments and avoid the pitfalls of easy fact/fiction dichotomy for issues that are murky, complex, or ambiguous. Counter-memetic strategy should expose how memes mistakenly create narratives of certainty in the face of situational ambiguity and complexity. That is, counter memes should avoid making new issue-based arguments themselves, and instead reveal the argument weaknesses in memes deployed to advance public argument. Simply put, memes can be used to demonstrate the argument weaknesses of memes. The repeated circulation of rebuttal memes to demonstrate the inferiority of memetic argument has potential to eventually decelerate reliance on the memetic form in public discourse. In addition, asking users to identify claims embedded within image memes during the stage of data processing and evaluation

(Figure 11), could induce a more critical or meta-cognitive engagement with the memetic content and its deficits.

Rhetorical analysis has traditionally focused on single cases. Advances in computational technology provide the possibility of scaling up rhetorical analysis, for at least certain kinds of artifacts, such as image memes. Such high-throughput automated possibilities are evident in AI software such as Project Debater [181] and SwarmCheck [182] which can make sense of voluminous amounts of argument data using argumentation principles. The integration of rhetorical analysis with ecosystem tracking into a SCADA can enrich the field of rhetorical study by growing data-driven rhetorical theory. In 1969, Chaim Perelman and Lucy Olbrechts Tyteca published the influential *New Rhetoric*—a comprehensive compendium of argument strategies that relied not on formal logic but on everyday rhetorical practices [183]. Their catalog was built upon meticulous collection and analysis of real specimens of persuasion. Likewise, with the building of the proposed SCADA, we have the possibility of identifying and cataloging argument patterns across large amounts of image meme data, in a partially-automated fashion. The incidental value to argumentation theory of tracking the emergence, interaction, proliferation, and demise of image memes through discursive ecosystems is significant. We can determine whether argument patterns in image memes replicate documented argument patterns or assemble new ones. We can assess whether the unique genre of the image meme privileges certain argument patterns over others. An over-reliance on certain argument patterns (like argument by exposing hypocrisy [17]) may signal epistemic trends that are being exploited in the digital public sphere because they make minimal attention demands. When audiences are conditioned to argue in certain ways, their receptivity to other argument patterns that demand more central processing may diminish. We may observe at scale, with the intelligence that emerges from the SCADA, that one significant answer to the epistemic crisis we are currently battling is to understand the problem not just through a content framework (e.g., the fake news-real news dichotomy) but rather to problematize the medium, in this case the rhetorical form of the image meme, as one of the primary drivers of the crisis.

Another way to address the crisis is by examining ethical frameworks for managing a resource commons. In ecological philosophy, the “land ethic” [184] captures a sense of duty and responsibility towards ecosystem interactions. In the eponymous book, Aldo Leopold contrasted the land ethic with alternative frameworks that might be used to guide decisions around resource use, such as economic valuation, pragmatic use, and libertarian or egalitarian ideology. The land ethic serves as a conceptual nexus that integrates actors with different interests, and bridges world knowledge traditions. The application of a land ethic to online spaces might help ground otherwise-abstract digital communities and give a framework for service through deep time to these spaces. The ecological land ethic begins from a scientific

foundation, then introduces insights from psychology and philosophy to characterize the nature of proper human-ecosystem relationships. In the case of a digital commons ethic, the system is physically grounded in the software and hardware that are the enabling architecture of the online platform. Framing an empirical (computational) basis as a starting point for studying online discourse could allow a “rhetorical commons” ethic to emerge, as driven and structured by psychological and ethical preferences.

Approaches to collective governance of ecological and resource commons have also integrated the economic insights of Elinor Ostrom and others [185]. As with these ecological commons, digital governance and economic systems could be designed with specified functions, performance metrics, and a stated collective purpose [186]. This model of “digital commons as public good” has already been applied to online communities [187,188]. Connecting the notion of “rhetorical commons” to the economic game theoretic setting of the “tragedy of the commons” helps connect the behavior of users, to outcomes at the level of the commons [189].

## Conclusions and Recommendations

Can an ecological framework layered on rhetorical analysis help bridge the world of meaning and the capacities of computational pipelines? The ongoing and changing nature of the epistemic crisis requires new technological approaches towards scaling the modeling and understanding of our rhetorical commons. Here we expanded on previous appeals to rhetorical ecology and observations of the fundamental similarities between these fields [37], to posit the foundation for a type of system which might be able to infer, model, and intervene in multimedia digital discourse. With such a system, it could be possible to move beyond syntactic and user-driven understandings of digital discourse, to better observe and codify cycles and patterns within it, and to make progress towards ecologically-framed platform policies which can be more clearly informed by social preferences and values.

## Recommendations

- Review best practices in improving information quality of crowdsourced subject-matter tagging in physical, digital, and rhetorical ecosystem contexts.
- Review and synthesize research on argument mining methodologies using crowdsourced annotations.
- Research the implementation and limitations of applications and web extensions for providing lenses (e.g., enriched augmented

views of an object) on content displayed on various electronic devices.

- Curate a list of qualitative and quantitative patterns in the rhetorical structure and use of image memes.
- Consider users a part of an information commons rather than simply affected by an information system in future work on misinformation dynamics.
- Ensure that the identity, privacy, and preferences of users are protected in rhetorical cataloging schemes.

## Funding and Acknowledgements

R.J. Cordes is funded by the NSF Convergence Accelerator Trust and Authenticity in Communication Systems Program (NSF 21-572), under award ID #49100421C0036 and is supported in research efforts through a Nonresident Fellowship with the Atlantic Council on appointment to the GeoTech Center.

Daniel A. Friedman is funded by the NSF program Postdoctoral Research Fellowships in Biology (NSF 20-077), under award ID #2010290.

## Chapter III

---

# Swarmcheck

## Crowdsourced Argument Checking for Improving Rational Public Discourse

Marcin Woźniak

### Abstract

---

The goal of this work is to develop a system for verifying reasoning and detecting disinformation based on the use of AI-assisted argumentation and reusable knowledge base of linked theses, to protect the public from manipulation. The system will collect information by means of a browser plug-in allowing users to extract information for verification, a web application for discussion and browsing the knowledge base, a widget and a chatbot for discussion and automation of informing users about detected fake news and popular misinformation in public discourse.

---

# Introduction

Online discussions are riddled with disinformation, poor explanations, fake news, propaganda, and malicious actors. Organizations operating for their own profit, with ideological motives, as well as being financed by government agencies of various countries, spread them with the help of paid "trolls" or bots, and they are further copied by unaware Internet users as they often have attention-grabbing headlines, affect emotions, and rely on uncritical reception of information and confirmation bias. The negative effects of fake news include increasing social polarization, manipulation, destroying social trust, and creating barriers for rational public discourse. They can also pose a threat to public health, as in the case of the response to the pandemics or national security in a case of stochastic terrorism.

Standard response of this problem (if noticed) is to support fact-checking, verification of identity, awareness campaigns, or influence social media companies to have better content curation. The problem with the standard approach is that fact-checking is too long, expensive and demanding in comparison to spreading disinformation. Fact-checking reports often don't reach enough people and can be too difficult for the average person to read. Therefore, it is based on trust in authority. Any type of cultural change, even if successful, is too slow and the technological environment, in combination with business models of social media companies, doesn't make it easier to achieve. In addition, there are many coordination problems that prevent institutions from taking effective actions.

The goal of this work is to develop a system for verifying reasoning and detecting disinformation based on the use of AI-assisted argumentation and reusable knowledge base of linked theses, to protect the public from manipulation. The system will collect information by means of a browser plug-in allowing users to extract information for verification, a web application for discussion and browsing the knowledge base, a widget and a chatbot for discussion, and automation of informing users about detected fake news and popular misinformation in public discourse.

The information checking process relies on algorithmic techniques (e.g., Deep Learning, NLP, web crawling, semantic networks) and collective intelligence, users of the system do not need to be experts in fact-checking. The system will integrate dispersed discussion fragments into a coherent database of validated and reusable arguments, which will be used to evaluate information. Algorithms will be created to classify argumentation patterns, a resolution algorithm that provides as objective as possible information conflict resolution, and argument mining algorithms that help automate distributed argumentation processes. In line with an ethical approach to AI development, the system supports human capabilities, transparently, and



explainably presents justifications for accepted classifications using argumentation graphs, and enables simple error correction using natural language.

The software-as-a-service (SaaS) freemium business model will provide free access for all users accessing the public knowledge base (open access). There will be a fee for using private discussions for organisations that want to use it as a decision support system, and crowdsourcing tool. The system will be integrated into Swarmcheck's existing activities - consulting services in group decision support, strategy development, qualitative research and participatory processes, and popular science and education activities with partner communities.

## Solution scope and overview

Our proposed solution is based on the combination of collective intelligence, artificial intelligence and argument mapping. Swarmcheck is a web application and knowledge building system based on collective deliberation. It stores data in a graph format, with nodes representing theses in natural language and edges representing argument relationship. Its key features include: reusable reasoning from the public, searchable knowledge base, crowdsourcing of argumentation from various interfaces, coherence and error correction mechanisms and visualisation of results for users of various platforms in digestible form.

The “Swarm” in Swarmcheck refers to a large group of people working together to achieve a desired outcome. This crowdsourced approach provides an inclusive approach to diverse perspectives in which even small contributions are accounted for. While building upon previous knowledge users can go much deeper than traditional discussions. The “Check” in Swarmcheck refers to enhancing the collective intelligence of our users to help them validate gathered information. With reusable argumentation about proper argumentation (meta-argumentation), systematic and transparent analysis of reasoning provided by groups is performed and conflicts are resolved.

There are few key ideas behind our approach:

**Reusability of Reasoning.** Argumentation in different languages can be optimized for reusability of single arguments without the need of understanding their initial context. Therefore, different discussions that share common theses are creating a unified knowledgebase of validated reasoning. One could compare it to Wikipedia, but for arguments in natural language on graph megastructures.

**Defeasible Reasoning** - even very connected nodes on ground ideas can be transparently falsified or discarded with proper

reasoning, especially with reusable meta-argumentation (argumentation about validity of arguments).

**Ease of Correction.** We are agnostic to the Truth, but we know how to help people correct mistakes with the least effort possible on their part. Reasoning can be distributed in a way that makes the whole process of collective critical thinking easier, especially with the help of AI.

**Distributed Conflict Resolution.** There is no voting per se, nor human authority controlling conclusions. Distributed discussions allow for proper use of language to be the ultimate guide of structuring argumentation and resolving conflicts or mistakes in data.

We believe that cooperation with the Verified Information Environments Program, which is proposing the use of common annotation and identity infrastructure, will be of mutual benefit to our work.

## Relevance to VIE

The Verified Information Environment Program is seeking feedback on a set of questions in order to guide the development of its architecture, here we offer our perspectives on those most relevant to our work.

### **How do we rapidly develop resilient trust in digital contexts?**

Ideally, we wouldn't need to trust the organizations, governments, or experts. There is a reasonable assumption behind distrust of authorities and appeal to authority is an argumentative fallacy. Even if the conclusion stated by authority is right, it is not based on the fact that authority conducted the reasoning, but on the reasoning itself.

Authority of a person or institution in public discourse is oftentimes seen as a shortcut to reaching the correct conclusion by adapting the position that authority has. Reliance on trust of authority often backfires when perceived flaws of authority that have nothing to do with the reasoning itself make people distrust their words, look for information that confirms contradictory views and fall into confirmation bias. On the other hand, it is important not to dismiss the loss of trust in authority

as unfounded. No person or institution is 100% right all the time. People use heuristics to assess their level of trust, because most often they don't have any other viable option. We propose that with the right technological infrastructure, any particular nuance can be brought to the level of public assessment, making appeal by authority an unnecessary persuasive tool in public discourse.

Our solution proposes creating a domain agnostic way to build a knowledge base of connected reasons in natural language. Argument graphs have the ability to precisely and transparently present the line of reasoning in multiple user interfaces suited for particular users. At the same time reusable graphs can address the objections and counter arguments at the level unavailable to any singular authority. Even the meta-level objections (e.g., working of the system itself) can be addressed this way. It is not necessary for people to familiarize themselves with every line of reasoning on the topic upfront, but the availability of criticism and precise answers when an objection arises is crucial in breaking the process of constructing uninformed beliefs and worldviews.

When we finish the current development stage of our technology, we want to apply a business model in which all public discussions are free and open access. Our expenses will be covered by companies that want to have private discussions (those wouldn't be part of the public discussion either way). The better open access data is, the more incentivised companies will be to use it as a decision-making support system. Which provides an ethical business model - key factor in long term success of mission of the project.

### **How can trust signals be tied not just to individuals, but to individuals in context with subject matter?**

Following the previous line of reasoning, if there is no need for trust in a particular person it must be replaced with something. This trust can stem from open and searchable data in natural language represented as argument maps. In our approach, every argument is transparently accounted for, no matter who formulates it; and the reasoning is structured and linked to previous structured argumentation from different discussions

in a way that allows for deep understanding, retention of meaning and error correction across long graphs of reasoning.

This presumes that the process contributing arguments (supporting or critical) can not only be freely available, but easy to engage with. Firstly, the system should help users to find and recognise a relevant thesis in its “natural environment” like social media, webpage, or communication app. This can be achieved by combination of good extensions of the system (interfaces or API) and supporting algorithms (or other users in earlier phases) that recognize theses in natural language. Secondly the system should help the user formulate an argument or reuse a relevant argument from the database. Swarmcheck developed a unique methodology of algorithmically guiding users to formulate or reuse arguments outside of its original context. It is worth mentioning that we conducted projects that increased the ability to explain instructions for users (e.g., a board game was created to test this idea) and increased inclusivity of our software for people with disabilities, such as navigation through argument maps by blind and visually impaired people, and we are open to further develop those methods with an extended community.

After an argument is taken as an input in the system there is no evaluation based on authority but with other arguments. Every argument can be challenged with argumentation (and meta-argumentation) which are also reusable. A very important aspect of the system is that failed arguments are preserved, so that anybody can see why some line of reasoning failed. Thus, there is no need to make the same mistakes again and again, and collective memory of the discourse will increase. In principle, even bots that mimic human argumentation can contribute, even if only to a “failed arguments” knowledge base. More productive bots can be used for argument mining to increase the number of arguments available to users on less popular topics. Parts of the knowledge base that form responses to bad arguments will emerge as reusable structures for anybody to use with the aid of Swarmcheck. This effectively allows for the emergence of neighborhood-watch like mechanisms in the digital information environment.

### **How do we design social systems that self-govern?**

Proposed technology can function as a discursive decision-making system in which conclusions that have required weight translate to decisions. Integration of these solutions with other projects, for example blockchain based DAOs (decentralised autonomous organisations) can change the way institutions are governed in the future. Swarmcheck itself is designed to be managed by the system itself via argumentation in its more advanced technological stages.

Assuming that the system will work according to described principles it is reasonable to predict that organizations that will use it to make decisions will make more informed decisions and outcompete organizations that rely on more traditional approaches. This process can gradually lead to more saturation of the discursive decision making in public awareness and the cultural shift to rational public discourse that other approaches are proposing will emerge. Surprisingly, at early stages of the system even more than companies, local municipalities are interested in Swarmcheck. It proved to be a useful tool for public participatory process, conducting research and informed policymaking. We are hoping that in that sector we can counteract negative incentives that stem from election cycles, increase the ability of democratic countries for long term planning and collaborative governance.

Currently we are using interactive argument maps as an interface which interestingly increases critical thinking<sup>1</sup> and retention of knowledge in users.<sup>2</sup> This creates an interesting feedback loop between the system and the users that increases the capabilities of both sides with time. Our R&D plan assumes that in 2023 we will be able to extract structured reasoning from various interfaces simultaneously. We believe that with scale and automation it is possible to run it even in the background of offline conversations, while those conversations could

---

<sup>1</sup> For discussion of the impact of argument maps on critical thinking, please see Twardy, Charles. "Argument maps improve critical thinking." *Teaching philosophy* 27.2 (2004): 95-116.

<sup>2</sup> Based on Swarmcheck's own research, currently unpublished.

influence high-level decisions and solve coordination problems, based on the merit of the argumentation.

It is worth pointing out that this technology can function as a medium that facilitates negotiations and discourse among agents with different levels of intelligence in natural language. This could constitute a large step in “AI-safety via debate”<sup>3</sup> strategy, and to some extent explainable AI.<sup>4</sup>

---

<sup>3</sup> For discussion of the “AI-safety via Debate” strategy, please see Irving, Geoffrey, Paul Christiano, and Dario Amodei. “AI safety via debate.” arXiv preprint arXiv:1805.00899 (2018).

<sup>4</sup> Swarmcheck is working as a solution to “visible thoughts project”  
<https://www.alignmentforum.org/posts/zRn6cLtxyNodudzhw/visible-thoughts-project-and-bounty-announcement?fbclid=IwAR28P4u5x6QoVLUHQMpBEA4EwXJjytkkqoKwM2f4hiPfhPF5GhHcx5Jmcyk>

# Proposed Next Steps

The proposed next steps are contained within a three-phase research and development cycle, based on the existing Swarmcheck system, which will be publicly available to Swarmcheck's test user community.

Each phase contains parallel research and development tasks that will complement each other in short, weekly feedback sprints. The research tasks will focus on the development of algorithms to support the recording and evaluation of information in the form of reusable argumentation and the expansion of the database. The development work will consist of implementation activities while creating a viable system that will be able to deliver the results analyzed in the research work with real users. The research team will continuously draw conclusions from the results and introduce further modifications to the ways of operation of particular algorithms, procedures and collected data, and will present them in the form of guidelines for development tasks (test driven development and RITE [Readable, Isolated or Integrated, Thorough, Explicit] approach).

In each phase, both research and development tasks will run for 12 months in parallel. Development work should be conducted from the beginning of each phase, since we will be basing on the system currently in use and being developed, and the project is a continuation of the system development vision. The course of work in the project assumes the development of new components to the system in accordance with the requirements and research and development goals for a given phase. The research work will also continue until the last month, as this time is needed for the overall phase results, database documentation, and description of the solution according to the guidelines of the expected results.

For the detection of fake-news, the system has to combine the rigor of logical, correct reasoning, with the freedom of performing user-understandable actions in natural language with a systematic approach to information analysis. The way algorithms, procedures, and data sets are created and analyzed requires interdisciplinary knowledge of argumentation analysis, logic, computer-aided argumentation, philosophy of science, cognitive science, semantic networks, NLP, expert systems, multi-agent systems, collective intelligence ethos analysis, data extraction and engineering methods, UX, qualitative information representation design, and ethics.

## Phase I

The aim of Phase I is to create, develop and prove the basic functional assumptions of the system (proof of concept) for classifying information into correct or incorrect. The final classification of information will be performed by the users of



the information system based on the implemented algorithmic procedures and a database of reproducible argumentation schemes. In all phases, argumentation specialists will animate and analyze discussions with test users (both passively and actively participating in them), which will allow mapping structured argumentation in natural language in the database for further testing and machine learning. Argumentation analysis will allow us to identify repetitive argument patterns present in disinformation analysis (in two languages: Polish and English), and on the basis of reuse of the identified patterns a model of argumentation evaluation will be created. During phase 1 a detailed standard and plan for data collection, storage and engineering will also be developed, which will be used throughout the project for statistical, qualitative, ethos and machine learning analysis. Argumentation will receive automated tags that will be used to integrate the internal knowledge base with available semantic networks and external knowledge bases (linked data approach). NLP models will be tested to search and analyze available Polish and English language data and to dissect statements and standardize theses. If the models in English are more effective, steps will be proposed for partially automated translation of the results between the languages to achieve the best results for Polish.

### **Phase II**

In Phase II, knowledge is gained about the limitations of the original prototype (and thus of argumentation systems as a type of knowledge evaluation systems), and an extension of the application categories for solving generic information classification problems. An interdisciplinary team will test systems for resolving data conflicts in argumentation that may arise under the intended use scenarios. New features of the system will be described in the form of user stories and tested with users after implemented in the application. The tests, which consist in an attempt to cheat the implemented methodology, will allow improvement of the existing algorithms and create new ones that will increase the system's reliability. In the development work, apart from the implementation of the research results, extensions of the system will be developed in the form of subsequent versions of browser plug-ins and chatbots. browser plug-ins, chatbots and integration with external systems API.

### Phase III

In phase III we will examine how the new functionalities increase the speed of argumentation and the accuracy of classification in comparison to competitive systems, ready-made examples and content independently evaluated by experts. We will apply methods of iterative prototyping and testing of algorithms from previous phases. In this phase, methods based on deep learning and web crawling will be used to support argument mining automation. Particular attention will be paid to usability testing among groups of recipients. Users will be surveyed about the usability of the proposed solutions and their understanding of the content presented by the system will be tested (knowledge test, A/B comparison tests, time measurement, number of clicks, user stories).

## Recommendations

The following recommendations are offered to the VIE Program:

**Scale up rational deliberation.** More intelligent systems Support and finance research and development and implementation of collective decision-making systems, select projects based on:

- Correct Reasoning
- Inclusiveness
- Scalability

**Transparent reasoning is a prerequisite for trust.** It is reasonable to switch efforts from ensuring the trust of authority to ease verification of visible reasoning in policymaking and science communication.

**Incentivize transparency of reasoning.** It is worth investigating how to make public organizations and governments more transparent in showcasing their goals and reasoning, and how to make them more accountable to changing when public understanding of their reasoning change

**Invest in XAI research.** Explainable AI (XAI) and transparency in algorithmic decision making can have great impact on society and require additional attention and funding.

Currently we are financing our team of developers, designers, and philosophers by providing services, mostly for public administration in Poland, especially through creation of participatory strategies and public programs. We have a good track record and our clients come back to us and recommend the usage of the system to other municipalities. Most of the procedures that are described here as algorithmic are conducted by trained philosophers and moderators in real time. This gives us initial validation of the concept and many insights on design of the next technological steps. We would be happy to share our experiences in this regard with other initiatives and researchers.

# How to Build a Truly Modular System or Organization

Łukasz Wilisowski

### Abstract

---

Everybody talks about the modular approach to building large systems, but what does it actually mean? In this article, modularity-related insights are discussed that may be of use both to systems-engineering related development teams and to organizations generally. Recommendations and general design principles are offered.

---

# Introduction

Everybody talks about the modular approach to building large systems. But what does it actually mean? In this article, I will try to share modularity-related insights that might be useful not only for development teams working on systems engineering solutions, but also to organizations generally.

The point I want to start with: making software is hard. The best way to prove it is to fully answer a seemingly obvious question: how do computer science experts, solution architects, and tech leaders choose technologies they recommend to the client?

The general assumption is that the experts:

- assess the full project scope and resources
- compare all available technologies by their quality and project fit
- minimize risk factors and overall client's costs

But in reality, they'd rather:

- choose technology that they already know (which is rarely the best technology out there)
- choose technology that they want to learn (yes, that is correct)
- discount risk factors and client's costs

Before you come to the wrong conclusion, let me elaborate. It is not that experts have bad intentions or lack competence; on the contrary they are just faced with a virtually impossible task. To explain, we can compare software development to road construction:

Both domains have their experts, and both domains have projects that require careful planning, including assessment of client and supplier resources, time constraints, and other factors. There are 3 differences in particular which are relevant for our discussion:

**Options.** There are relatively few technologies (techniques and materials) used to build roads in common use. In programming, you have dozens of well-known languages, each of them having thousands or more libraries, with overlapping functionalities. On top of that, you have many parallel design patterns which

you use to build and scale up the system. The number of possible ways in which you can build a system is gigantic.

**Novelty.** Road construction is a relatively repeatable task. You can compare one solution to another. In software, each application is unique, you never build the same system twice. You cannot directly compare different approaches, although some techniques such as prototyping allow you to revoke bad technological choices at an early stage (according to the fail-fast principle). But generally, building new software means dealing with an unknown.

**Stability.** Road construction techniques are relatively stable, they evolve slowly, and changes in the process are introduced rarely. In programming, the whole industry is in a state of constant change. Programming languages evolve, thousands of new software libraries are published every day. There is no top-down control over the community. Because of that, every programmer (no matter how proficient) needs to spend the majority of one's stuff learning new things. You just simply never stop learning.

If mastering one programming language, technology, or library takes years, no project is the same, and the number of options is endless - how can we expect that a single expert can analyze and compare every possible solution? We can't.

## Approach

To address this problem of dealing with unknowns, three guidelines have been established:

- Be Agile and Ready to Change
- Allow for Experimentation
- Reuse Good Practices

Good practices (also called design principles) are replicable across different programming languages and technologies, hence their success. Examples of such principles include microservices and plugin-module architecture.

However, practices are subject to trends. For example, microservices are a very powerful way to build scalable applications, but their utility in small to average systems is somewhat disputable. This problem of fashionable theories and

approaches taking over the landscape is not a software-specific problem, and can be found elsewhere - for example, this problem has been argued to be present in the success of String Theory in the domain of physics despite the lack of substantial evidence confirming the theory.<sup>1</sup>

Returning to the primary topic of the article, Modularity - to understand what a module is, we can represent a software system on 2-dimensions, technical and business (see Figure 1). From this perspective, *everything* can be considered a module. The software system can be built of horizontal slices (technical modules) and vertical slices (business modules).

We can think of modules (especially technical ones) as compressing information and creating hierarchies of complexity. A good example of this would be TCP/IP network layers. Each module provides an interface (which can be referred to as a contract, abstraction, or generalization), which contains a compressed definition of how to use it. For example, when you shop online and pay for the goods, the PayPal service promises to pay to the store the transferred amount, without you having to worry about how it all works. It hides all the unnecessary details from you and promises that none of those details will ever affect your user experience. In a situation when it does, such as where a general contract depends on implementation details, we call it a leaking abstraction.

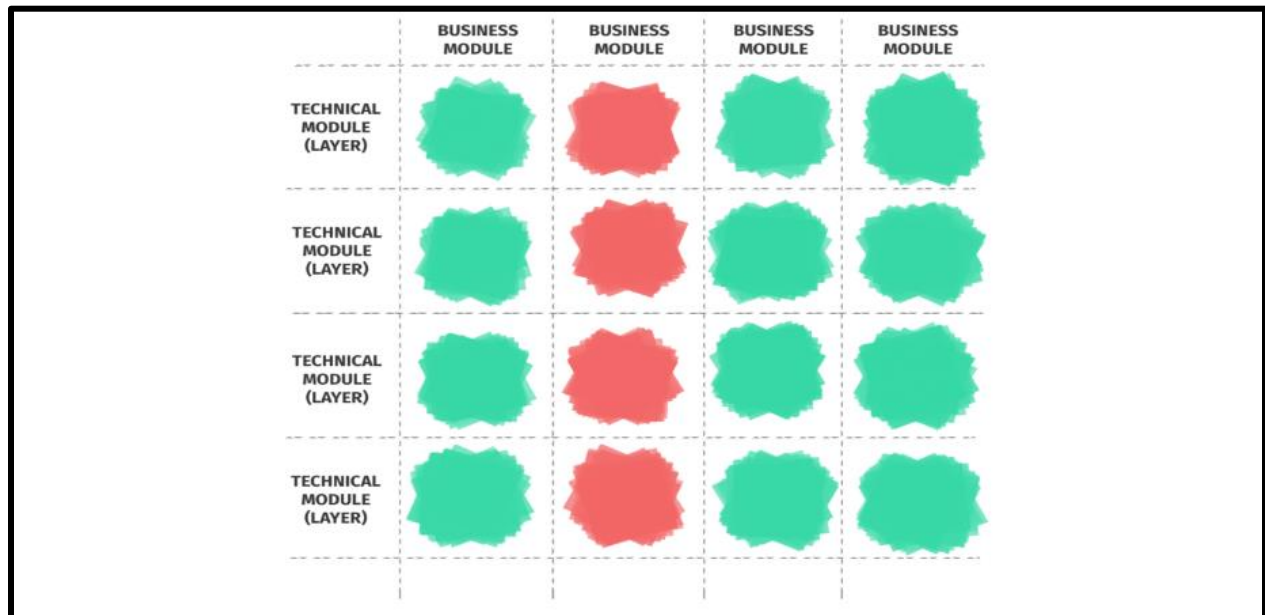


Figure 1. 2-Dimensional Module Representation

---

<sup>1</sup> For discussion of the problem of fashionable theory and the success of String Theory, please see the book *Fashion, Faith, and Fantasy*, by Sir Roger Penrose.



## Standard vs Modular Approach

Now, we turn to a very important question. What is the difference between a standard approach and a truly modular approach? In traditional programming of technical modules (layers), we have had these long-established assumptions:

- upper layers in hierarchy depend on (extend) lower layers
- lower layers in hierarchy do not know anything about the upper layers

The truly modular approach requires that we bring these assumptions into the business domain (see Figure 2).

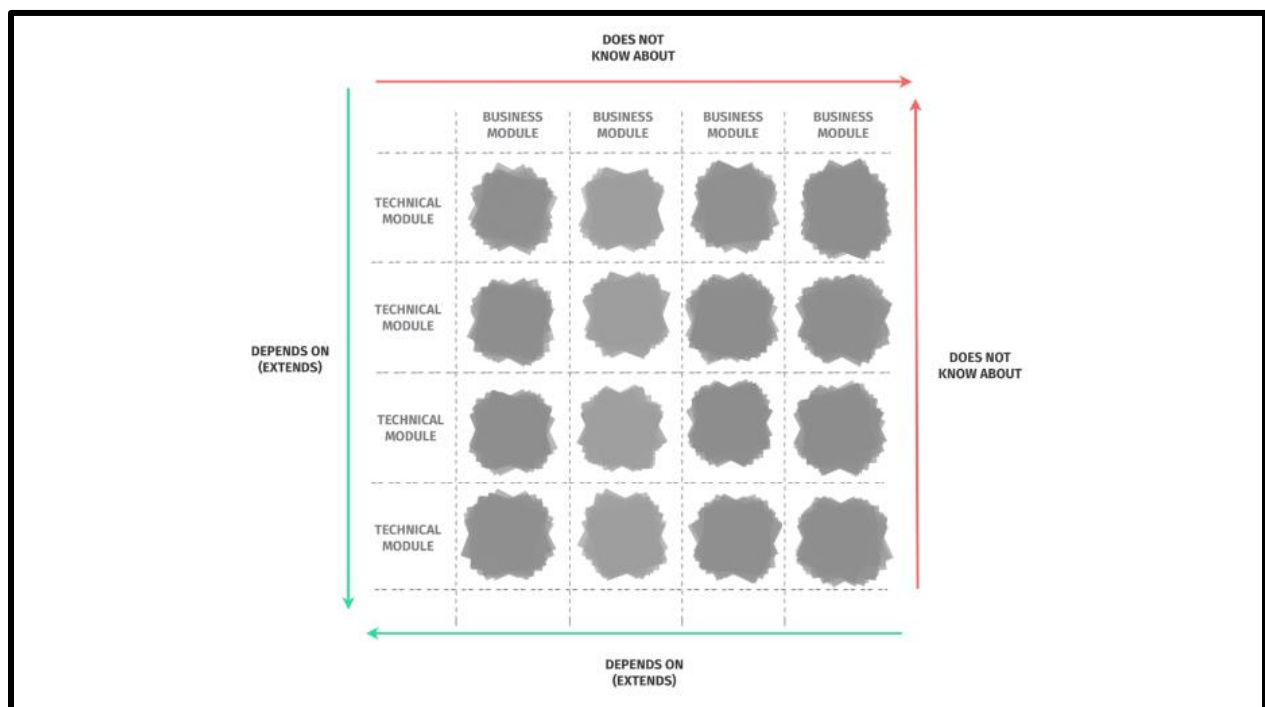


Figure 2. Extensibility of Modules in the Business Domain

Modules, following these assumptions, will form a dependency tree, with branches. Within this tree, any given module will be aware only of the modules it uses. Incidentally, by using this definition, one has a simple method to test for modularity:

- A true module should be able to be removed with one click.
- A false module will require a lot of changes to be fully removed, because of dependencies hard-coded within its parents.

For example, the Lego block-building system has a truly modular structure. Most parts of the model can be removed in a single step. A Lego piece does not know what other pieces are connected to it (the information about an extension is part of the extending piece, not

an extended one). On the other hand, classic software approaches rarely result in a modular structure. The website's home page knows all its sub-components. Removing any sub-component (child) often requires modification of the home page (parent).

## Modularity in Social Systems

This way of thinking about modularity is useful beyond the field of computer science. For example, in terms of employment, a modular employee might be considered one who can leave the company or take time off without any loss for the company, without describing the employee as replaceable.

To build on this example, consider two hypothetical employees, Mark and John:

- Mark has the best knowledge in the company about technology X. He knows the details that other people do not. When there are problems with technology X, everybody turns to Mark. Mark has created effective processes for years and everything depends on him.
- John shares all the knowledge with his team members. He uses his experience to teach and guide others. He wants to create the next experts. He contributes to creating processes but makes final decisions together with his team. He has a boring habit of keeping documentation clear and up to date.

If John takes time off or the team ran into common errors while he was not there the company will continue to operate on a normal basis. If Mark takes time off, work could come to a halt, or worse.

## Concluding Recommendations

There are many ways to implement a modular architecture, depending on the technology that is in use. The following are general design principles that are reusable and language or use-case agnostic:

- Think of a module as an addition to the existing code-base (for example a new folder with the source code). Your architecture should be constructed in a way that allows making any possible change by adding code only as opposed to removing old code.
- If there is only one use case associated with a particular module, invent a second one. This could be a hypothetical use case or a different kind of client. This process will improve the abstraction and generalizability of your modules.

## **Structuring the Information Commons**

### IV – How to Build a Truly Modular System or Organization

- Do not tie yourself to specific technologies, because they are currently best on the market (database, front-end, back-end). This keeps your architecture flexible.
- Use dependency injection and inversion of control. Specify contracts for extensions, but do not import the specific ones.
- Your architecture should allow you to add more than one module at the same level within a dependency tree. Parallel modules are only possible when you have properly implemented extensions.
- Make the code extensible. It will allow you to outsource work to many sub-teams without them interfering with one another.

## Chapter V

---

# Active Inference in Modeling Conflict

## A Framework for Modeling Conflict in Business, Operations, Legal, Technical, and Social Contexts

Scott David, R.J. Cordes,  
& Daniel A. Friedman

### Abstract

---

In this paper, we integrate conflict studies with Active Inference, a developing framework which provides an integrative and systems-level perspective on cognition and behavior. This formalization, the Active Inference Conflict (AIC) model, situates conflict in terms of a multiscale process of communication, trust, and relationship management enacted by interacting entities. The AIC model helps capture and extend the insights of previous models applied to aspects of conflict and war, such as OODA loops (observe-orient-decide-act), the generations of warfare model, and the Rumsfeld Matrix. The AIC model aids in the analysis of pertinent aspects of modern conflict, such as cyber, psychological, biological, informational, financial, and ideological conflict, that are not amenable to coherent or consistent analysis using traditional models of human conflict. AIC is demonstrated to be of use in both monitoring and studying conflict, as well as in designing systems intended to facilitate controlled or managed conflict in scenarios characterized by business, operations, legal, technical, and social (BOLTS) components. Insights and implications from qualitative use are used as a foundation for offering recommendations for future research and social systems design.

---

Active Inference in Modeling Conflict was originally published in the COGSEC 2021 volume "Narrative Information Ecosystems: Conflict and Trust on the Endless Frontier".

# Introduction

Human-scale conflict constituting “war” in its various incarnations has been studied from a variety of perspectives, including, but not limited to, statistical, ethnographic, logistical, sociological, legal, and philosophical frameworks. However, with the notable advances made in the capabilities of weapons systems and the introduction of global defense pacts made in the 20th Century, the risk calculus of triggering an official declaration of war has changed. The resulting dramatic increase in costs and displacements of kinetic war compels state and non-state actors to pursue their conflicting interests through alternative means. The resulting complex threat surfaces are not always well-described or modeled by existing frameworks for conflict (which usually have a military or domain-specific focus), which further amplifies risk even in tractable scenarios [1]. In this paper, we make use of Active Inference (ActInf), a framework which provides an integrative and systems-level perspective on cognition and behavior, to propose a new formalization of conflict in terms of a multiscale process of communication, trust, and relationship management enacted by interacting entities. This application of ActInf to questions of conflict, called the Active Inference Conflict (AIC) Model, extends recent work on Active Inference and human-robot trust system [2], cyberphysical systems [3], and societies as cognitive agents [4] to the domains of human conflict in expanding shared information environments.

The AIC model is grounded in several previous frameworks for action and conflict from military science, including the generations of warfare (GW) model, observe-orient-decide-act (OODA) loop, and the Rumsfeld Matrix. Additionally, the AIC extends these models to better describe, frame, and offer recommendations for the current and projected future nature of war and other forms of conflict, which is increasingly non-kinetic. The AIC model is intended to offer generalization beyond conflict itself, helping not just to describe nation-state conflicts, but also complex multi-scale conflicts involving individuals and communities in contexts characterized by their business, operations, legal, technical, and social (BOLTS) components. The essential historical insights gleaned from the GW model offer a useful foundation from which this paper’s ActInf framing can be understood, and establishes a new chapter in the GW model’s framing of the timeless yet ever-changing aspects of human conflict.

In this paper, we begin by offering a survey of past qualitative and quantitative models of conflict and the insights they provide. After this survey, we consider the essential features of the reviewed models, and highlight the need for models which offer more interoperability and generalization in order to stay relevant in the face of an ever-changing expression of conflict. We then offer a primer on the ActInf framework in terms of core terms and features. Following this description, we

explore how the AIC model can extend previous models such as OODA and GW while still capturing their essential insights. In this exploration, special attention is given to how the AIC model relates to the Rumsfeld Matrix, and what this relationship may reveal about Rumsfeld's oft-neglected quadrant, the "unknown-knowns". We suggest that management of relationship and conflict with a prioritization of the often neglected "unknown-knowns" quadrant provides a pathway to multi-scale risk mitigation and leverage points for human interactions online. In summary, AIC is revealed to be more than just a powerful new model of war and conflict. AIC framing also invites consideration of how humans can harness the destructive energies of prior conflagrations of conflict at all levels into constructive systems that can perform useful "work" by converting the underlying information differentials of conflict into new forms of value the benefits of which can be distributed in managed ways to maintain the generative AIC apparatus (analogous to how an engine extracts useful work from heat gradients). The AIC model is an applied Active Inference approach for mitigating risk and enhancing value from the ever-increasing informational component of modern interactions. Finally, we conclude with a summary of insights and recommendations for future research and application.

## Previous Models of Military Conflict

Being of obvious, existential importance to state sovereigns, war and conflict has been a subject of interest to historians, scholars, and artists since the birth of civilization. As evidenced by the hundreds of thousands of books written about the American Civil War alone [5], and a history of scholarship which extends back to some of the earliest books ever written [6], the subject of war has an unfathomably large literary and oral corpus. The vastness of the body of literature on war suggests that even if only a small fraction of the corpus is dedicated to generalizing and modeling war (the rest being historical documentation and analysis of instances of war), it would still constitute a significant body of literature in itself. For purposes of this article, and in the interest of presenting a referenceable review of past models and generalizations of war (while acknowledging that it is an impossible task to describe them all), we present past models of war and conflict in the following categories:

- Narrative Models
- Quantitative Models
- Conflict Information Flows and Decision-Making Models

## Narrative Models of Conflict

The term narrative model is used here to describe formal and semiformal models of conflict which were intended to provide guidance and actionable insight to strategic commanders through the use of qualitative, non-technical methods such as storytelling, aphorism, historical example, parables, and slogans.

## Collections of Heuristics

The earliest attempts to create and compile informative representations of conflict and war do not offer integrated models in a modern sense, instead they offer collections of axioms, idioms, recipes, rules, principles, and patterns - rules of thumb, based on insights drawn from the experiences of the offeror. One of the oldest examples of these collections is Flavius Vegetius Renatus' *Epitome Rei Militaris*, or "Epitome of Military Science" [7]. It is one of the few surviving Roman-era works on military science and art from its time and was routinely used during the Middle Ages to augment and inform writings on warfare [7].

Though much of its content deals with specific questions about routine situations in which Roman commanders may have found themselves, such as in what kind of places camps should be built or how a suitable place might be chosen for battle, a section of the Epitome titled "General Rules of Warfare" also supplies "basic principles in an unspecific form which could be adapted to serve a great variety of military situations" [7]. These include:

- "It is difficult to beat someone who can form a true estimate of his own and the enemy's forces"
- "He who spends more time watching in outposts and puts more effort into training soldiers, will be less subject to danger"
- "Never lead forth a soldier to a general engagement except when you see that he expects victory"

[7]

Examples from other well-known collections of timeless heuristics relating to war throughout history and across cultures provide similar sorts of insights, such as the following:

From Sun Tzu's Art of War

- "A skillful soldier does not raise a second levy"
- "In order to kill the enemy, our men must be roused to anger"



- “If equally matched, we can offer battle; if slightly inferior in numbers, we can avoid the enemy; if quite unequal in every way, we can flee”
- “If you know the enemy and know yourself, you need not fear the result of a hundred battles. If you know yourself but not the enemy, for every victory gained you will also suffer a defeat. If you know neither the enemy nor yourself, you will succumb in every battle.”

[6]

From Moltke’s Art of War

- “Excessive extension of the front brings danger of a breakthrough.”
- “Engagements in forests last for a long time”
- “One must immediately prepare supporting points captured in an engagement for defense in order to thwart the enemy’s efforts to recapture them”

[8]

Countless other works elaborating the art of war, provide detailed rules, patterns, and axioms of human armed conflict, such as those by Mao Tse-tung, Machiavelli, and Sun Bin [9–11]. When these collections are viewed as part of a common ensemble of axioms, bundled together, they may be argued to constitute nascent narrative models of warfare, helping generals, real or armchair, better understand the complex and challenging scenarios of conflict they are encountering, simulating, or studying.

Also included within these collections of heuristics are later works from the 1800’s, such as Antoine-Henri Jomini’s Art of War [12] and Carl von Clausewitz’s On War [13]. While both these books provide their fair share of axioms and rules like earlier works, they also move beyond simple heuristics in an attempt to capture more generalizable models and frameworks for understanding and describing the underlying causes and motivations of warfare as an aid to formulating strategy and tactics for engagement. These developments signal an increasing awareness of the behaviors of war as part of the larger set of behaviors associated with human interactions and the conflict that they generate.

For example, Jomini provides the following frameworks for understanding the nature of conflict, moving beyond a mere description of the practices of war to its underlying contexts of conflict to encourage an enhancement of the understanding of how best to engage [12]. Several of Jomini’s classification schemes are excerpted here:

Eight types of motivations for states to engage in warfare:

- “To reclaim rights or defend them...
- to protect and maintain the great interests of the state...
- To uphold neighboring states...
- To fulfill obligations...
- To propagate political or religious theories...
- To increase the influence and power of the state...
- To defend the threatened independence of the state...
- To avenge insulted honor...
- From a mania for conquest.”

Two kinds of international Intervention:

- “Intervention in the internal affairs of neighboring states...
- intervention in external relations”

And four kinds of war which result from such an intervention:

- “Where the intervention is merely auxiliary, and with a force specified by former treaties...
- where the intervention is to uphold a feeble neighbor by defending his territory, thus shifting the scene of war to other soil...
- A state interferes as a principal party when near the theater of war, - which supposes the case of a coalition of several powers against one...
- a state interferes either in a struggle already in progress, or interferes before declaration of war”

[12]

Clausewitz offers similar context-enhancing frameworks for war, but goes farther, arguing that even more generalizable analysis is needed and that those who “never rise above anecdote” will “never get down to the general factors that govern the matter... indeed they will consider a philosophy that encompasses the general run of cases as a mere dream” [13]. Clausewitz recognized that theory informs practice, and that awareness of context and causation of war as a form of human conflict

provides valuable insights into the strategies and tactics for its effective engagement. Clausewitz was well aware of the limitations of prior descriptions of warfare, and made explicit the benefits of more comprehensive and multi-dimensional models that situated warfare among other forms of human conflict.

## Trinity of War

Carl von Clausewitz, in pursuit of deeper generalizations, proposed what may be the earliest framework for describing warfare that is recognizable, on its face, as a generalizable model. He suggests that war is an extension of state policy, and as such, it is ruled by a “paradoxical trinity” of forces [13]. His description of this trinity is excerpted here:

*“The first of these three aspects mainly concerns the people; the second the commander and his army; the third the government. The passions that are to be kindled in war must already be inherent in the people; the scope which the play of courage and talent will enjoy in the realm of probability and chance depends on the particular character of the commander and the army; but the political aims are the business of government alone.*

*These three tendencies are like three different codes of law, deep-rooted in their subject and yet variable in their relationship to one another. A theory that ignores any one of them or seeks to fix an arbitrary relationship between them would conflict with reality to such an extent that for this reason alone it would be totally useless...*

*Our task therefore is to develop a theory that maintains a balance between these three tendencies, like an object suspended between three magnets.”*

[13]

The trinity of war model captures the multi-node complexity that yields the nonlinear aspects of what motivates and channels the expression of those motivations in kinetic conflict. Further, it helps described certain non-combat oriented insights regarding conflict, such as war being conceptualized as an extension of political conflict [14], that it is motivated by state interest or *raison d'état*, and is moderated by a state's ability to channel the motivations of both civilians and military personnel toward conflict [15].

What may be the most important aspect of Clausewitz's model however, is that it was far ahead of its time in framing war as something akin to a complex system

rather than a mechanistic process, in which a trinity of “chance, uncertainty, and friction... will make anticipation of even the first-order consequences of military action highly conjectural” [16,17].

## **Military Revolutions Model**

Among the various categories of qualitative planning and descriptive models which have come into (and gone out of) fashion within the United States military was a collection of models centered on “revolutions in military affairs”, which grew to “increasing prominence in Washington’s Byzantine budgetary and procurement struggles” in the 1990s [18], and served to rhetorically bind together technical and modeling advances. Initially just a reference by Western historians and Soviet military theorists to the notion of key historical inflection points in which there were unforeseeable, “fundamental [and] systemic” changes in the expression of war, the “military revolutions model” was picked up by the US defense community as a concept that was also considered valuable for doctrine and planning [18]. Since that time, numerous attempts have been made to model and chart these revolutions in order to help military leadership better understand their place both in history and in current affairs, and to help them plan for the future. Some examples of these models are surveyed below.

### **Krepinevich Model**

The model presented by Krepinevich was one of the earlier attempts at formalization of the historical revolutions in military affairs. While the revolutions specifically noted by Krepinevich have been greatly modified or even abandoned in later models, his formalization of the elements underneath military revolutions has stayed relevant [18]. These elements were said to consist of technological change, systems development, operational innovation, and organizational adaptation [18,19]. The historical revolutions noted by Krepinevich, in chronological order, are as follows:

- Infantry Revolution
- Artillery Revolution
- Revolution of Sail and Shot
- Fortress Revolution
- Gunpowder Revolution
- Napoleonic Revolution
- Land warfare Revolution

- Naval Revolution
- Revolutions in Mechanization, Aviation, and Information
- Nuclear Revolution

[19]

Krepinevich's model is unique among the other historical revolution models for its focus on warfare alone. Notwithstanding the focus on war, he recognized that changes in technology, which are themselves generated by the larger social and historical context, affect the nature of engagement in war. In a sense, he saw technology as the vehicle through which large scale social and historical changes affect war. Among the more valuable insights he derives from this model is that technological innovation does not guarantee a revolution in military affairs - instead, these revolutions occur when states change their process, systems, and organization in order to incorporate those innovations [19].

### **Knox and Murray Model**

Knox and Murray's take on the revolutions in military affairs model [20] was built from its predecessors, incorporating key elements from Krepinevich, which they considered "typical" and fundamental to models of this kind [18]. What sets Knox and Murray's model apart from its predecessors however, is three-fold. First, they explicitly included non-military systemic changes within the scope of revolutions in military affairs, such as those related to economies beyond the ability to supply armament. Second, they see each of the revolutions as reflecting, not just the innovations of its time, but also the novel combination and integration of the innovations and resulting changes of its predecessors. Third, they include two separate tracks of revolutions, seemingly inspired by Krepinevich's suggestion that the inflection points in expression of warfare were separable from the implementations and incorporations of technological innovations. One was termed "military revolutions", the other, "revolutions in military affairs", referring to abstract inflection points and revolutionary implementations, respectively [18]. A summary of their charting of revolutions is included here:

- Precursory, or “anticipatory” Revolutions in Military Affairs
  - The introduction of the longbow, gunpowder, and fortress architecture
- Military Revolution I: The Modernization of the State and its Military Institutions

Associated revolutions of military affairs:

- Dutch, Swedish, and French tactical and organizational reforms
  - Britain’s financial revolution
- Military Revolutions II and III: The French and Industrial Revolutions

Associated revolutions of military affairs:

- Napoleonic warfare and the complete battlefield annihilation of the enemy’s armed forces)
  - Transportation: railroads, steamships
  - Armament: combination of quick-firing small arms and artillery
  - Communications: telegraph
- The Fisher Revolution
    - The introduction of “all-big-gun” battleships
  - Military Revolution IV: The First World War and its Irrevocable Combination of Preceding Revolutions

Associated revolutions of military affairs:

- Combined Arms Tactics
  - Blitzkrieg Operations
  - Carrier, Submarine, and Amphibious Warfare
  - Radar and Signals Intelligence
- Military Revolution V: Nuclear Weapons and Ballistic Delivery Systems

Associated revolutions of military affairs:

- Precision Reconnaissance and Strike
- Stealth Systems

- Increased Lethality of Conventional Munitions

[20]

### **Hoffman Model**

Hoffman, a former US Marine Corps infantry Officer with 4 decades of experience as a national security analyst, offers one of the most recent models of military revolutions which expands on and challenges aspects of the Knox and Murray model [21]. Hoffman focuses on what comes after the five revolutions within the Knox and Murray model through the lens of the Clausewitz trinity, considering how human-machine teaming, the end of the “heroic age” of the military, and automated systems might affect various aspects of war, social stability, and public sentiment toward policy [21]. He expands the Knox and Murray model to seven revolutions, with a more explicit emphasis on non-violent phenomena, such as ideological extremism [21]. A summary of the Hoffman model of military revolutions (and their key features) is included here:

- Westphalian System
  - Revenue generation, banking and taxes, and the introduction of professional militaries
- French Revolution
  - National mobilization and levy en masse
- Industrial Revolution
  - Mass production, standardization, and large-scale economic exploitation
- World Wars
  - Combined arms, armored blitzkrieg, carriers, bombers, and jets
- Nuclear Revolution
  - Nuclear weapons and intercontinental ballistic missiles
- Information Revolution
  - Command and control, connectivity and global reach, imagery, and ideological levy en masse
- Autonomous Revolution



- Autonomous weapons, swarms of robotic vehicles, self-organizing defense systems, big data analytics, and deep-learning systems.

[21]

### **Generations of Warfare Framework**

In the late 1980s, William Lind and a collection of US Military officers from the US Army and Marine Corps presented what is now known as the “Generations of Warfare” (GW) framework in an article published in the Marine Corps Gazette [22]. It is notably similar to the military revolutions model both in terms of its intentions and structure. The GW framework is built on the notion of linear sequential development over time, marked by key inflection points driven by technology and ideas. The GW framework has arguably achieved broad use and has received a great deal of commentary and adaptation, for example the projection of a fifth generation of war (5GW) beyond the four initially described [23]. A summary of the initial conception of the four generations of warfare is provided here:

- First generation: Line and Column Tactics
  - Driven by technological changes
  - Operational Art practiced by individual commanders
  - (e.g., Napoleon)
  - Reliance on indirect fire (e.g., artillery)
- Second generation: Fire and Movement
  - Driven primarily by technological changes, but also by ideological changes
  - Operational art practiced by high-ranking officers
  - Reliance on massed firepower, and manpower
- Third generation: Nonlinear Tactics
  - Driven primarily by ideological changes, but also technological changes
  - Operational art practiced by low-ranking officers (e.g., tank commanders)
  - Reliance on maneuvers and non-linear tactics
- Fourth generation: Whole of Society

- Driven primarily by ideological changes
- Operational art practiced in small-teams and in the gray zone between military and civilian
- Reliance on gray zone warfare (e.g., psychological and informational operations, targeting a society's culture)

[22]

### **Gradients of Warfare**

The “gradients” of warfare model (xGW) proposed by Daniel Abbott is a reimagining of the generations and revolutions models of framing changes in warfare [23]. Although the gradient and generation are often used interchangeably, the gradient model abandons chronological development (generations) and instead describes movement along a single finite, abstract axis, representing an arbitrary gradient of diffusion or concentration related to a particular conflict [23]. The gradients described by Abbott [23] are summarized below:

- The Zeroth Gradient
  - Genocide and all-of-society warfare (e.g., ant colonies, ethnic cleansings)
- The First Gradient
  - Physical concentration of resources (e.g., chimpanzee border patrols, medieval warfare)
  - Placing troops in the same place at the same time
- The Second Gradient
  - Concentration of effort (e.g., coordinated fire)
  - Directing effort toward the same place at the same time
- The Third Gradient
  - Coordination and concentration of operational art (e.g., blitzkrieg)
- The Fourth Gradient
  - Focus on “degrading the opponent into an earlier generation of warfare”

- Decentralized gray zone conflict
- The Fifth Gradient
  - Coordination and concentration of ideology

[23]

### **Kohalyk's Projection of xGW**

An interesting result of abandoning chronology as a primary axis and replacing it with axes related to abstract state features is that Abbot's gradients may be "projected" onto other models to yield additional insights from existing models. For example, Kohalyk, based on Abbott's assertions about the nature of the gradients, projects the gradients onto John Boyd's famous OODA (observe, orient, decide, act) loop (see Figure 1) [24,25]. This exercise demonstrates that Abbot's gradients can be repurposed, not just to describe levels of diffusion, but also the basis for that diffusion and the changes to that basis over time, providing a more stable view on the generations of warfare model that gradients were originally intended to replace [24]. This projection can be summarized as follows:

- The First Gradient
  - "Characterized by prioritizing the transition between decision and action"
- The Second Gradient
  - "Characterized by prioritizing the gap between orientation and decision"
- The Third Gradient
  - "Characterized by prioritizing the disruption of orientation"
- The Fourth Gradient
  - "Characterized by prioritizing the gap between observation and orientation"
- The Fifth Gradient
  - "Prioritization of the disruption of observation itself"

[24]

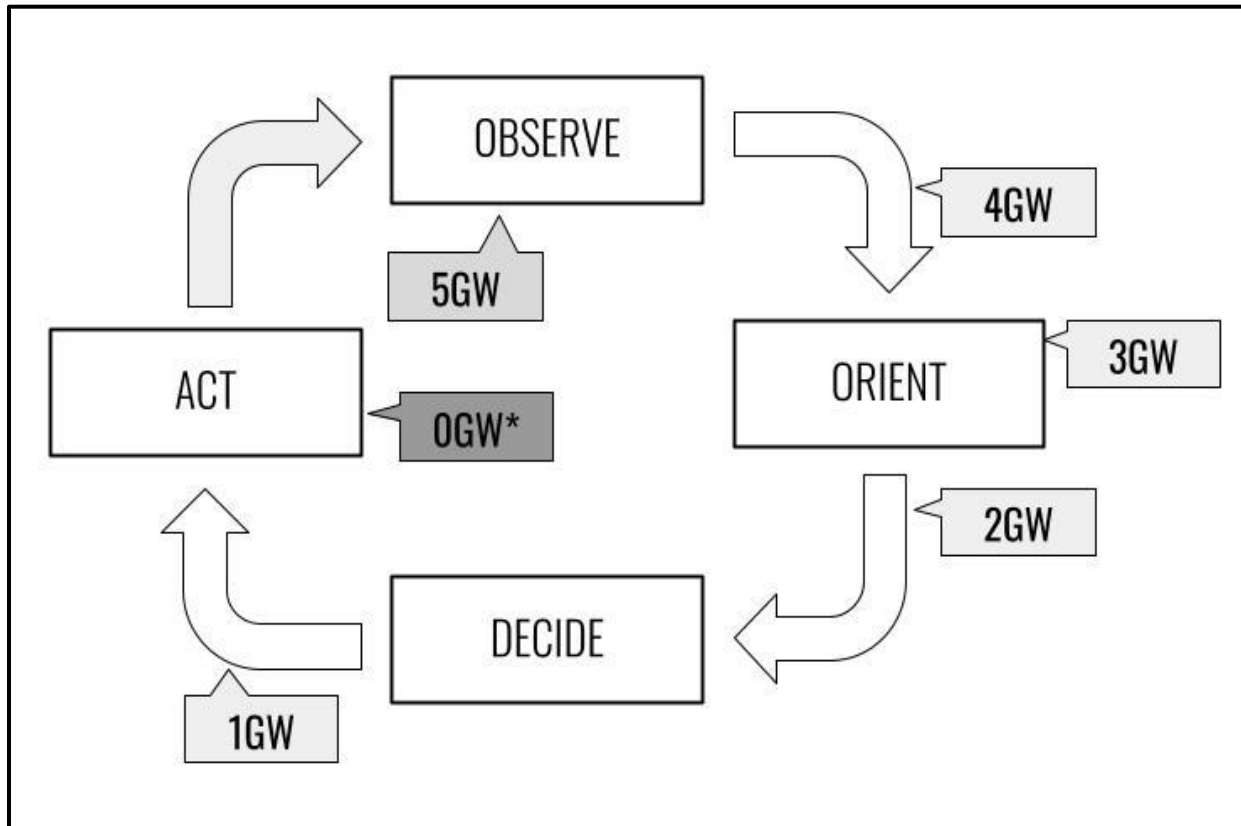


Figure 1. Abbott's Gradients of Warfare projected onto John Boyd's OODA loop. Adapted from [25]. 0GW not included in original figure.

## Linn's Model of Strategic Narrative

Breaking rank from chronologically or technology driven models of war, Linn offers a heuristic model of approaches to modeling war and the narratives which accompany those approaches. He proposes three general, abstract narratives encoded into the theoretical groups which would hold them: guardians, heroes, and managers [26]. Guardians are those who model war primarily as a science that is "subject to laws and principles" which can offer the means to predict the consequences of specific policies. Heroes model war primarily as an art, dependent upon military genius, experience and training, morale, and discipline. The final group, managers, model war as a "logical outgrowth" of politics and economics, dependent on logistics, mobilization of resources, standardized and effective equipment, and the assignment of well-educated professionals.

## Quantitative Models of Conflict

The term quantitative models of conflict is used here to describe the models of conflict which sit in clear separation from qualitative and narrative models, attempting to frame conflict in terms of formalized mathematics and computational structure. Several of these models are summarized here.

### Lanchester Models

The Lanchester model is likely the earliest substantial quantitative model of warfare, being introduced in the early 1900s in the book *Aircraft in Warfare: The Dawn of the Fourth Arm* by Frederick Lanchester [27]. Lanchester introduced a series of quantitative rules, such as the N-squared law (“the measure of the total of fighting strength of a force will be the square of the sum of the square roots of the strengths of its individual units”), and differential equations to describe concepts like attritional dynamics and predict the likelihood of outcomes of engagements [27]. In addition, he used geometry to illustrate the resulting models of these equations in numerous examples across air, naval, and land warfare with consideration for various kinds of armament [27]. Though introduced in the early 20th Century, Lanchester models are still being adapted today to represent things such as force ratios and information importance in guerilla warfare and insurgencies [28] despite the model’s shortcomings in describing real-world dynamics [29].

### Fault Tree Analysis

Fault tree analysis was developed to decompose potential failure states of a system or operation into subevents to better understand potential for cascading failures [30]. Each of these subevents can be given probabilities and relationships with other events, allowing risk analysts to calculate the probability of compound events and specific outcomes [30,31]. Using fault tree analysis, conflicts can be modeled in terms of various system states and their likelihood to trigger undesired system states or cause cascading failures via complex threat surfaces [1].

### Effects Based Operations

Effects Based Operations (EBO) planning is a form of course of action planning for military operations which is characterized by its use of Bayesian graphical models (“Bayes nets”) and models of complex systems [32]. While EBO is primarily a planning tool, it embraces a systems warfare approach by modeling an area of operations as a series of components which may be acted on to generate effects which cascade throughout the system. As a consequence of this approach, conflict becomes more general and less weighted with connotations of violence, instead being better described as friction or disruption, making it particularly useful for planning within and describing gray zone and narrative warfare [32,33].

## DoDAF

The US Department of Defense Architecture Framework (DoDAF) and its variants are “military architecture” frameworks intended to improve planning, procurement, and the deployment of various military systems [34,35]. While it is not intended to model conflict explicitly, the DoDAF system incidentally generates a model of conflict consistent with Linn’s conception of a “Manager’s” view of war [26] as a consequence of its modeling of future military needs. Under this view, various kinds of conflict can be described and analyzed by modeling the resources, sub-organizations, missions, and logistics of a military organization itself as a system-of-systems interacting with constraints and limitations (e.g., adversaries and their military organization).

## Systems Warfare

Western network-centric warfare, Chinese systems confrontation warfare, and the Russian Gerasimov Doctrine are all examples of modern updates to military doctrine necessitated by the rise of gray zone warfare. Each focuses on permanent conflict, a fusion of hard and soft power across numerous domains, and describing war in terms of whole-of-system conflict over networks, such as those of influence (media) and exchange (supply chains and economies) [36–40]. While the details and documentation of modeling approaches for describing systems of interest within Chinese and Russian doctrine are not easily available [38], those used within network-centric warfare are extensive and often make use of agent-based, Bayesian, and complex system-of-systems modeling methodologies to describe and analyze the structure and risks of abstract conflicts [40–42].

## Models of Conflict Information Flows and Decision-Making

The preceding categories of conflict models focused on the historical and qualitative (Narrative Models of Conflict) and the quantitative and data-driven (Quantitative Models of Conflict). In this section, we describe models that have been developed with a behavioral focus, whether they take a qualitative or quantitative approach. These models of information flows are not just explanatory - they are used in national militaries to inform design and decision-making and as such, they have real impacts and need to accurately and appropriately describe systems [39]. Many information flow and decision-making models have been considered for use within national militaries, such as Shewhart’s Plan-Do-Check-Act (PDCA) model [43], Wohl’s Stimulus-Hypothesis-Option-Response (SHOR) model [43,44], and the Endsley model [43,45] (see Figure 2). However, two models in particular, the Observe-Orient-Decide-Act and Rumsfeld’s Triad of “Knowns,” have seen broader adoption and adaptation than others. Here, these two models are summarized.

## Observe–Orient–Decide–Act (OODA) Model

The Observe–Orient–Decide–Act loop (OODA) model is among the most familiar and commonly used decision-making frameworks in modern times and is used “ubiquitously throughout the branch-specific and Joint doctrinal publications of the US Military” [46]. While the OODA loop is now contained within a scholarly corpus, its creator, John Boyd, never directly published on the topic, instead choosing to share the ideas behind OODA primarily through his presentations [46–49].

The OODA loop was originally designed to help describe and inform real-time decision making by pilots, wherein a “pilot observes the variable and surrounding, orients the aircraft to an advantageous position... [decides] the following course of actions in order to engage” and then acts them out (see Figure 3) [50]. The generalizability and simplicity of this “loop” of factors in decision making led it to enjoy reasonably high levels of adoption, not just in the military, but also in areas such as business and healthcare [50]. However, this simplicity, paired with the lack of published clarifications and formalizations by Boyd, means that it is constantly being reinvented, reconsidered, reinterpreted, and modified to fit various situations leaving it lacking consistent definition and coherent development as a model that could further enhance its usefulness [43,50,51].

## Rumsfeld Matrix of Knowing

The Rumsfeld “Matrix” [52], “Paradox” [53], or “Quadrants” of knowing, was not initially formally proposed as a framework for action and perception, but rather was merely a response provided by Secretary of Defense Donald Rumsfeld to a question asked about the lack of evidence of weapons of mass destruction in Iraq:

*“Reports that say something hasn’t happened are always interesting to me, because as we know, there are known-knowns; there are things we know we know. We also know there are known-unknowns; that is to say we know there are some things we do not know. But there are also unknown-unknowns – the ones we don’t know we don’t know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.”*

[54]

Though Rumsfeld only offered 3 informational states in the direct quotation, the suggestion of known-knowns, known-unknowns, and unknown-unknowns implies a combinatorial requirement for an additional fourth state: unknown-knowns, which has led this framework to be referred to as “Rumsfeld’s Matrix” [55]. Interestingly, many analyses ignore the presence of this 4th implied category [53,56–59].

While other decision making and information flow frameworks discussed above focus on linear steps in the decision-making process itself, the Rumsfeld Matrix of known-knowns, known-unknowns, unknown-unknowns, and unknown-knowns is different. The matrix is invoked to help describe the static abstract information spaces and voids that decision makers must navigate and explore (see Figure 3) with gradients of greater or lesser information and lack of awareness of degrees of ignorance - a double hurdle to situational awareness.

Rumsfeld's strategic categorization has since been adopted as a rhetorical framework for considering information gathering and prioritizations in planning and decision making in the military and elsewhere. The Rumsfeld Matrix, like John Boyd's OODA loop, enjoys an informal rhetorical ubiquity - it is a popular reference across other fields, such as in science [59,60] and energy infrastructure [52].



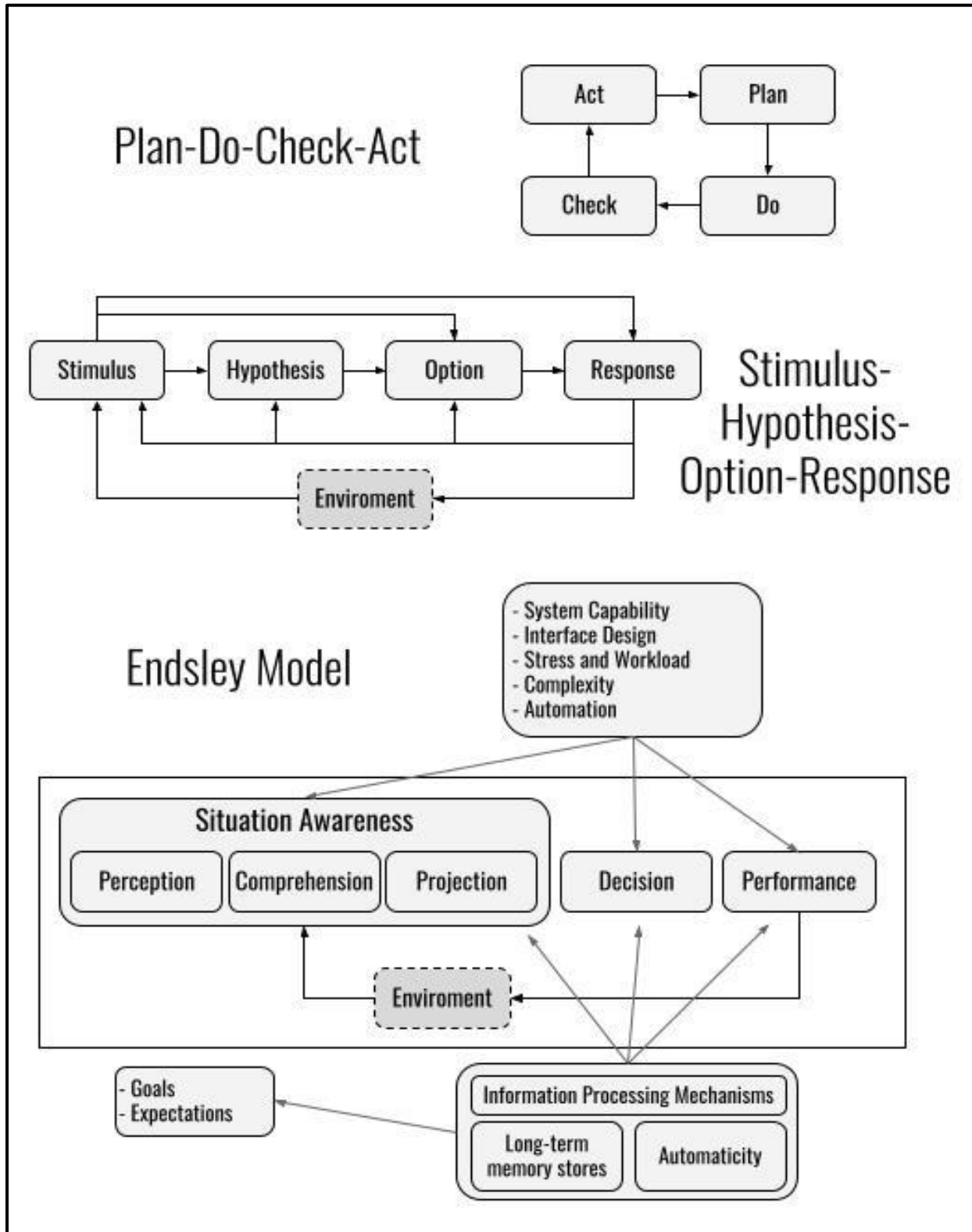
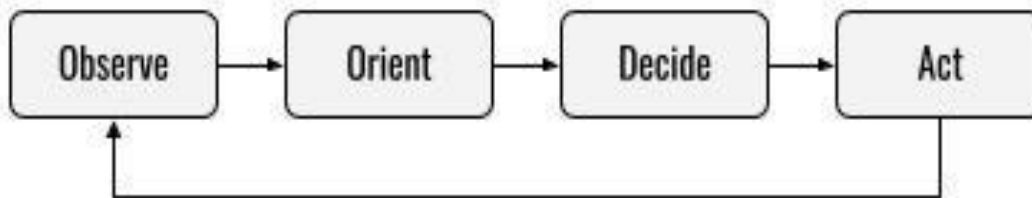


Figure 2. Various Decision-Making Models. Plan-Do-Check-Act Model from [43], Stimulus-Hypothesis-Option-Response from [44], Endsley Model from [45].

## Observe-Orient-Decide-Act (OODA)



## Rumsfeld Quadrants

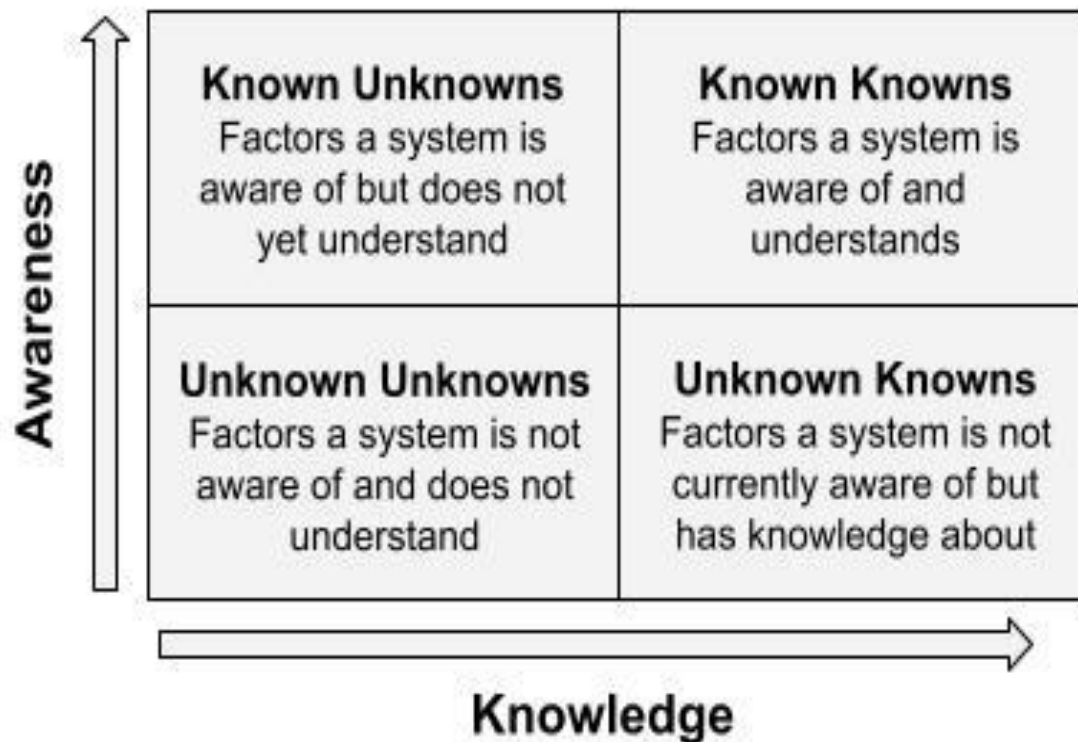


Figure 3. OODA and Rumsfeld Quadrants.

## Essential Features and Limitations of Past Conflict Models

This brief survey of conflict-oriented models used within military contexts reveals an arc of abstraction across time from simple pattern collection, to formalisms, and finally toward generalized models. The survey also reveals a persistent challenge through time of the problems of change management in the conduct of warfare (i.e., of inconsistency and adjustment to new paradigms and changed historical circumstances). While each of the models described had an important place in the history of the development of theory and within military scholarship, each also suffers from weaknesses which prevent it from offering sufficiently comprehensive predictive and descriptive power in the gray zone conflicts of the 21st century and beyond. However, each prior model has strengths and offers insights which should be captured by new models. Below, we consider some key insights to be preserved and brought forward from previous models. These insights will inform the AIC model introduced herein.

### Changing Expression of Conflict

Numerous models show signs of aging as the expression of conflict changes. As a first example, aspects of Clausewitz's trinity are still quoted as a basis for informing doctrine at the highest levels of the US Armed Forces [17] in a way which is consistent with Clausewitz's view of his theories as a "basis for study, not as doctrine" [15]. However, even when used in a limited way as a basis of study or theory, it still faces serious challenges in capturing significant aspects of modern conflict. While some argue many aspects of the trinity may be applied through analogy to asymmetric and low intensity conflict, the model may have to be somewhat contorted to be applied in many conflict scenarios; for example, in the conflicts between the Medellin Cartel and the Colombian Government [15]. Further, Clausewitz's trinity simply cannot explicitly or sufficiently describe the categories of conflict most relevant to modern organizations, such as narrative warfare and terrorism where many actors may be individuals motivated by ideology [14,32]. Even within defenses of the trinity model and of Clausewitz we find the suggestion that attempting to torture the model into explicitly describing aspects of modern conflict may be "profoundly confused" [61] and stem from the likelihood that Clausewitz "has been more often quoted than read and understood" [14]. While the underlying components of reason,

genius or strategy, and passion are still valid, the central tension, or as Clausewitz described it, the “balance between these three tendencies”, will no longer express itself in the same way and may need to be paired with other models in order to continue to provide value and insight [21].

While the Clausewitz trinity has seemingly received the most attention in terms of adaptation for the changing expression of war, approaches such as Lanchester models and Generations of Warfare, have also seen numerous adaptations in order to fit new paradigms. Replacements, such as models of conflict within the purview of network-centric warfare, fare far better in describing these new paradigms but might make a polemologist or military historian wonder if they describe old ones well. Even with the Generations and Revolutions of Warfare models, which are intended to capture the development of war historically, may unfortunately create a unidimensional or linear view of war as consistently developing in sophistication. Further, they place all conflict prior to the first millennium as “precursor activities”, creating a paradigm of study and thought similar to that which is found in “traditional Western historiography, in which all of prehistory — the bulk of the history of our species on earth — [is] consigned as an afterthought on the far left side of any historical diagram — the historical terra incognita before classical antiquity” [62].

It is important to consider how models built for new expressions of war might represent old ones given what is suggested by Abbott’s Gradients of War: that the expression of war may degrade in sophistication rather than increase linearly. There is a need to address how we represent conflicts within abstract space in order to capture not only the essence of previous and current expressions of warfare, but also to help project and consider what may come next.

### **Limited Interoperability**

The value of a model of a system might be derived not just from how well it handles updates to information about that system, but also from how well it interfaces with other models. How does a system reflecting one model come to “know” what is already “known” to a different model? For example, it would offer tremendous value via interoperability to be able to project

or map models onto each other. However, among all the models considered above, only limited capacity for backwards- or forwards-compatibility was found (the exception proving the rule was the mapping between OODA and the Gradients of Warfare in Figure 1). Though some models seem quite general, they have poor interoperability with others, for example, the value of computational systems such as those within EBO and Lanchester models is siloed from the insights within information flow and decision-making models. Though some work has been done elsewhere to map heuristics and narrative models to computational frameworks in gray zone and narrative warfare through the use of “pattern languages” [63], or collections of practice and risk heuristics which can be layered into EBO-like frameworks, it isn’t apparent that any substantial work has been done to generalize this approach to conflict in general [32].

Separate from attempts to map relationships among narrative models and their computational and informational counterparts, there is also significant dissonance within each of these categories. For example, Lanchester equations, by merit of their structure, cannot easily interface with EBO or systems warfare models. Further, within narrative models we find rampant disagreement on how to describe conflict in terms of priority. In addition, within informational models it is unclear how models like OODA can scale from local or single-actor tactical decision-making to strategic or multi-actor decision-making with adversaries in-the-loop as EBO or systems warfare models would indicate may be required. Inconsistencies or incompatibilities within and among models hinders the ability of applied composite models to provide superior insights into the origins and operations of human conflict.

There is a need for a computational integrative framework that connects tactical (micro) and strategic (macro) timescales, and builds on the strengths of narrative, quantitative, informational, and decision-making focused (meso) models. In the next sections, trends in the understanding of human interactions generally are brought to the challenges of analyzing human conflict, including war, and the synthesis introduces multiple new metrics of system performance from previously neglected contiguous domains of human behavior from which a richer,

and more extensible, computational model of human conflict and war emerges.

### **Generalization**

In addition to being able to handle updates to information and interface well with other models about a particular system, the potential value of a model might be further discerned based on how faithfully it is able to describe and integrate with other systems with one or more similar attributes. The history of conflict modeling, as illustrated in the summary of warfare literature above, reflects an ever-increasing awareness and integration of variables from the studies of interactions in conflict beyond those traditionally classified as “war.” As humans migrate their interactions from physical space to abstract online “information” space, the potential for integration of other knowledge about managing interactions and conflict in non-warfare contexts becomes increasingly relevant - and increasingly possible.

In fact, as the human species migrates an ever increasing portion of its interactions from physical interaction pathways to information-rich digital and online networks, the nature of conflict, including war as conflict, is changing. In traditional interactions and conflicts, the physical landscape and kinetic actions of stakeholders had the greatest influence on the models used to study those systems. In digital online information interactions, the “landscapes” are not physical, but instead are conceptual, narrative, and even memetic [64]. At one level, conceptual conflict might be seen as more amenable to dissipation without resort to irreversible destruction of rivalrous physical objects of value. On another level, abstract spaces lend themselves to myriad different simultaneous characterizations, each of which can provide pathways to conflict resolution, together or in combinations.

In the past, conflict might be explained with reference to people speaking different languages or seeking control of rivalrous physical territories. Increasingly, however, conflict can be described with reference to different paradigms, argots (trade languages), and risk concerns. Much as prior conflict might arise between speakers of different languages, so too might future conflict be analyzed as conflict between and among the

different languages, paradigms and interactions patterns of business, operating, legal, technical and social domains (BOLTS).

Since war is a subcategory of human conflict, BOLTS-based parsing can also help to introduce potential pathways to integration for models of nation state level conflict, including war. As the proportion of of conflicts between and among people, organizations and nations becomes less focused on violent physical conflict, it is increasingly better described as occurring over surfaces characterized using business [65,66], operations [67], legal [68,69], technical [67,70], and social [32,71] (BOLTS) components. As the case for traditional battlefields, the ability for modern models to capture both violent and nonviolent aspects of conflict at varied scales of organization in myriad contexts, digital and physical, becomes existentially important. BOLTS has become an approach to analyze this continuation of (information) warfare by other means.

While the popular models of conflict described thus far tend to focus on describing and providing insight into violent conflict, outside of the warfare-oriented corpus there is fortunately a rich history of models developed in an effort to understand and address non-violent, non-physical, or indirect conflict [72,73]. These traditional models of human conflict management are nonetheless non-traditional models of warfare. As warfare is migrating from physical to informational domains these non-traditional models present themselves as candidates for integration with traditional models of warfare.

Unfortunately, at first glance, these non-warfare models of conflict tend to appear to be focused on interpersonal and intragroup conflict, rather than inter-organizational or violent conflict, and some may explicitly avoid discussion of these topics [72,73]. However, within this corpus of non-warfare conflict work, concepts have been developed that can be helpfully brought to the study of war. For example, non-warfare conflict research includes research on negotiation and intragroup organizational conflict presenting concepts which are ripe for generalization to interorganizational business and legal contexts [73–75], research on task and process conflict directly applicable to understanding larger scale operations



frictions [73], and research on relational and diversity conflict which has already been applied to better understanding cultural and social frictions [72,73].

Other potentially useful non-warfare models of human conflict and its management include those models that analyze conflicts within a “commons”, which has its own storied computational and narrative corpus. Research on commons management focuses on conflicts which can arise in markets (both abstract and real) and the access to resources in which varied groups and actors have individual interests but collective ownership or stake [76,77]. For example, the oceans, the polar regions, the atmosphere, outer space, and non-earth heavenly bodies, are beyond the direct control of any nation, but provide resources and spaces in which nation states, and their resident citizens and companies, increasingly interact. In those spaces, conflicts of interests among stakeholders are bound to arise as competition for resources and conflicts of interactions emerge.

Elinor Ostrom won the Nobel Prize in Economic Sciences in 2009 [78] for her work in describing co-management regimes for addressing conflict in historical settings such as the conflicts that arise in the context of shared grazing and forestry resources, fisheries, and riparian (water) rights. Her work has been instrumental in the international management of fisheries and other resources in international waters, and for models of managing both outer space and knowledge space as well. Hess and Ostrom, in their book, *Understanding Knowledge as a Commons* [79] lay out eight principles for “robust, long-enduring, common-pool resource institutions”, which are:

- Clearly defined boundaries
- Rules that are well matched to local needs and conditions
- Individuals affected by these rules can participate in their modification
- The right of community members to devise their own rules is respected by external authorities
- A system for self-monitoring members’ behavior has been established
- A graduated system of sanctions is present



- Community members have access to low-cost conflict-resolution mechanisms
- Nested enterprises - the “appropriation, provision, monitoring and sanctioning, conflict resolution, and other governance activities” are organized in a “nested structure with multiple layers of activities”.

To help communicate the impact of these principles, Hess and Ostrom present the “Institutional Analysis and Development” (IAD) framework (see Figure 4). This framework presents a map of the relevant variables to the expression of friction, or conflict, within what it calls the “Action Arena” and represents a key example of a model comprised of elements which are generalizable to a great number of kinds of non-violent conflict. In addition, it makes use of narrative models regarding common “patterns of interaction”, such as “freeriding or misuse”, which can be layered into the model with probabilities and expectations about outcome, offering implications for how narrative models and pattern collections may be generalized to interface better with computational models.

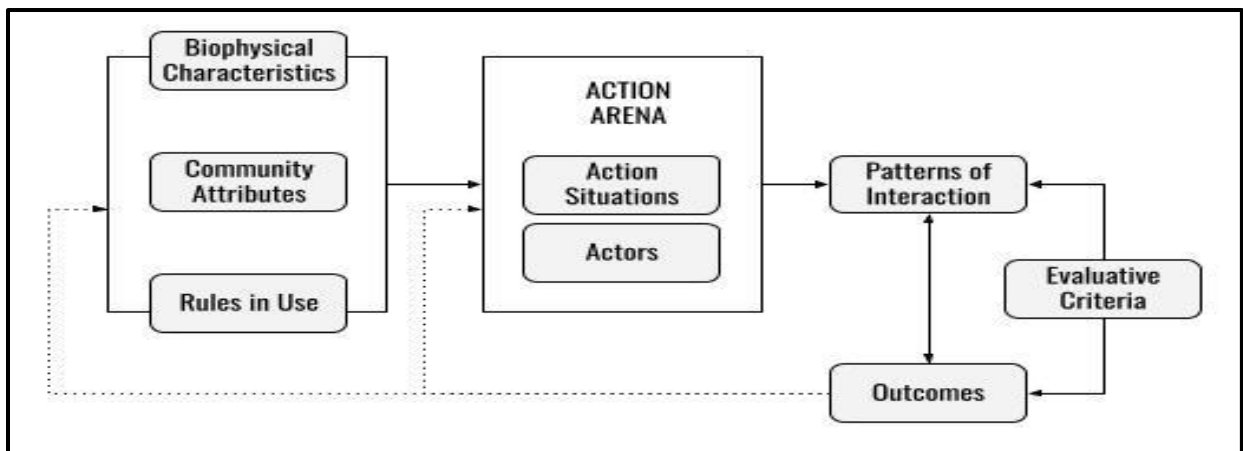


Figure 4. *Institutional Analysis and Development Framework, modified from [79]. Biophysical characteristics refer to ideas, artifacts, and facilities, the relevant factors which relate to the physical or quasi-physical affordances, boundaries, capacities, and limitations of a particular commons. The attributes of the community, refer not just to measurable qualities of the community, but also to those which comprise it, such as users, consumers, providers, and policymakers. Rules in use refer to administrative procedures, legislation, and contracts, as well as other activities considered to be pro forma even where they may not be codified or observable.*

With this discussion about models of warfare above, there appears to be a need to account for new frameworks that encompass modern expressions of conflict, are interoperable across domains, and generalize well enough to encompass peaceful and rapidly-changing times as well as classical forms of conflict and related

operations other than war (OOTW). An open challenge is for a computational model to capture the value and insights provided by various forms of previous narrative, quantitative, and information flow models of conflict. In the following sections we address this need by proposing a framework based on Active Inference. Active Inference is a framework arising from cognitive science, which has had demonstrable value in unifying certain aspects of cognition and sensemaking, and which may be used both computationally and qualitatively at different scales (e.g., single agent or multi-agent) [80–82]. The following sections present an overview of Active Inference, followed by its application towards the domain of conflict – the Active Inference Conflict (AIC) model.

## Active Inference Conflict Model

Here we propose a framework for modeling modern multiscale conflict, based upon an application of Active Inference (ActInf). ActInf is a behavioral modeling framework that integrates perception, cognition, and strategic action under a common currency – the reduction of expected free energy. As discussed below, expected free energy has several different compatible phrasings which facilitate its use in decision support in different systems and situations. Across these formal phrasings of free energy, a commonality is the emphasis on selecting actions that finesse both the epistemic (knowledge-oriented) and pragmatic (utility- or reward-oriented) aspects of action. Broadly, ActInf can be considered an approach that builds on quantitative approaches to action (e.g., cybernetics and control theory) with modern insights from cognitive sciences [83,84]. This action-oriented view casts the active sensing of systems as fundamentally about reduction of uncertainty. The sensemaking process goes wrong when inappropriate uncertainty-reducing behaviors are implemented, or the variability of the area of operations is too variable to be tracked effectively.

The Active Inference Conflict (AIC) model is an approach which unifies some aspects of previous models of conflict, and generalizes conflict in order to help capture business, operations, legal, technical, and social aspects relevant to modern gray zone warfare. Additionally, the AIC model has sufficient flexibility to be used both qualitatively or quantitatively across different timescales (e.g., tactical, strategic), structural scales (e.g., individuals, organizations, communities, and states), domains, and scenarios. Recently it has been suggested that autoethnographic organizational approaches (e.g., reflection upon one’s own experiences and surroundings) provide an amenable on-ramp to the ActInf framework [85]. Multiple informal and technical introductions to ActInf and the broader Free Energy Principle exist [81,86–90], here we introduce some of the salient features and descriptions of key terms within the ActInf framework which predisposes it towards effective application to the domain of conflict and for use within AIC.

From a military science perspective, AIC provides a bridge between single-agent real-time tactical decision-making models (such as OODA), and broader strategic analyses (such as those provided by the GW framework). As ActInf itself is a development on Bayesian graphical modeling to accommodate multi-level cognitive processes, the AIC model can be seen as the integration of this ActInf framework with other key existing models of conflict and models of cognition more broadly. Due to its descriptive bottom-up modeling approach, AIC also provides an avenue for integrating the analysis of military, non-military, and non-kinetic models of conflict (as well as cooperation, and other categories of interactions). Below, we provide a primer on ActInf with a focus on how key ideas are applied in the AIC model. Figure 5 summarizes the scope of AIC and Table 1 provides a map for the territory we explore in the following sections (the core terms and features of ActInf as deployed in AIC).

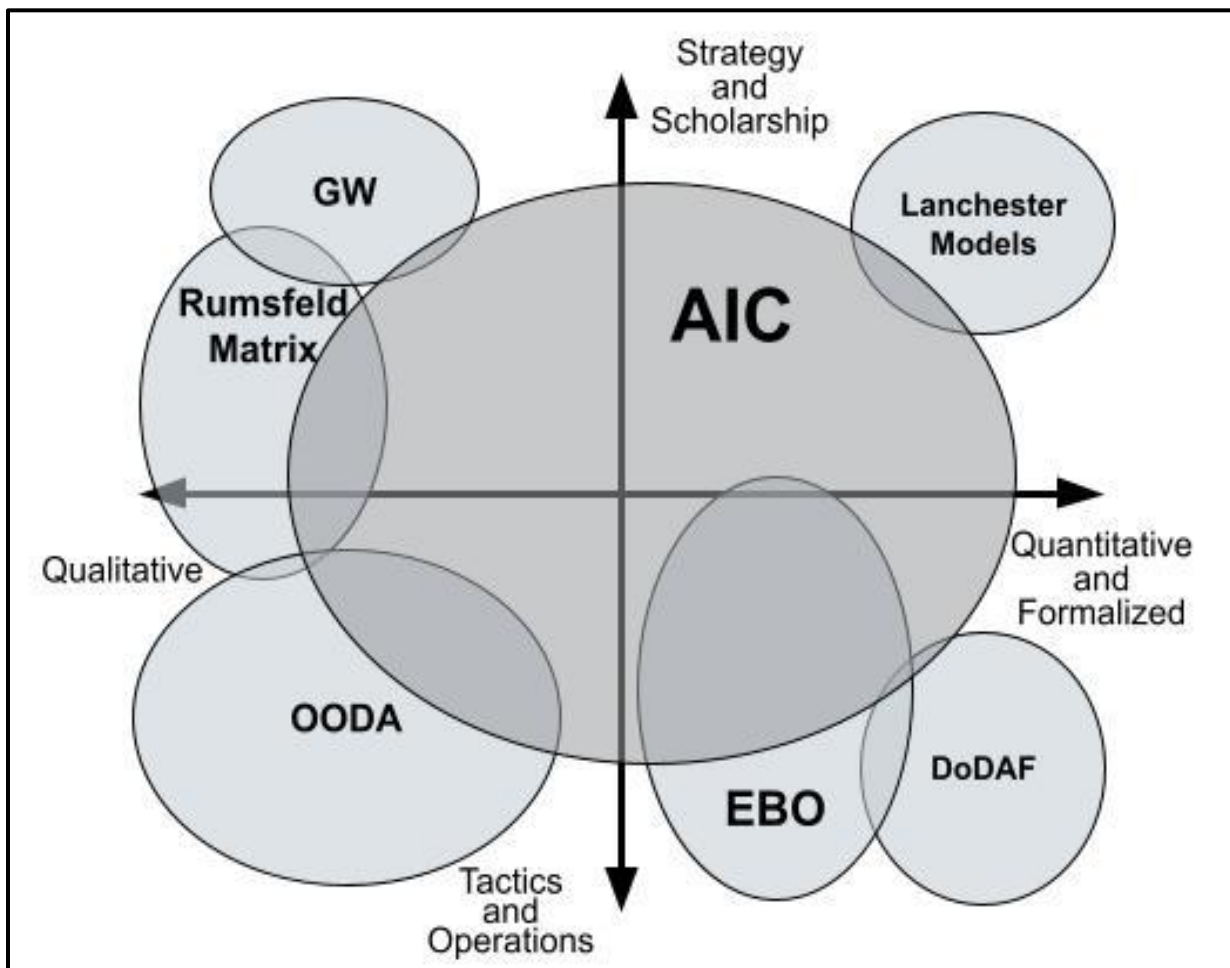


Figure 5. Scope of Active Inference Conflict (AIC) model along the dimensions of qualitative to quantitative (X-axis) and tactical to strategic scale (Y-axis). From the top-right and going clockwise: Lanchester models, DoDAF (Department of Defense Architecture Framework), EBO (Effects Based Operations), OODA (Observe-Orient-Decide-Act), the Rumsfeld Matrix, and Generations of Warfare (GW) model.

**Structuring the Information Commons**  
V – Active Inference in Modeling Conflict

| ActInf Core Terms              | Usage at Tactical scale in AIC model                                                                               | Usage at Strategic scale in AIC model                                                                                        | Tactical and Strategic scale in AIC model                                                                                                                                                                      |
|--------------------------------|--------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Entity                         | Human, human – with tech in the loop, squads, teams                                                                | State or non-state group                                                                                                     | Larger entities are made up of smaller entities                                                                                                                                                                |
| Generative Model               | Short-term expectations for a given scenario, enacted & embodied by a tactical agent                               | Long-term expectation at a diplomatic or geopolitical level, of a strategic agent                                            | Strategic-level generative models constrain/enable the function of tactical-level models.                                                                                                                      |
| Perception & Action            | Perception: Bodily senses and (meta)cognition.<br>Action: Physical movement including tools                        | Perception: Informational ingress, observations from e.g., markets, environment<br>Action: Communications, operations orders | Scales can interact and influence each other through Action (one scale/system's action is another's Perception)                                                                                                |
| Affordances & Policy Selection | Decision of which button to press, what to say, or which route to take                                             | Decision of which sanctions to apply, communications to release.                                                             | Large scale outcomes (movement of a legion) are jointly influenced by top-down and bottom-up implications and decisions                                                                                        |
| Expected Free Energy (FE)      | Implicit or explicit prediction over a time horizon of the uncertainty associated with a given sequence of actions | Implicit or explicit prediction over a time horizon of the uncertainty associated with a given sequence of actions           | Systems driven by tactical minimization of FE may not achieve strategic aims.<br>Strategic minimization of FE may entail novel regimes for tactical elements (e.g., waking up early, or experiencing surprise) |
| Action-Perception Loop         | Continuous flow of bodily sensory information and personal physical movements                                      | Continuous flow of organizational/informational inputs and outputs                                                           | Action loops of tactical entities are faster/smaller nested systems within strategic entities (like players on a soccer team)                                                                                  |

Table 1. Core terms in ActInf (left column) and their usage in AIC (right column)

## Active Inference Overview, Terms, and Features

There are several core features and relevant terms from ActInf that are necessary in communicating the AIC model (Table 1). Here we provide an overview of ActInf topics and terms, with an eye towards how the concept will be applied in the AIC model and the general implications for the term's quantitative and qualitative use.

### ActInf Terms

Here, the terms necessary for communicating the AIC model are described.

#### **ActInf Entity**

An ActInf entity is the minimal system description or model that is partitioned off as a separate (but interacting) thing from its environment or niche. The “thing-ness” of the system is specified by how relevant system variables are partitioned into several kinds of states. The scale of the entity might represent, for simulation and modeling purposes, anything from individuals to communities [91–93].

Some presentations and applications of ActInf differentiate two categories of Entities: “Mere” and “Adaptive” [94,95]. A “Mere” ActInf entity is one that passively synchronizes or reacts to external stimuli or causes. Relevant Mere ActInf entities in a model of conflict might include inanimate objects, smart contacts or blockchains, or any system with a well-defined, passive, or completely understood input-output relationship. In contrast an “Adaptive” ActInf entity is one that interacts with its environment in an embodied, agentic, anticipatory, cybernetic, and anti-dissipative fashion. Relevant Adaptive Entities in a model of conflict might include humans, teams, organizations, companies, countries, and non-state groups.

ActInf entities can be considered “generic” patterns that partition the statistical dependencies of agents into internal, external, and blanket (incoming: sense, and outgoing: action) states. This characterization of a generic entity type is useful for several reasons:

- ActInf entities have tractable interfaces for lateral interaction as well as nesting within other ActInf entities, allowing for modeling of complex heterarchical synthetic

intelligence, or macro-cognition and organizational behavior [3,80,96].

- So long as ActInf entities have action affordances which can interface with external entities and sense affordances which interface with external stimulus, the representation of their internal state and policies can be modified in any way appropriate for the nature of that entity and the simulation or modeling task at hand.
- Even without full quantitative integration, the process of framing a system in terms of its entities and nested entities can help illuminate its structure as exercise in system modeling and sensemaking [85].

### **Generative model**

The generative model of an ActInf entity refers to the ongoing creation by internal states of expectations, for example the states that the organism or organization expects itself to be in. Entity actions are selected in order to reduce uncertainty about the realization of those expectations, as the generative model includes expectations over sense, action, internal, and external states. In application across systems, the imperative for behavior in ActInf entities is not the maximization of reward but rather the reduction of uncertainty [97]. Reduction of uncertainty is always in reference to a specific generative model possessed or enacted by a system of interest, be it an organism or organization [3,92].

### **Perception & Action**

ActInf entities are continually engaged in perception and action. ActInf builds on the predictive processing, embodied cognitive frameworks, as well as other Bayesian and computational models of perception [98,99]. Perception is the ongoing process by which sensory observations are predicted or inferred by the generative model of an ActInf entity. Action refers to the enacted outcomes or outgoing statistical dependencies of the system, whether they are digital, social, financial, or physical.

### **Affordances & Policy Selection**

Policy selection, or action selection, is the process by which the entity will (act as if they) decide upon a course of actions (a policy). For a body, the action states might refer to the exact



angles of each joint, while the policy selection “to walk” could entail a complex sequence of changes to action states. The space of possible policies for an ActInf entity at a given time is known as their affordances (opportunities for action and interaction in the niche), drawing on the use of the term in ecological psychology [100]. Policy selection is carried out in light of a preferences over sensory observations (e.g., having a preference for warm temperatures over cold, and then acting to light a fire to reduce surprise about temperature). Thus policy selection is cast not in terms of finding highly-rewarding states, but rather inferring which option from a given limited set of affordances is expected to lead to the lowest expected difference between expectations and experience (lowest expected “free energy”) through time, in terms of pragmatic (utility) value as well as epistemic (uncertainty-reducing) value. When these expectations and preferences are for rewarding states, then ActInf models can recapitulate behaviors found in other kinds of reward-maximizers and reinforcement learners [81,97]. The selection of policy is in ActInf because entities can rapidly transition from utility-oriented behaviors to epistemic actions, as the flow of received information changes moment by moment.

### **Expected Free Energy**

This expected free energy quantity used for policy selection, can be variously framed as achieving evidence for a successful self, resistance to dissipation, or the general reduction of uncertainty [98,101]. Several useful mathematical decompositions and equivalences exist for expected free energy, for example energy minus entropy (similar to Gibbs free energy), surprise plus informational divergence, accuracy minus complexity (as used in Bayesian statistics and machine learning) [102]. Classical decision-making constructs such as expected utility, informational foraging, risk-sensitive policy inference, and optimal control are special cases or derivations of more general formulations of ActInf entity behavior [81,103].

### **Action-Perception Loop**

The action-perception loop in ActInf describes how Internal states (constituting the generative model of an entity) update in response to incoming sensory stimuli, and how actions (outgoing influences of the entity on the niche) define the

outcoming interfaces of the systems. This problem of real-time control occurs in the domain of robotics, public health, command and control systems, and elsewhere. To model these heterogeneous yet structurally-analogous scenarios with an ActInf entity, the entity can be modelled as a Partially Observable Markov Decision Process (POMDP) [88]. This POMDP specification is a Bayesian graphical model that lays out all variables required for minimal modeling of an ActInf agent (Figure 6). At each timestep of the POMDP model, the entity receives new observations from the niche, updates the parameters of its internal generative model, performs policy selection, then enacts an action consistent with the selected policy.

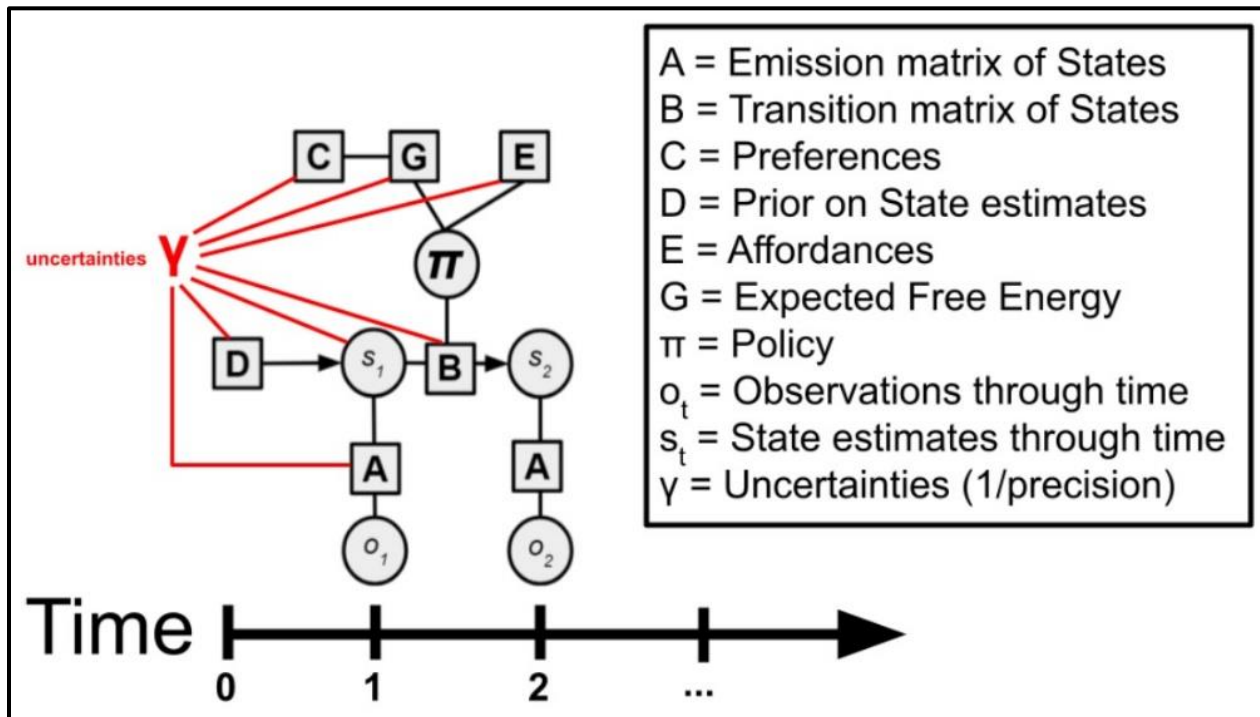


Figure 6. Partially Observable Markov Decision Process (POMDP) model of an ActInf entity.

## ActInf Features

The ActInf framework builds on the key terms towards several essential features. These components and generalized structures offer myriad affordances to researchers and analysts. Here we discuss several ActInf core components, placing them in the context of the AIC model as a formal model of interacting systems in conflict.



## Interactions with the Niche

Niche refers broadly to the surroundings or context of an entity, be it biological, social, or informational. The niche is the unobserved generative process that passes sensory observations to the entity (akin to how the location of the sun is not directly observed, but is instead inferred from the angle and type of impinging photons). ActInf entities interface with their niche through sense (incoming stimuli) and action (outgoing effects) states. Entity actions can modify their niche, reflected by changes in how the states of the niche change through time (for example tightening a screw so it doesn't wiggle in the future). This type of active co-construction between entities and their surroundings is known as niche construction or stigmergy [104]. This partitioning of the Internal, Action, and Sense states of the system of interest (the "particular states" [105]) entails that all features or data outside of the system of interest are external or niche states. We can consider the POMDP of the ActInf entity from Figure 5 as it interacts with its niche (Figure 6). The internal states of some system of interest can be modeled such that the external states provide observations ( $ot$ ) to the entity, and the selection of policies ( $\pi$ ) are upstream of the enactment of action state.

## Interacting Entities

This same ActInf framework can apply whether the external states (external from the point of view or partitioning of the entity) are of a very different kind than the entity (e.g., an ant colony inferring a raincloud) or a similar kind (e.g., two humans and their mental models of each other). Interacting entities can select policies with long-term expectation of net-positive interactions (e.g., trusted interactions from a game theory perspective), and this framing can suggest the formation of new kinds of organizations. The concept of Thinking Through Other Minds (TTOM) describes how the internal general model includes each Entity's own actions as well as the actions of the partner [106,107].

## N-Dimensional Modeling of Abstract Space

The advantage of a domain-flexible description of entities and their interactions, is that it facilitates the modeling of high-dimension interaction spaces, and detection of patterns across different interfaces or types of observations across BOLTS

surfaces in way that may be considered analogous to the integration of different kinds of neuroimaging data (fMRI, EEG, and MEG) in the Statistical Parametric Mapping (SPM) framework [108]. General ActInf modeling, along the lines of complex systems models described above, can capture the dynamics of classical cooperation/conflict situations as well as extend to model heterogeneous, unconventional, and yet-unseen mechanisms and patterns. With the use of an event reporting framework, this ability to capture cooperation and conflict across myriad surfaces may help to identify not just yet-unseen mechanisms and patterns, but also difficult to detect opportunities for strategic attention and action [109,110].

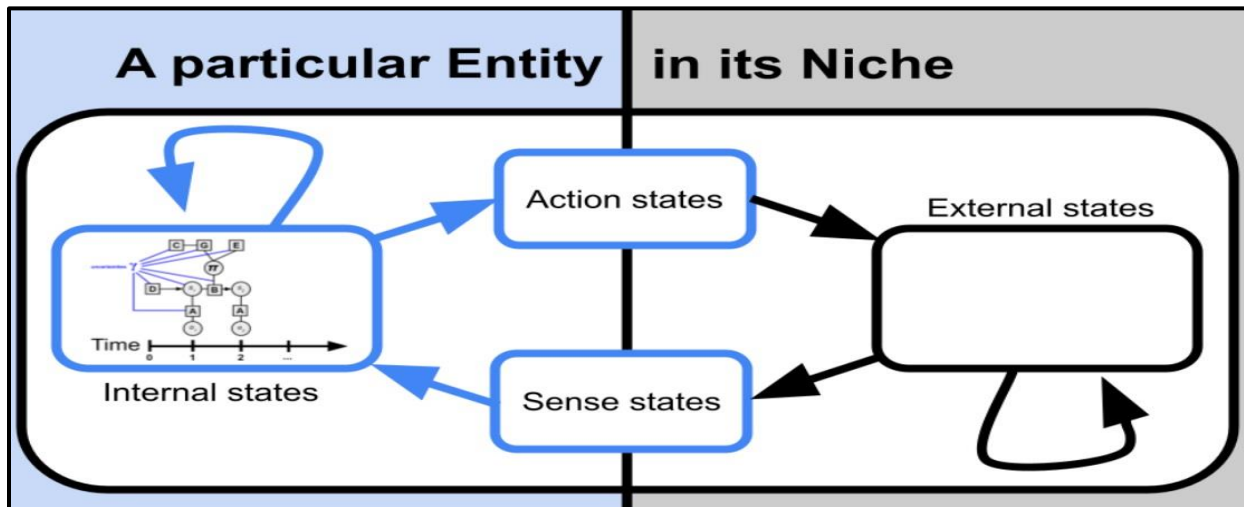


Figure 7. ActInf entity interfacing with external states. At right, external states are influenced by entity action states, and also external states may have endogenous dynamics. External states pass observations to internal states via entity sense states. Uncertainty in the flow of incoming sensory observation can be reduced through updating the internal model of the entity (learning) and action.

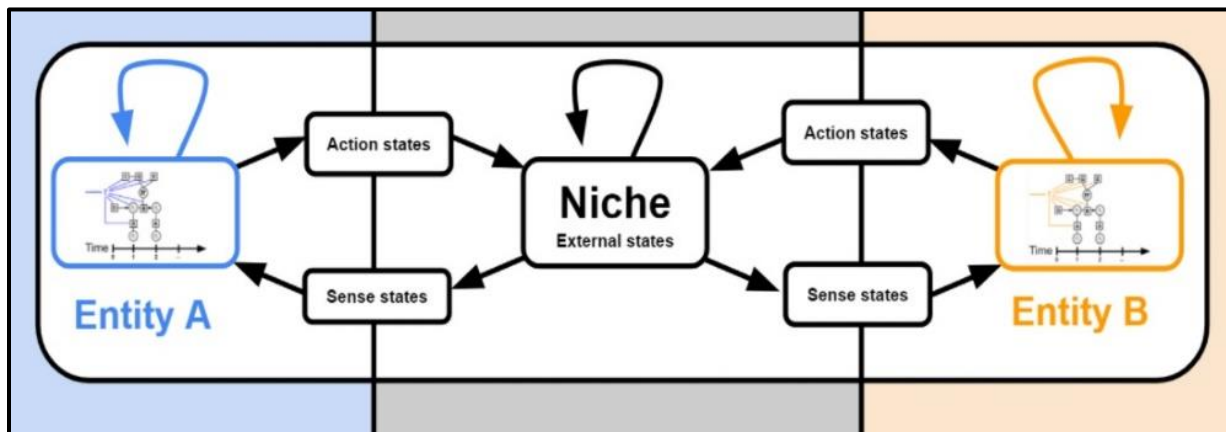


Figure 8. Two ActInf entities A and B, interacting via a shared niche (ecological, informational, or otherwise). The generative process of the niche is influenced by endogenous dynamics as well as actions from both entities.

## Use of the AIC Model

Here we build on fundamentals and recent applications of the ActInf framework to work towards new models of systems in managed and unmanaged conflict, cooperation, and every sort of human and institutional interaction in between.

### Entity Action Loop and Alignment with OODA

To understand the cycle of inferences and actions entailed by each timestep for an ActInf entity, it is helpful to consider this ActInf model and POMDP specification alongside the stages of the OODA model (discussed above). In contrast with OODA, the ActInf framework provides a model for “regimes of attention” [111,112], niche modification, and long-range/predictive/anticipatory policy selection in deep or nested generative models.

In both OODA and ActInf, the perception-cognition-action cycle is continuously unfolding, and can be thought of as beginning with the perception of new observations. Here we align ActInf terms and framings with the OODA sequence, with reference to Figure 9.

**Observe:** incoming observations (o) are received by sensors, sense organs, measuring tools, or other signal channels.

**Orient:** These observations are integrated with prior beliefs (D) about hidden causes or states of the world (s) through the bidirectional Bayesian mapping (e.g., constituting a generative model and recognition model) of the matrix (A) connecting observations to hidden states.

**Decide:** The updated Internal generative model of hidden states is used to perform inference on action, akin to other cybernetic or control theoretic framings. This selection of policy proceeds by the integration of preferences over outcomes (C) and constraints over action possibilities (E) in the calculation of expected free energy (G) in terms of pragmatic and epistemic value, as conditioned on different possible policies.

**Act:** Having selected the policy with the lowest expected free energy over the time frame of analysis, action states are updated.

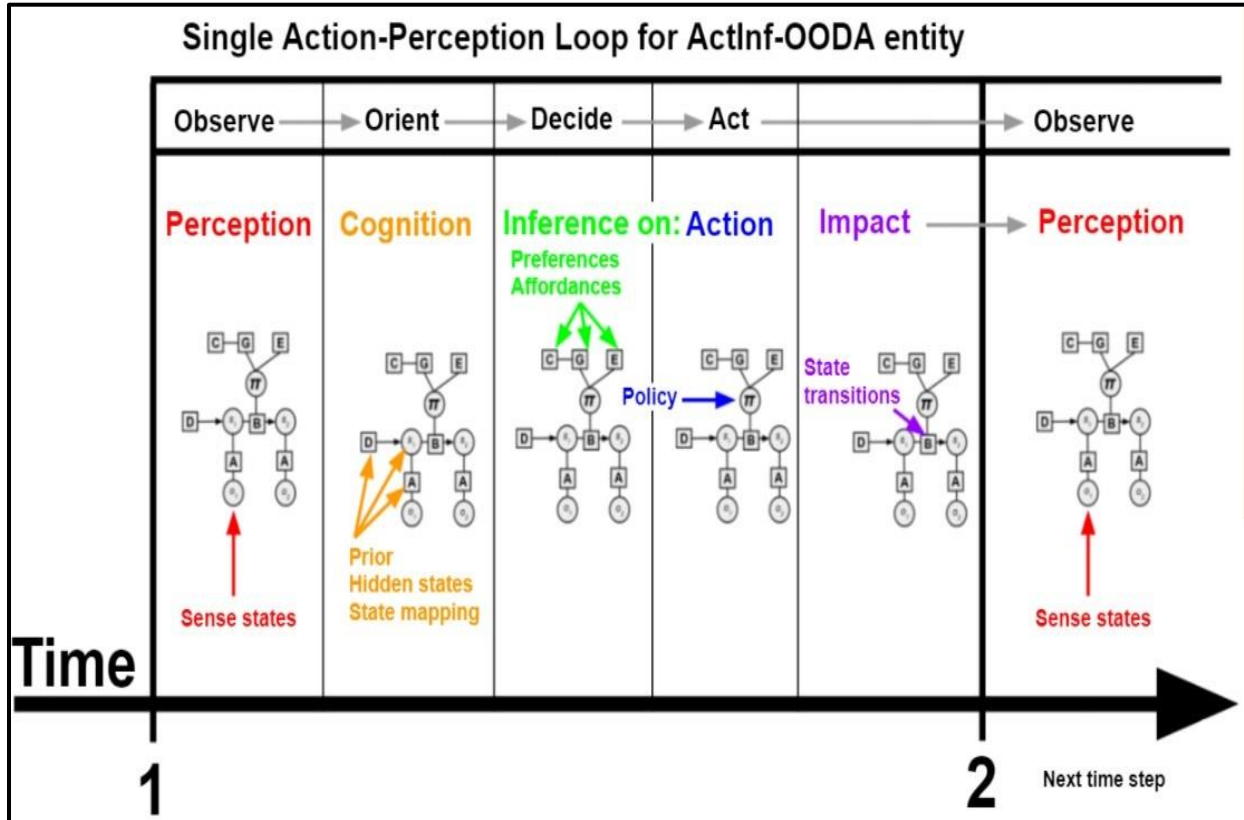


Figure 9. Comparison of Action-Perception loops for ActInf and OODA entities

## Unifying Quantitative and Formal Models of Conflict

The AIC model does not replace prior quantitative models of conflict, it instead integrates them and offers a new medium for their expression (as well as a new environment for testing and formal development). For example, given that AIC can be nested into and applied in agent-based models [80,113,114], methods such as game theory matrices and Lanchester equations can be calculated at snapshots and be used to predict and project the outcomes of simulations and iterated games - as well as test other formalizations and counterfactuals. AIC isn't limited to integration with agent-based models, it can also plausibly be nested into EBO and network-centric warfare graphs and planning cycles. Additionally, given that ActInf is a development on Bayesian graphical network methodologies, AIC itself, without any integrations, can be represented as a graph akin to those found in other graph-based models. Further, it can extend these quantitative and formal approaches (EBO for cognitive effects) or provide a surface for interoperability between them (e.g., Lanchester variations for both infantry- and artillery-driven conflict within the same larger model) in myriad conflict settings.

## Moving Beyond Generations of Warfare

The AIC model has the capacity to model structurally flexible, nested, and interacting entities and embedded decision-making processes. This allows for standardized and formal representation of conflict, whether it be between ant colonies or nation states, or between ant colonies and nation states. This formal representation allows n-dimensional measures of features and organizations within historical conflicts and thus opens the door to methodologies such as component factor analysis, which can allow for classifications and archetypes of conflicts that aren't limited by their place in history or by their placement on a single dimension. The analysis provided by AIC does not necessarily render previous narrative models of conflict classification obsolete - instead, it may offer opportunities to support and extend, and offer more insight into the similarities between these various models (for example returning to Flavius Vegetius Renatus' aphorisms discussing estimation, uncertainty, and expectation). In this same vein, AIC can be used to generate new narrative models akin to Generations of Warfare, as war evolves and adapts along numerous axes - for example, along axes such as the relative distribution of decision-making or the growth of intelligence requirements.

The decisions that are made today in this period of rapid transition will affect human conflict for many years. In this regard, AIC offers a potentially useful paradigm that can be extended, beyond the Generations of Warfare Model, into the past, anchoring it as a potential analytic tool to help predict efficient and effective strategies for future conflict analysis and resolution at multiple scales.

## Modeling and Discovering BOLTS Conflict

As discussed, modern conflict is coming to be better characterized as occurring over surfaces with combinations of conflict measurement and risk mitigation structures drawn from multiple, previously-isolated domains. In this paper, we have applied the rhetorical mnemonic device "BOLTS" to invite simultaneous consideration of multiple separate paradigms, measurements, and languages to a given conflict use case. The analytical parsing encouraged by BOLTS is one of many possible mechanisms for such a multi-faceted analysis, and is useful because the individual B-O-L-T-S components are broadly familiar, and the conflicts among the silos (e.g., technological vs. legal considerations of data use, business vs. social goals of online social networks) are well known - even if they remain unresolved. The business, operations, legal, technical, and social components therefore provide a familiar backdrop against which AIC can be rendered more accessible. The visual integration of AIC with BOLTS is shown in Figure 10. Below, we note examples which emphasize each of these aspects and consider AIC's use in these settings.

## Business

Business and economic relationships have always influenced human interactions from the earliest agoras to today's global online markets. The emphasis on metrics is driven by systems of risk mitigation and leverage associated with such business phenomena as production, resource accumulation, monetization, zero-trust trading, remote dealing, financialization, and myriad other "Business" concerns. Consider, for example, the many current structural global conflict surfaces that can be fruitfully analyzed as artifacts of the long-term implications of once-admired cost cutting strategies (such as foreign production of domestic goods) associated with the historical transition from physical to information dependencies. For example, the domination of China in manufacturing (and the consequent dependencies of consumer societies such as the US) is a product of US companies seeking lower labor costs (and compliance with environmental, labor, and other domestic laws) in the past decades. The US became dependent on information and finance to maintain access to and control of such remote production activities, creating a period of relative order (in terms of environment and labor gains within the US), but deepening the dependencies on access to foreign labor and production apparatus - which creates disadvantages for the US in the event of conflict with China affecting trade. AIC can be applied to analyze, consider and identify developing price leverage within larger business and economic structures and their relationship to economic policy, or to help infer internal model or policy of adversaries (based on their policy "pings"), and can also be of use in identifying de facto adversaries that may not have coherent structure under the law or be otherwise be detectable through standard business or legal metrics (e.g., informal consortium-like entities, such as a category of businesses operating within a common niche, nascent cartels, mutually-dependent trading arrangements, online Distributed Autonomous Organization [DAO] structures). Looking at Business interactions through an AIC lens helps to reveal existing and potential interactions and their respective threats, vulnerabilities, and opportunities for new value creation, which will drive innovation in multilateral risk mitigating structures and in business entrepreneurship and innovation.

## Operations

The concept of “Operations” in BOLTS overlaps with other BOLTS notions, but its separate consideration yields novel insights into conflict, particularly when brought together with the AIC model. Operations includes concepts such as supply chains, scaling of operations, organizational change management, operating efficiencies, human resources, and a host of other notions of human organizations that reflect attempts by humans to manage conflict for rule-driven behaviors across interactions at arbitrarily-large scales. In these contexts, the AIC model provides a coherent and comprehensive lens through which to analyze “operations in conflict.” For example, consider that many current “supply chain” related conflicts and challenges are a result, in part, of “just-in-time” manufacturing, lean inventories, and other less-capital-intensive forms of doing business ushered in by the enthusiasm for outsourcing in the mid-1980s, and accelerated by the “bricks-to-bits” commercial information revolution. Those trends have continued and been accelerated by the overall migration from physical to information-based interactions and transactions. Consider that there is a large and still growing set of operations protocols that eliminate the need for organizations to maintain large and expensive inventories. The continuing advances of the information revolution allow the virtualization of internal supply chains and of the provision of access to parts, ingredients and subassemblies when as needed further disintermediating previous supply chain interactions - which changes can lead to conflict. With respect to the labor element of operations, the “outsourcing” of labor to other, less regulated, countries is also a part of this cost-cutting effort. The modern expression of this outsourcing is found in innovations such as eBay, UBER, or Lyft where the value steps in the management and structure of inventory and service provision, routing, and delivery are becoming increasingly separated. AIC can be used to model the structure and distribution of decision-making processes both in BOLTS and traditional conflict arenas and developing points of affordance and access leverage in relation to policy. Further, it allows for the modeling of operational niche and the processes and protocols associated with managing the potential conflicts within a given niche.



## Legal

The laws of physics are universal, but the laws of people are not. The technology of the Internet is based on physics, but the regulation of the internet is not based on the laws of physics. The result of all this is that the Internet has the potential to be deployed globally (and beyond) with technical standards, but the laws of the 195 sovereign countries which are not globally standardized, creates conflict. Of course, it is not just the laws and regulations themselves that are in disputes, but also the interactions of the billions of individuals and organizations acting every hour of every day under such laws. The legal focus is fruitful in measuring and managing conflict since that is the intended effect of all legal systems. However, non-legal conflicts, such as political, economic, social, cultural, aesthetic, and other non-legal interactions, are beyond the reach of the risk mitigating help of legal systems. AIC applied with BOLTS can help to bridge the gap by bringing legal forms of conflict management into closer contact and interoperability with other BOLTS forms. In addition, legal confrontations in civil, criminal, and international disputes are in and of themselves conflicts which can be modeled by AIC. However, law is not just a source of conflict mitigation - it is increasingly a source of agenda-laden conflict engagement. Consider that beyond its role in helping to resolve individual conflicts, confrontations that apply law as a sword (and not just as a shield) are increasingly becoming a chosen avenue for conducting gray zone conflict and disruption between and among nation states and other entities. In the case of nation states, each of which as a sovereign can, by definition, create its own laws, legal warfare or "lawfare" [68,115,116] can be said to be composed of the development, amendment, and mobilization of "domestic and international laws" for geopolitical and military gain [117]. These forms of aggression are not typically characterized as "war," but rather in such forms as trade negotiations, immigration policies, tax and financial regulations, bilateral treaty negotiations, regional pacts, cartel arrangements and other similar forms. The development of legal standards for the protection of statutory and contractual rights, the enforcement of legal duties and the reliance on predictable legal processes when exploited as a means of deterring, binding, and protecting individual and organizational interests' actions in conflict with others is often difficult to detect in the churning and dynamic

landscape of legal conflict. While legal notions such as “abuse of process” are intended to curb excessive and socially-destructive application of law as a sword, the subjective and contextual aspects of legal forms of conflict can obscure root causes and intentions of conflict in many cases. AIC, with its affordances for modeling and inferring internal models and policy, could be of use in classifying and detecting patterns of legal actions and consequent leverage within myriad interaction niches in order to more effectively measure, moderate, and manage legal conflict affordances at tactical, campaign or theater, and strategic levels.

### **Technical**

Technical infrastructure, standards, and protocol are bounded by both computational and legal rules. The dynamic technical edge between these two areas is of particular importance for the future of conflict as human attention turns from a focus on data secrecy as a basis for conflict mitigation strategies, toward a focus on information integrity as a pathway to reducing information risk and interaction conflict.

Data plus meaning yields information. Data security is necessary, but insufficient, to yield information reliability and distributed security. “Meaning” security is needed to complement data security to manage information network conflicts. While data security is the focus of technical features of the Internet and modern computer science, “meaning” security is the focus of law. Consider that all contracts and laws can be viewed as enforceable “stories” about the past, present, and future. When those stories are agreed upon and acted upon, they de-risk future interactions in ways that no one person can achieve by themselves (for example the laws and technical specifications that interact to de-risk otherwise hazardous situations such as highways and exchanges). Such enforceable stories are the way that humans achieve “meaning security.” Contracts and laws are all promises that we make to ourselves and others about the future, and the law is a mechanism to test our performance against those agreed upon parameters. In this way, it is not unlike technical specifications that set rules of general application for the technical performance and behaviors of engineered systems.

As the desire for verifiable information integrity supersedes yesterday's satisfaction with data security, the human and organizational components of systems will be increasingly recognized as critical system components, not just as users of systems. Legal and technical paradigms are tightly intertwined in information systems, where technical specifications help assure data system integrity and legal rules help assure meaning system integrity, with the result of enhanced information system integrity. Such "tools and rules" leveraging will be accelerated through application of AIC framings that will quickly reveal the potential alignments of such systems. Such analyses will be critical to the advancement of various information integrity structures to help manage the conflicts that are bound to arise through the introduction of such new distributed information integrity structures as decentralized management of intellectual property, the introduction of digital "twins", smart contracts, computer-aided governance, and the progression of data privacy- and information integrity-related legal structures.

Emerging interaction structures provide a sense of the challenges and opportunities that reveal themselves at the intersection of technical and legal interaction and conflict management use cases. Historically, notions of intellectual property law (involving copyright, patent, trademark, certification mark, and trade secret) have always blurred the boundaries between physical and intangible value of goods and services in commerce. In terms of decentralized management of intellectual property, consider that nation states and the Westphalian system are based on physical boundaries and borders, hence the exclusivity (rivalrousness) of ownership of real property (e.g., land). At its base, the Westphalian paradigm of enclosure and exclusive jurisdiction may be fundamentally inconsistent with the infinite duplication that is possible with information. This may mean torturing new technical expressions of intellectual property to fit this previous legal, business, and operations paradigms, for example through primarily interpreting and designing non-fungible tokens (NFTs) as an expression of ownership of a given represented object (e.g., a particular artistic image), or by developing new systems which acknowledge these changes, for example through primarily interpreting and designing NFTs as an expression of rights, stake, and affordances related to some given

represented object. In terms of digital twins, the notion of the identity entanglement between the referent human and their digital extension, as well as tangible and intangible property and their digital extensions (e.g., NFTs), introduces just one category of many fundamental shifts ushered in by the transition from physical to digital worlds - similar in potential impact to the introduction of corporate depersonalization or personhood, or of nation states themselves. Further, consider the introduction of decentralized autonomous organizations (DAOs) which may be composed of both human and adaptive autonomous entities and what this means for legal accountability, internationally and domestically. The legal handling of these transitions is thoroughly non-trivial - as one path might lead to serious implications for nation states and the foundation of their sovereignty (e.g., no one can force or coerce a public blockchain to grant and revoke an affordance) while another might lead to a substantially more powerful, and consequently, dangerous foundation for sovereignty (e.g., governments able to computationally force or bar interaction in a digital-focused society).

### **Social**

Simulation and modeling of narrative and social conflict can be notably difficult due to underlying challenges in accurately characterizing and modeling situational features that are relevant for ActInf agents [32]. AIC's nested ActInf entities and their affordances for flexible representation of internal models and policy offers a common avenue for various extant and new approaches in representing ideological, psychological, narrative, and memetic conflict, as well as deterrence. Recently various models of dyadic and collective social interactions have been implemented using ActInf entities [112,118–120], suggesting a strong possibility for these kinds of models to be deployed in the case of conflict. The implications of using AIC in work on cognitive security and narrative management is discussed further in the discussion of modeling cognitive security.

## **Modeling Cognitive Security**

Cognitive security (COGSEC) here refers to the study, development, and implementation of “practices, methodologies, and efforts made to defend against social engineering attempts - intentional and unintentional manipulations of and

disruptions to cognition and sensemaking” [121]. COGSEC is difficult to measure and model for the same reason simulation and modeling of narrative and social conflict is - there are distinct, underlying challenges in representing and predicting the effects and attributes of internal states. AIC, as previously stated, offers opportunities for representing internal states of entities in relation to external conflicts, emphasizing impacts on cognition and sensemaking. However, AIC’s potential uses in the study of COGSEC go further: recent work on scripts and context-driven reflexes in ActInf [119] rely on the same structure and methodologies as AIC and have great potential in being applied to better understanding relevant threat surfaces, given that so much of the threat surfaces relevant to COGSEC and social engineering are related to development and exploitation of reflex for both offensive and defensive purposes [122]. COGSEC methodologies found in social engineering and counter-deception literature could be simulated and considered using AIC, to better model and measure COGSEC and also consider how traditional methods such as the “reduction of the complexity of problems, introduction of routine and bureaucratic procedures, the choosing of satisfactory solutions rather than optimal ones, [and] giving preference to partial solutions at the expense of comprehensive ones and avoidance of new problems”, and more recent approaches such as narrative information management [123], common vulnerabilities and exploits (CVE) databasing of narrative influence techniques [32], and engagement with narrative content [64,124] might affect state and expression of COGSEC in a variety of entities.

# Implications from Use: Future Information Structures and Rumsfeld's Neglected Quadrant

Usage of AIC to represent modern conflict and the BOLTS structures which interact within it provides insights beyond the projection of winners and losers in iterated games. Of particular interest are implications regarding the nature of the BOLTS structures themselves and the prioritization of their objectives in the reduction of uncertainty in their niche. Here we consider these implications before concluding and offering recommendations for future technology development.

One of the implications of the move of the human species from physical toward information-based interaction landscapes is the reduction in efficacy and relevance of those historical institutions that provided reliability and protection for humans in physical spaces. As conflict becomes more abstract and less obvious, these traditional institutions are revealing their lack of fitness for governing in non-physical domains. While physical existence still precedes and is prerequisite for the achievement of other states and satisfaction of other needs, as reflected in Maslow's hierarchy of needs [125], human interactions will continue to be increasingly dependent on the information landscapes in which nation states, and other organizational structures, are struggling to replicate the status quo. This struggle of legacy institutions to understand and manage conflict in an inherently incompatible information landscape, is forcing individuals to seek alternative structures of risk reduction to help them navigate.

Using AIC as a qualitative lens renders all conflict as a form of information generation for the participating agents, with violent conflict constituting a "costly ping". In the past, the information generated from conflict might have been found in the numerous post-mortems and experience-informed treatises after campaigns [26] or in what could be called proactive intelligence, information about the enemy assembled after engagements [126] - however, now that conflict is increasingly situated in the information landscape and that the underlying "assets" and "territories" that are the objects of social, political, economic, and legal attention have shifted from physical emphasis to information emphasis, new structures are offered the opportunity for unparalleled management, monitoring, and facilitation of conflict. As well as the opportunity to define, via BOLTS norms, rules, and infrastructure, how conflict can be approached and resolved. AIC may be of use in both the design and implementation stages in these pursuits, and can provide alternative pathways that can be applied in those settings.

Another consequence of this move from physical to information emphasis is the non-rivalrous nature of informational assets. Physical property (whether real estate or tangible personal property) is rivalrous - its use and enjoyment cannot be simultaneously and exclusively enjoyed by multiple parties. Territorial expansion and the plunder of property reveal the rivalrousness of historical nation state conflict. In terms of digital materials - it is possible for two people to enjoy the use of the same software simultaneously, to read the same book, to watch the same movie, or to access the same data for different uses in different contexts without diminishment of the use and enjoyment of another. Further, physical assets are generally scarce and increase in scarcity over time - whereas the amount and complexity of information which can be generated as well as the rate of its growth is infinite. Both are expanding rapidly and creating structural hurdles to both individual and organizational situational awareness - the ability for organizations to manage this information effectively is strained [123].

Using Rumsfeld's Quadrants, which frame the information spaces and voids of value to organizations, as a lens over conflict both between organizations and between organizations and abstract phenomena (e.g., "war" on cancer, drugs, COVID-19), highlights Rumsfeld's neglected quadrant, "unknown-knowns". Further, it suggests that this neglected quadrant is a doorway from the static to the dynamic perspective on knowledge systems. The first three quadrants are described from the perspective of a centralized hierarchical party or bureaucracy - things are either known or not to that party, without reference to interaction with other parties that might alter the status of knowns and unknowns. On the other hand, this neglected quadrant appears to be a paradox: How can a given party not know a given "known"?

For any individual ActInf entity, an unknown-known appears to be an impossibility - its known-knowns and known-unknowns are accessible within its internal state and its unknown-unknowns represent relevant voids within its internal state that it does not yet identify as such - which begs the question: Where is there room for an unknown-known? The AIC model helps to formalize several situations in which unknown-knowns exist:

### **Known but Inaccessible**

An ActInf entity may hold relevant information that goes unused in policy formulation as a result of it not being immediately accessible.

### **Failure of Curation**

An ActInf entity may hold relevant information that is technically accessible but goes unused because of poor cues or the absence of indications of relevance.



### **Back Turning**

An ActInf entity may ignore relevant information because it may contribute to policy formation which conflicts with some other existing policy, prior belief, or contextual model.

### **Selective Disclosure**

An ActInf entity may have information that is accessible but will not access it in the interest of security or efficiency.

### **Known but Undeciphered**

An ActInf entity may have latent information available which has not yet been deciphered, extracted, or codified.

### **Insufficient Communication Dynamics**

An ActInf entity composed of nested Entities, each with their own internal models, may fail to make use of relevant information due to insufficient internal communication dynamics.

Most important among these several dynamics, is the notion of unknown-knowns within multi-agent systems. The moment that the ActInf entity interfaces with another in cooperation, they become a new perceivable entity, each with internal states that may be more or less synergized. Each has known-knowns and known-unknowns that the other is not necessarily aware of, constituting unknown-knowns in the context of the organization. The AIC model provides support for the argument that, in a turbulent and information-rich environment, top-down management of information dynamics is no longer sufficient - that Rumsfeld's initial prioritization of unknown-unknowns must give way to a prioritization of unknown-knowns, where "more than sufficient knowledge" exists but goes unused or misused in policy formulation due insufficient communication protocols, leading some to call for knowledge and rhetorical ecosystem approaches in the design of more decentralized information systems [123,127].

In this vein, the primary focus of the field of knowledge management might be considered to be addressing the problem of unknown-knowns. As has been addressed elsewhere, when the information management system in question begins to include parties outside the confines of a traditional organizational structure, the management of trust becomes a key concern [123]. The AIC model, in its use as a lens, demonstrates the value of trust in sharing unknown-knowns in a knowledge ecosystem in the form of several notable insights:

### **Trust is Synonymous with Reliability**

Through an ActInf lens, trust is best characterized as projected reliability (e.g., high precision, or low uncertainty) of both other ActInf entities and indicators which inform projection.

### **Trust can be Externalized to Interfaces**

ActInf entities don't necessarily need to trust one another, but instead, can externalize trust to interfaces and related protocols among them in their niche to reduce costs of communication.

### **Trust can be Externalized to Symbols and Signals**

Given that trust is best interpreted within an ActInf context as projected reliability, symbols and signals can thus be "trusted". For example, traffic signals allow drivers to externalize their trust to signals which inform the projection of other drivers' behavior, as opposed to being left to develop trust with other drivers in order to share the road.

### **Trust is a Prerequisite for Efficient Information Sharing**

ActInf entities that question the motives or quality of communications, have high costs in interpreting or accepting those communications.

### **Trust is a Prerequisite for Collaborative Enterprise**

ActInf entities require trust, commensurate with associated risks, in order to engage in collaborative enterprise. Recently this has been explored in the context of human-robotic interactions [2].

We argue that these insights about unknown-knowns, trust, and the non-rivalrous nature of the objects relevant to modern conflict should inform the development of new structures and systems. We distill these insights in order to offer recommendations in the discussion below.

# Discussion

In this paper, we have briefly surveyed models of conflict, considered their strengths and inadequacies, proposed a unifying model based upon the application of Active Inference (ActInf), and considered the implications of use of the Active Inference Conflict (AIC) model. The initial survey revealed that the study and modeling of warfare progressed generally through time from inventories of tactics toward more theoretical and ultimately more abstracted and context-informed analyses. That evolution of the models could be framed as mirroring the parallel development through time of human understanding of human structures of information, as well as structures of cognition, organization, and interaction across the sciences and social sciences, including patterns of conflict in those disciplinary domains. For example, as discussed above, early quantitative models of conflict such as the Lanchester model used mathematical tools that were modern at the time, such as linear regressions and differential equations.

Today, similar analytical and paradigmatic (r)evolutions are taking place in research and understanding about human commerce, behaviors, political governance, and other related domains, ultimately positioning the subset of behaviors and interactions associated with “war” as categories of a subset of patterns of human history and society - albeit patterns that are a non-linear in relation to others. Clausewitz’s observation about politics and war is consistent with this notion of the evolution of the human understanding of the human condition, but following the results of the survey, we contend his famous quote, that “war is the continuation of politics by other means”, is incomplete within this context as it would appear that both war and politics are a continuation of conflict by other means (and, in fact, conflict is a continuation of the normal function of living systems in just another analytical framing).

The survey revealed an increasing abstraction and formalism in the modeling and study of conflict and war, evolving from catalogs of physical battlefield heuristics toward broader and more detailed analytical framings of context and motivations for physical forms of conflict. However, it also indicated that many of the models are underdeveloped for current applications and struggle to address the changing expression of war and the migration of human interactions from predominantly physical interactions (i.e., kinetic warfare) toward abstract, symbolic, and intangible interactions within information landscapes. Further, the survey disclosed that existing warfare models did not have the necessary generalizability to be broadly applicable to the relevant expressions of conflict in other social contexts, and that the models are rarely interoperable.

Following this survey, we proposed the use of ActInf methodologies and terms in modeling conflict and named this application the Active Inference Conflict (AIC) model. The AIC model is intended to represent a needed updating of conflict framing to reflect changes in human interaction patterns, and also provides built-in mathematical rigor that could be used to facilitate the organization and operation of future conflict management architectures. The AIC model, as a consequence of it being founded on the matured quantitative models of ActInf, is tractable to simulate, can incorporate empirical data, and also can immediately be implemented qualitatively to produce novel insights about various forms of conflict. We discussed how this approach, with its affordances for sense and action loops, multi-entity interactions, entity nesting, and policy selection offers old models a new medium for their expression and interoperability while also providing avenues for generalizing conflict modeling which can capture relevant aspects of modern conflict.

Specifically connecting the AIC model to OODA and GW demonstrated the relevance of integrating previous tactical and strategic frameworks within a single multi-scale formal model. Of particular interest was the consideration for the ability of AIC to capture conflicts which have business, operations, legal, technical, and social components, to move beyond generations and gradients of war and offer a new medium for capturing metrics for classifying and clustering myriad forms of conflict, and to model emerging conflict surfaces involving cognitive security and narrative warfare.

Finally, we considered broader implications suggested by qualitative application of the AIC model to conflict generally. We reflected on the state of the information environment, noting the difficulties that traditional institutional and governance structures are having in handling modern information-based conflict and that new, alternative structures for risk reduction are being offered the opportunity to provide value. In addition, we reflected on the non-rivalrous nature of information-as-asset and the infinite potential for information growth, and how these factors affect organizations - mainly in terms of processing information load - which is a useful surrogate for risk. Within these reflections, we suggested that the AIC model is not just useful for the study of conflict but also in the design of systems which manage it. Finally, we applied the AIC model to reveal latent insights about trust and knowledge environments within the Rumsfeld Quadrants, specifically regarding its oft-neglected quadrant, “unknown-knowns”.

The AIC model, as previously discussed, provides an avenue for formal modeling of systems - but this same affordance also facilitates design and evaluation of the design of systems, and to implement and test BOLTS norms and rules. This is to say that a socio-technical system modeled with the AIC model can effectively be a “twin” of that socio-technical system, and thus can be used for more than just studying its conflicts, but also for managing and facilitating endogenous information conflict and

friction itself. It took humans millennia to figure out how to convert the random motion of atoms energized by heat from fire into useful “work” through the use of heat engines. The AIC framing invites consideration of how the equivalent of a “combustion chamber” might be configured for converting the friction of disagreement into useful work within a knowledge environment in terms of developing new information, repairing faulty or incomplete information, discovering unknown-knowns and unknown-unknowns, and helping entities within develop trust and healthy information flows. Within this context, de-risking of interactions in which information exchanges occur could be seen not as a state, but as an ongoing process - which places pressure on designers of information systems to develop simple rules and effective protocols.

Past work has considered how humans and human organizations collaboratively organize, annotate, and structure claims as a form of narrative information management [64,123,128], and could be of use in conjunction with the AIC model to build tools which document, facilitate, and resolve informational conflicts with an objective dimensionality from the AIC model that leverages existing approaches. Further, these pairings of approaches could help give new life to the older narrative models of conflicts and unify them with the work on commons management [79], as it could provide a new medium for formalizing, documenting, and sharing of heuristics and practices for risk mitigation [32].

As the rate of information growth continues to explode outward in both volume and complexity, the AIC model reveals that the search for unknown-unknowns or known-unknowns may need to be deprioritized, as this information may fail to be disseminated and integrated - rendering most relevant information as unknown-knowns. Where “hope” was left in Pandora’s box, it might be said that “trust” was left in Rumsfeld’s matrix. The AIC model helps to demonstrate and codify the value of trust in knowledge ecosystems which facilitate the sharing of unknown-knowns, and demonstrates how trust can be externalized to protocol and signals through their being reliable indicators of quality and behavior. Ultimately, a primary suggestion of this work is that facilitating mutual interdependencies, interfaces, and trust-management frameworks, key prerequisites to sharing unknown-knowns, could attract an increasing subset of information conflicts into generative structures (perhaps best framed as structures which operate in the manner of what might be called a “risk commons”) which can capture value and grow trust, rather than accelerate discord. Below, we distill these and other insights within this work into recommendations for future research and the design of new systems:

## Recommendations

- Develop more work on the use of the AIC model in extending the value of OODA loops in simulation and decision-making models. This could utilize complex systems modeling software such as cadCAD [129], and those specifically for ActInf such as ForneyLab [130] or infer-actively [131].
- Explore the use of the AIC model in modeling past conflicts as a basis for measuring various attributes of those conflicts, and the use of those attributes for new classifications and “generations” or gradients of conflict.
- Explore the use of the AIC model and the integration of commons management principles and compensating controls across business, operations, legal, technical, and social (BOLTS) surfaces.
- In the design of information exchange systems:
  - Acknowledge de-risking as an ongoing process, rather than as a static attribute.
  - Consider trust as synonymous with perceived reliability.
  - Make use of the fact that trust can be externalized to signals and symbols so long as those signals and symbols are reliable indicators of behavior and state.
  - Consider disagreement, inconsistency, and incoherence as events which can be mined for value via shared protocols and standards rather than creating an illusion of security through attempts at their removal.
- Across many domains (e.g., war, scholarship, and design), reprioritize Rumsfeld’s neglected quadrant of unknown-knowns.

# Contribution Statements

**Administration and Facilitation:** R.J. Cordes

**Writing, Editing, and Revision:** All authors made substantial contributions to writing, editing, and revisions across all sections.

## Funding and Acknowledgements

Scott David is funded by the NSF Convergence Accelerator Trust and Authenticity in Communication Systems Program (NSF 21-572), under award ID #49100421C0036.

R.J. Cordes is funded by the NSF Convergence Accelerator Trust and Authenticity in Communication Systems Program (NSF 21-572), under award ID #49100421C0036 and is supported in research efforts through a Nonresident Fellowship with the Atlantic Council on appointment to the GeoTech Center.

Daniel A. Friedman is funded by the NSF program Postdoctoral Research Fellowships in Biology (NSF 20-077), under award ID #2010290.



## Chapter VI

---

# **An Active Inference Ontology for Decentralized Science**

**From Situated Sensemaking to the Epistemic Commons**

Daniel A. Friedman, Shaun Applegate-Swanson,  
Arhan Choudhury, R.J. Cordes, Shady El Damaty,  
Avel Guénin—Carlut, V. Bleu Knight, Ivan Metalkin,  
Siddhant Shrivastava, Amit Kumar Singh,  
Jakub Smékal, Caleb Tuttle,  
& Alexander Vyatkin

## **Abstract**

---

In this work, we examine science from the vantage points of blockchain technology and its connection to decentralized science (DeSci). We consider science as a collective process using Active Inference, an integrative framework that models the cognitive processes of perception, planning, and action selection in terms of Bayesian probabilities and updating. We present the Active Entity Ontology for Science (AEOS) as a composable and versionable system for modeling various science systems, using the Active Inference entity partitioning. Further steps for developing and utilizing AEOS in the context of scientific ecosystems are provided.

---

# Driving Questions

## **Centralized Science (CeSci): Operations, Community, and Practice**

- What are successful, effective, and productive scientific practices for communities and individuals?
- What frameworks and tools can help form and support communities of practicing researchers to promote long-term collaboration and impact?
- How are preferences and expectations aligned and communicated in ecosystems of scientific collaboration?
- How do we avoid perverse incentives in systems of scientific inquiry?
- How do we ensure effective and transparent resource allocation and funding as research becomes more interdisciplinary, interorganizational, and international?
- How can scientific careers be started and nurtured, increasing the accessibility and vitality of science as a community and body of knowledge?
- What are fair and effective means of elevating voices to scientific thought leadership roles?
- What are good decision trees to identify strategic priorities for scientific funding, intellectual capital allocation, and focused development?

## **Decentralized Science (DeSci): Coordination Engineering for Knowledge Building**

- How can decentralized teams and organizations coordinate research using effective distributed mechanisms?
- What scientific outputs are within the scope of Decentralized Science (DeSci)?
- How will DeSci produce traditional research products such as papers, as well as other outcomes, such as products, services, platforms, tools, and protocols?
- What artifacts and design patterns will stimulate the development of integrated DeSci systems?
- How can Distributed Autonomous Organizations (DAOs) scaffold and catalyze research, grants, education, and community development?
- What methods for value capture and financialization improve the process of scientific discovery, and which methods result in poor outcomes that are at odds with an open science community?
- What improves the reliability of outcomes and accountability of funds in emergent collaborations?
- What are the relations between the trajectories of scientific governance, and social/ecological/political change and collapse?
- How can scientific careers become more accessible to researchers who are non-PhD holders, and those with other commitments such as family, participation in industry, etc?
- How can we use blockchain technology and signature economies to cultivate cultures of fair work attribution, through records of contribution to knowledge artifacts and proof-based protocols?
- How can we maintain a favorable signal-to-noise ratio for DeSci work output?
- Can the Active Inference framework and ontology be used to understand complex sociotechnical systems such as epistemic communities?

# Introduction

Science is a cumulative, collective endeavor that augments a single human's sensemaking ability with tools, practices, and processes we use together to observe, measure, and understand the world around us. Even before the invention of modern scientific instruments such as microscopes and particle accelerators, humans relied on their senses to observe the environment and on their brains to make inferences about these observations. This is still true today. Any study of how knowledge is accumulated must consider how the brain works when making predictions based on observations and incomplete data. Additionally, as the information required for scientific research extends far beyond the cognitive capacity of a single unaided human, tools for scientific organization, knowledge management, and communication become increasingly necessary to consider as well.

## Science and the Epistemic Commons

A recent shift in science is the move towards greater recognition of the complex nature of systems such as brains or societies [1,2]. For example, cognitive science has largely moved away from the ambition to discover universal laws of behavior to focus more on the causal structure of the brain and its patterns of activity in different scenarios [3]. Utilitarian scientific models today strive to explain natural phenomena by piecewise integration of generative models expected to match the causal structure of the target system at the relevant scale of behavior [4,5]. The development and validation of mechanistic models is based on multiple cognitive strategies, for example involving computational modeling, empirical observation, statistical inference, and domain-specific theoretical considerations. Despite this, the social structure of science has not yet shifted to embrace integrative practices. While the scientific community has long-standing practices which they optimize for, many of these parameters have lost their validity, efficacy, or legitimacy [6,7]. This results in suboptimal integration of complex transdisciplinary knowledge by favoring the partitioning of research along the lines of domain-specific methodologies, organizational processes, thematics, and vocabulary.

Indeed, the modern era has seen the emergence of an unprecedented hegemony of administrative institutions over all other forms of human social organization. Political anthropologist James C. Scott explains the emergence of top-down structures for complex problem solving as a default reflex to compress complexity. Scott claims that the capacity of bureaucratic states and administrators to understand and manipulate the world relies on a systematic reduction of the world onto standardized, legible tokens (teachable facts), such as accounting, and measurement units or legal identities based on a permanent patronym [8]. The need

for this "view from above" has placed pressure on research institutions to adapt both their objectives and operating activities to meet the needs of administrators. This "technoscientific" logic has left a complex and pervasive legacy in the organization of contemporary research, including in the way we understand and practice scientific research. In worst-case settings (which are unfortunately common), research quality is evaluated on superficial measures with no principled consideration of deeper relevance, leading to a raging competition between researchers on the number of words written, sheer number of publications of any kind, or the ability to use technical language [9–12], without considering the positive impact of the work. Entire research programs can be funded because of insider networks, or correlation to a hype cycle, while more thoughtful and critical (and ultimately constructive) research may easily go unnoticed. In other words, the key metrics for tracking global scientific progress have become increasingly divorced from their ability to replicate or have meaningful impact [13,14], and more correlated with social and cultural signaling within the scientific community. Further, when the professional scientific community is only a small fraction of society, there is the possibility of detachment or isolation from broader goals.

Given the incentive misalignment and power imbalances between administrators and practitioners, the scarcity of resources, competition over resources both between researchers and research institutions, and the inflexible nature of the bureaucracies which connect and govern extant research institutions, there are few mechanisms available for improving traditional research institutions aside from increasing funding. Luckily, human prosociality has taken many forms throughout our evolutionary history from which we can draw inspiration. We could develop tools and protocols which treat the outcomes of research as contributions to an epistemic commons, a web of informational or knowledge-oriented (epistemic) systems which would allow individuals with disparate motivations, skill sets, and beliefs to collaborate at the intersection of their interests and tackle specific questions or systems with the most contextually relevant tools, in the best interest of cumulative, collective scientific knowledge. In other words, we argue that relaxing the constraints of legibility associated with bureaucracy and its "view from above" could help develop the "view from the ground" (i.e., pragmatic understanding by practitioners) as well as the "view from within" (i.e., scientific knowledge, in the classical sense of an objective description of system organization and activity).

Human prosociality has contributed to the persistence of our species. The emergent structure of social organization is variable but can develop towards hierarchical, top-down control which are transactional and mutually beneficial [15,16]. However, hierarchical control can lead to the sequestration of wealth and power over time, and a centralized motivation that does not always benefit all members of the community. Furthermore, hierarchical structures develop cognitive biases over time that limit the ability to parse multiple, conflicting streams of information that do not

fit into standard decision making templates. Decentralized structures may be more capable of acting on divergent information streams, however decentralization alone is not a mechanism that provides intelligent coordination for effective collective action [17,18]. Decentralized Autonomous Organizations (DAOs) are an emerging form of coordination that enables new kinds of cyber-physical communities mediated by immutable, cryptography-protected code [19,20]. DAOs allow for programmable digital social structures, embedded mechanisms for consensus, intermediation of trust-requiring actions, and independent operation of workstreams with multiple leaders instead of a top-down hierarchy. Coordinating such decentralized communities of practice, or more generally governing "complex commons", is novel, difficult, and high-stakes work, often without the scaffolds and norms that traditional offline governance can benefit from [21]. For example, systems that are decentralized in principle may still be functionally centralized in terms of power or asset distribution [22].

We propose that epistemic drives and norms can grow organically from communities of practice, defined by open questions, methodological synergies, personal affinities, diverse contributions, and serendipitous encounters. We model the similarities and differences between Decentralized Science (DeSci) and Centralized Science (CeSci) through the Active Inference framework, understood here as a framework addressing how living systems create an understanding of their environment [23]. The Active Inference formalism affords an integrated account of the multiscale dynamics shaping research, from the constitution of epistemic affordances shaping scientific practices [24] to the coevolution of epistemic beliefs and communities [25], as a process of uncertainty reduction entailed by the descent of a gradient of variational free energy. As applied, this model implies that cognitive understanding derives from adaptive control in agentic engagement rather than explicit representation [26–28], and that the activity of epistemic communities is shaped by the sociocultural constraints defining their structure rather than individual states of mind [29]. Based on these considerations, we articulate a conceptual framing and Active Inference entity ontology model for Decentralized Science.

Before we dive deep into head-on contrastive definitions of CeSci and DeSci, a primer on decentralized systems is crucial. We would like the first definition not to be simply an anti-definition of its counterpart, as we aim to incentivize effective research and set up collaborative relations between emergent DeSci and incumbent CeSci, not simply to ignite controversy. Therefore, even though here we will use the terminology of decentralization, considering nuance in the definition [22,30–32] as well as alternative terms, 'Decentralized' is a good first step. One interesting perspective to apply here is that of "Self Certifying Systems" [33]. These self-certifying systems are those in which an entity validates its own existence in the system, and the incentive mechanism of the system is intrinsic to it. For instance, a dataset in a database system such as InterPlanetary File System (IPFS) [34,35]

validates its own legitimacy by hashing the contents using a cryptographic hash function. In these dynamics, the root of trust lies within the network entity itself, and not necessarily the legitimation from sources outside of the system.

## Centralized Science (CeSci) and Decentralized Science (DeSci)

Decentralized Science (DeSci) is a relatively new term, introduced by Web3 collectives to describe the use of recent digital tools for funding, training, planning, coordination, execution, dissemination and archival of scientific activities and assets by digitally connected communities. DeSci is an emerging area with multiple, potentially even incompatible, forming senses, so all explorations here should be considered preliminary and partial. The term DeSci suggests, by contrast, the existence of Centralized Science (CeSci), which would stand for the continued status quo organization of science as a highly bureaucratic activity managed by select academic and private institutions. Some motifs or patterns that are found in CeSci and DeSci can be found in Table 1 and Figures 6-8.

DeSci can be defined as the use of Web3 technologies to introduce markets through the deployment of open source finance tools. In this view, DeSci represents the introduction of commodity markets for scientific assets and services where such markets were previously impossible (e.g. financialization through tokenization of intellectual property, scientific platform governance, peer review or curation services, or access privileges to data or infrastructure). In another perspective, DeSci is seen as a set of mechanisms for bottom-up individual sensemaking. From this perspective, DeSci is the capability of individual agents or communities of practice to make sense of the world autonomously, by defining their own questions, language and methodology. These disparate views represent the distinction (but not preference or judgment) between DeSci as a set of emerging Web3 tools and DeSci as a research ecology, respectively.

Tool use, including markets and organizational technologies like DAOs, are fundamental to the ecological view of science [36]. However, proponents of DeSci do not conceive of this distinction as a purely technological one, for example simply describing the fact that DeSci organizes with the Internet while CeSci utilizes paper as well as digital media is an insufficient distinction. For example DAOs don't replace off-chain or offline human organization, they augment it with on-chain guarantees and some automatic processes for human organization (i.e. voting is still a human process, but vote security may be automatic). The broader perspective on DeSci is that it aims to reform the organization of scientific activity itself, which is expected and preferred to translate into an increased ability for Science to fulfill its mandate and align with social values [37]. We fully agree that this expectation is warranted,



at least in principle, as the socio-epistemic dynamics we call Science clearly derive, at some level, to the concrete organization producing it. We will therefore attempt to characterize the distinction between CeSci and DeSci in terms of their respective socio-epistemic organizations.

Both CeSci and DeSci aim at knowledge discovery and curation, but they differ in their means to this end in terms of their incentives, structure, and norms, or more generally, in terms of their research ecology. In CeSci, the language, methods for discovery and strategic priorities are imposed from the top down by a core group of decision makers acting out the mandate of government policy positions. Core groups of decision makers also receive external input and constraints, for example in the case of peer review. In DeSci, methods and norms emerge from the bottom-up interactions of a web of loosely connected communities of practices with diverse coordination and prioritization mechanisms. Therefore, CeSci and DeSci diverge significantly in terms of their sensemaking; one imposes specific meaning, understanding and other cognitive constraints to scientific activity so as to keep it legible to the "view from above", the other cultivates the "view from the ground" by allowing those same cognitive objects to unfold spontaneously throughout the course of scientific activity.

Though there is fundamental overlap, and perhaps more similarity than not between CeSci and DeSci at this incipient stage, those terms stand for a fundamental divergence in the social and epistemic dynamics shaping the organization of research. Here we explore some of those dynamics by considering key differences between DeSci and CeSci.

## CeSci: centralizing institutions against situated sensemaking

CeSci is characterized by the presence of centralizing agents (such as government institutions, philanthropic foundations, private businesses, or universities) having the power and willingness to establish and enforce rules, goals, ontologies, and other informal constraints within the wider scientific ecosystem. Its purpose can be understood as pragmatic rather than purely epistemic, in that CeSci institutions are funded and motivated, internally and externally, to build epistemic value (i.e. knowledge) that can assist other institutional actors and stakeholders understand and manipulate their reality in terms of their objectives, for better or worse. An example of this is the 19th Century development of statistics, as it was explicitly intended to help the developing modern States to perceive and control their population through the collection and analysis of demographic data, and effectively shaped what it meant to "think like a State" [29]. Due to extrinsic mandates, CeSci orients the sensemaking activity of its participant researchers toward specific technical and pragmatic purposes.

Although the goal of scientific research has largely extended beyond war and administration, the modern logic of centralization and technoscience is largely continued in contemporary research. The managerial role over the epistemic commons was granted to late modern academic institutions, and complemented with the role of supporting technological development through research on the property of chemicals, mechanical systems, and the like. Techno-scientific institutions have a core mandate of assisting outside institutions in their pragmatic endeavors. Their overarching goals are to accumulate epistemic value in that direction. Influence over research institutions by external organizations can be direct or indirect. In the first case, it may mean cadres of administrators exercising direct control over what research objectives should be, what funding agencies can be or should be approached, and recruiting based on compliance with external mandates. Indirect influence can be exerted, for example, by external organizations offering funding based on use of specific language or alignment with pragmatic research outcomes.

Accordingly, the production of knowledge artifacts, such as journal articles, is currently enacted according to a hierarchical structure of funding agencies, scientists, institutions, and publishers. There are in-groups and out-groups, as well as unspoken rules to obtain funding for and publish results of scientific literature. Moreover, there are agendas driving all agencies that fund scientific research, which are not necessarily transparent. Many basic research questions are framed in a way that adheres to these agendas. For example, researchers trying to characterize a mechanism by which a brain protein folds will need to emphasize its supposed role in curing Alzheimer's disease, when in fact any actionable outcome may be unclear or distant from the research [38]. This situation provides a strong incentive for researchers to leverage the information advantage they have over administrators to grab their attention by making incredibly grandiose, possibly dishonest claims on the meaning of their research. Administrators simply lack time or incentive to check on such claims, and active researchers lack the power to modify the incentive structure and would risk their job if they refused to follow them.

One aspect of the difficulty in communicating between administrators and active researchers is related to a discrepancy between the relevant time scales and perspectives for their respective activities. Bureaucratic funding agencies (private or governmental) have specific, long-standing research directives that cannot be modified easily, both because of long-term commitments on organization structure or mandate and because of the limited time officials have to understand, evaluate and integrate feedback. For example, an academic administrator will tend to lack domain expertise in the details of protein folding, and to lack motivation to discuss how category theory or information geometry could lead to major scientific breakthroughs. Administrators, in virtue of their function, are motivated to discuss how the research they manage contributes to the competitiveness and prestige of

the institution they answer to, and why it is therefore worthy of more funding. They may encourage their employees working on protein folding to claim in bad faith that their research will solve cancer, to publish a lot of papers into highly competitive (although not necessarily domain-relevant) journals, or to collaborate with companies for patents. If, for example, experts on protein folding want to express the benefits of future conceptual advancements, the importance may not be conveyed among fellow researchers and administrators due to a lack of common language, and therefore an inability to make decisions in a framework of mutual understanding.

Because of this fundamental asymmetry in power, information, and incentives, the centralizing hierarchical structure of science interferes with the rapid adaptation and evolution of scientific ideas. The need for researchers to make their ideas and objectives legible from the perspective of administrators conflicts with their ability to develop a context-relevant ontology and to focus on actually understanding the target system, or aiming to understand phenomena beyond the scope of expected research narratives. Centralizing institutions need to build largely artificial semantic and incentive silos, both as a way to compete against rival institutions and to measure and control the activity of researchers. Centralizing institutions organize science around outcomes which are legible and manageable, therefore hampering the kind of situated and transdisciplinary sensemaking which enables and grows our understanding of the world, and our ability to ground this understanding within a common language.

## DeSci: governing epistemic commons

In contrast to the centralizing institutional view presented above, research can be organized as a decentralized federation of communities of practice, each striving to address specific issues in a contextually-relevant way. Such communities, by definition, lack a direct dependence on any centralizing institutions, or at the very least such institutions lack the ability to step in and micromanage their activity. Of course, the mere fact that those communities do not need to adapt their language or objectives to the expectations of administrative cadres does not make their activity scientific in nature. However, decentralization offers specific opportunities to scientific communities by allowing them to focus on developing pragmatic solutions to concrete problems and/or purely advancing the epistemic understanding of the natural world, free of centralizing institutional bias. As decentralized communities of practice lack the coercive power which defines centralizing institutions, their continued existence relies entirely on the voluntary participation of members, and therefore necessitates an institutional structure facilitating and incentivizing participation. Consequently, and also due to the lack of semantic silos (see below), decentralized scientific communities will have a strong incentive to cooperate with each other, to make relevant issues discoverable to

wider sections of the population, to allow participation purely out of voluntary interest, and to develop unifying languages easing their integration.

The opportunity for knowledge appropriation (via legal enclosure through intellectual property institutions as well as artificial semantic siloing) relates the development of scientific knowledge to a broader, well-studied category of collective action problems: the governance of common goods [39]. Common goods (commons) are defined as resources that are rivalrous or non-rivalrous yet non-excludable (one cannot be forbidden from accessing them). The potential for misuse or pollution of the commons consequently necessitates governance at the scale of communities of stakeholders, and regulation needs to follow certain conditions that were initially formalized in economic sociology [11,12] and recently generalized in evolutionary theory [40–42].

Roughly, the implementation of governance systems for commons entails:

- the unambiguous definition of a problem to solve;
- the construction of a community of stakeholders willing to help solve this problem;
- the institution of a decision system allowing participation on an equal footing by all members of the community;
- agreement on a set of rules (which are interactive and mutable with validity checks over time), which fairly reward active cooperation within the community;
- the iteration of (1)-(4) at higher scales where the community as a whole is confronted with a situation requiring governance.

Although the definition of knowledge as a commons (a public good) entails a discussion of how exactly knowledge can be appropriated, the relation of decentralized science to common goods management is rather straightforward. Communities of practice organize around specific problems or questions, define rules for participation, and coordinate with each other to define and solve meta-problems - such as the development of standard tools or a common language or unifying framework which facilitate progress within each community.

Thus far, the distinction we have made between Centralized and Decentralized Science is largely organizational. For example, early European academia emerged from the clerical legitimation of scholarly "guilds," professional associations similar to contemporary trade unions or cartels, which effectively restricted who could practice a certain craft or access a specific kind of knowledge within a certain area

(generally a city). Thus, patterns of sensemaking were immune from any outside oversight, and were imposed by the community on each and every one of its individual members. Centralizing institutions robbed community members of important aspects of their intellectual autonomy, and clearly enacted autonomous agency (see Thinking like a State [29] for a discussion of institutional autonomy). In contrast, decentralization is facilitated through processes that facilitate broad access to the epistemic commons, as described above in the example of the printing press. When we highlight the opposition between CeSci and DeSci, we are prefiguring a discussion on how specific kinds of organization can facilitate differences in the integration of information about the world over multiple dynamical scales, and what it means for the scientific community.

In the contemporary age, digital technologies provide clear opportunities for proponents of DeSci. Indeed, they facilitate the widespread diffusion of information with a relatively low entry bar, and therefore the voluntary cooperation of multiple communities of practice engaged in situated sensemaking. This encourages the decentralized adoption of standardized methodologies and languages, transparency over code and data management, and willingness to produce information of clear value to outsiders. Indeed, because semantic silos cannot be enforced in order to increase one's influence, the digital dynamic tends to encourage collaboration in the interest of the broader system because it is the only way by which an actor can fulfill their objectives (see DeSci Token Engineering below). We see the success of such a decentralized yet unifying project in Wikipedia, which is beyond any doubt the most important example of a largely standardized, accessible and transparent knowledge system, and is based entirely on voluntary and decentralized cooperation. At the moment, there is however no appropriate, analogous system to use in the organization of scientific activity, as an alternative (or complement) to the currently dominant role of centralizing institutions.

## Implementing Decentralized Science in the Web3 era

Contemporary digital tools have the potential to establish a distributed, transparent, scientific community with open participation, where every participant has root level access to all knowledge and the rules for its production. Online communities have seized the opportunity to develop and utilize such digital tools for funding, training, planning, coordinating, executing, disseminating, and archiving scientific assets and activities, therefore establishing the blueprint of a Web3 DeSci. There is no defining characteristic of the whole set of Web3 DeSci tools. Nonetheless, some features characterize a larger subset of these tools and are worth highlighting.

First, Web3 tools allow for more interactions (including economic transactions) between economically- and geographically-diverse groups. Second, Web3 tools can increase the transparency of products, activities, and interactions that are themselves digital, or that can be represented digitally. Third, Web3 tools allow for the introduction of novel markets that, in some pockets, might increase decentralization but in others, might increase centralization. Instead of using general terms to elaborate on each of these features, we hope to demonstrate their reality and significance as we overview more specific features of and developments toward a Web3 DeSci implementation.

## Web3 for DeSci

Before heading into some details about the implementation of DeSci (which may just be one iteration or genre of the infinite game of Science), the reader should be equipped with familiarity around some basic Web3 terms: <sup>1</sup>

- Web3
- Blockchain
- Smart contract
- Token (fungible and nonfungible)

### **Web3**

First, "Web3" is an imprecise term referring to an ethos that values digital self-sovereignty and, more important for our purposes, to recent developments in peer-to-peer protocols that service this ethos. Part of the Web3 ethos is to value open source software and the personal control of one's data [43]. These values have influenced and continue to influence the design and adoption of these protocols. Indeed, the term "Web 3.0" was introduced by Gavin Wood [44], a key figure in the development of the Ethereum and Polkadot protocols, who proposed Web3 as a "post-Snowden" Web, a decentralized Web similar to the Web1 of the early internet and in contrast to Web2

---

<sup>1</sup> The listed terms are just current technological means to building a desired framework, and are presented here as introductory, not as comprehensive reviews or final answers

where a small number of centralized agents control people's data, access to websites, and the web applications that many use every day [45]. Web3 protocols are diverse, but the ones at the heart of most functional systems are so-called "blockchain" or distributed ledger technology (DLT) protocols that support complex smart contracts.

### **Blockchain**

Recent developments in peer-to-peer (i.e., decentralized, distributed) protocols, particularly blockchain protocols such as ones that facilitate Distributed Autonomous Organizations (DAOs [20,21,46]), are the biggest enablers of a Web3 DeSci implementation. Without going into detail, a blockchain protocol enables a network of self-interested, competing computer nodes to agree on the state and immutable history of a shared ledger. Anyone with an internet connection and enough funds can append the ledger or verify its integrity, though it is still possible to implement access control with smart contracts. Blockchains are a natural fit for building currencies, which historically emerged as an extension to bookkeeping by institutional actors [47], and currency building in fact constitutes their most widely known application today. However, the principles of data integrity and access apply just as well to running any record or application that must be tamperproof. Blockchains can be used to establish ownership of digital assets (e.g., datasets or intellectual property [48]), mediate between pseudonymous actors (e.g. during peer review), establish decentralized identities used for reputation (e.g. when applying for grants), prove an actor has done something (e.g. stored a dataset or run a compute service), and more. The key here for DeSci is that blockchains enable actors to keep track of scientific activity in permissionless systems, and to build systems of exchange and accounting around any kind of digital information – which includes much of the products and tools of modern science.

### **Smart Contract**

Blockchains can establish digital records via programs called smart contracts. A smart contract (in the context of blockchains) is a computer program whose code is publicly available but cannot be changed by anyone, is executed on computers across a network, and can be invoked (i.e., started) by anyone [49,50].



Smart contracts run on blockchains. When the blockchain's record is updated by a smart contract, the update is verified by the distributed nodes on the blockchain network. Despite the presence of "contract" in the term, smart contract does not refer to a legal contract, though it is still possible to implement certain legal agreements in smart contracts, such as the triggering of payments in the case of predetermined on-chain activity. Not all blockchain protocols support the complex smart contracts that enable the features listed in the previous paragraph.

### **Token (fungible and non-fungible)**

One kind of smart contract is called a token. A token is a digital asset. It can be bought, sold, or transferred on the blockchain. The implementation of a token in code is little more than a record of which addresses (i.e., users, loosely speaking) own how many units of the token. When units of a token are exchanged, the record is updated, not unlike what happens in a bank's database when two of its customers use the bank to transfer money to each other. Tokens come in various forms, broadly separable into fungible and non-fungible (though see [51] for a more complete discussion). A token of the latter form is often referred to as an NFT (Non-Fungible Token). The salient difference between these types of token is in the name; there are many interchangeable units of any given fungible token, and there is only one unit of any given non-fungible token. Because it is easy to use tokens to record ownership, they have been used by DAOs to designate voting rights (similar to shares of stock in a corporation) or provide proof of participation [52,53]. Tokens have many other use cases and have even enabled the emergence of a discipline known as token engineering [36], which studies the use of tokens to incentivize certain behaviors within a market.

## **DeSci Organization: DAOs, Blockchain, and Markets**

The blockchain technology affords certain advantages over alternative systems in terms of transparency, decentralization, and tamperless regulation. However, its development was quite explicitly motivated by the implementation of market mechanisms at the exclusion of other forms of social organization, and therefore focused heavily on the implementation of monetary transactions. The issue with that position is straightforward. The very existence of multiplicative investments in a

monetary economy (as entailed, e.g., by the legal status of companies in capitalist countries) entails the emergence of an extremely polarized power law distribution in wealth, and can (even in the absence of strategic interaction or information asymmetry) lead to the control of a large proportion of the economy by a few individuals [26]. In reality, strategic interaction and information asymmetry are ubiquitous, and the power entailed by the wealth of a few aristocrats gives them the capability to build a population of dependents and influence the rules of the game to their advantage. A complete disinvestment of public powers or otherwise democratic institutions from economic redistribution and regulation would therefore expectedly lead to the development of a feudal-like economy, where a few wealthy individuals use their economic leverage to yield near absolute political power over the many.

The fundamental vulnerability of marketization-oriented Web3 technologies is deeply aggravated by preexisting asymmetries in wealth and power. Any token that can be traded freely (while not pegged to a stable asset) is vulnerable to price instability entailed by speculative investment, which would make it essentially worthless as a currency (i.e., a means of holding and exchanging value). On the other hand, any token that is not coupled in some way to the mainstream economy cannot be used to pay for basic goods such as food or shelter, and therefore lacks any economic value. Historical currencies were instituted by States through 3 main leverages: by creating demand through taxation, by backing their value to some external value holder (generally rare resources such as precious metals or recognition of debt by institutional actors), and by forcefully crushing the non-monetary economy and the communal institutions regulating it [47]. None of these leverages are accessible to decentralizing collectives, at least not without significant State / capital backing and an open betrayal of their explicit mandate.<sup>2</sup> Although the problem of monetary value is specific to currencies, the very demarch of marketization entails the ability to create economic value - or, in other words, to find a buyer for what you're trying to sell. For example, if no one is willing to pay you (in fiat currency, crypto, or some other value) the equivalent of a living wage for cleaning data, you are not going to make a living cleaning data.

Despite the aforementioned shortcomings, marketization can still play a substantial role in scientific decentralization. Markets offer an opportunity for large scale economic integration between peers, and therefore for the basis for an "exodus" from the grasp of centralizing institutions [55]. For example, tokens could provide an explicit and unfalsifiable account of traditionally less-valorised work, such as data management. However, blockchain-powered marketization cannot provide a sufficient solution for a successful decentralization of scientific activity. The

---

<sup>2</sup> see [54] for a general discussion of value generation and regulation in cryptocurrencies.

blockchain only ensures decentralization in the limited sense that it makes breaking the rules that were written onto the blockchain extremely difficult, and therefore helps enforce the libertarian notion of liberty, defined negatively by the absence of explicit coercion [56]. Building DeSci, on the other hand, entails a broader shift toward positive liberty to do science in a decentralized way, i.e. the empowerment of individuals and collectives to organize scientific activity around decentralizing values and institutions. This process entails that we pay attention to the actual dynamics of decentralizing systems, and ensure that individuals and collectives can effectively enact their formal freedom of collaborative sensemaking.

DAOs, whether enforced by blockchain technology, or more classical means, provide the direct structure for their own activity. DAOs are organizations defined by an explicit mandate and specific rules for participation and decision, and are usually open to anyone willing to accept them. In the extent that these rules are reasonably fair, and that individual participants get a say in the decision system and overall organization, DAOs constitute a natural fit for the management of commons.

Of course, none of this excludes the value problem discussed above: science mandates that scientists be paid a living wage while they focus on their work, or otherwise to have enough time and resources to be able to conduct scientific work. In consequence, a radical shift to DeSci would likely necessitate broader social reform, such as the institution of a right to work in the research domain of one's choice, or perhaps a solution such as Universal Basic Income. Some cryptocurrencies, like the Ğ1 currency [57], try to build such change from the ground up by distributing a predefined rate of monetary creation equally between members, and explicitly grounding the networks to social webs of trust and exchange. But implementing such changes, especially at a global scale and within a reasonable time frame, would be entirely beyond the ability of contemporary DeSci communities. In the remainder of the document, we will therefore focus on steps that DeSci can practically take right now, many of which are related to tokenization and marketization mechanisms.

## DeSci Vision and Open Questions

In our society today, a relatively small number of individuals have full-time careers in scientific research. Those with family commitments, a significant role in industry, or an aversion to academic social systems are largely absent from the process of creating and curating accessible knowledge artifacts and truth-seeking using the scientific method.

Can we support and develop communities that train researchers from many diverse backgrounds and provide opportunities for scientific collaboration, funding, and publication outside of the narrow constraints of CeSci career trajectories? How can we motivate and reward productive scientific engagement in this system, where there are less rigid barriers for who is allowed to conduct research? Can we benefit from neurodiversity and large-scale harmonization of different perspectives to reach epistemic consensus states beyond the reach of the narrow confines of what constitutes a "traditional scientific mind," or collection of such minds, in our society today? Decentralized communities of practice focused on research are already self-organizing, and will begin to answer these questions over the coming decade.

## DeSci Token Engineering

Blockchains and tokens can be used to create incentive mechanisms to promote desired behavior and punish undesired behavior, and these incentive mechanisms can be used in DeSci to encourage valuable scientific activity. The practice of developing incentive mechanisms that utilize tokens is referred to as "token engineering" [20]. There are many ways to use tokens to reward behavior. In Filecoin, for example, a storage node is rewarded tokens for storing data only if it submits successful cryptographic proofs demonstrating it has stored a unique copy of the data over a predetermined period of time; the node's collateral is reduced if it fails such proofs. Not only can DeSci utilize systems that use token engineering, but DeSci can create token-engineered ecosystems. Token-engineered systems might be created for research-specific use cases, for example [58]. Token engineering can also be used to reward researchers and establish a value flow for a scientific ecosystem, a topic which we turn to now.

In the case of scientific journals, recent adoption of blockchain technology in the field of data economy can serve as an inspiration for designing a decentralized knowledge ecosystem. Blockchain offers the tools to observe, design, and control complex system behavior through the means of token engineering [59]. Were we to model entire communities as a collection of interacting agents which transact, perform work, and make economic decisions (e.g., investment, staking), this simulation could provide insight into the effectiveness of the overarching incentive structure, the resilience of the system to various risks, and the capacity of the organization for interoperability. We leave this as an opportunity for future study.

Token system engineering might ask questions like:

- How are stakeholders rewarded for meaningful participation?

- What is the value flow? Where does value start/enter?  
Where does value go/exit?
- What is the value distribution, internal and external to the system of interest?
- How does the system shape patterns of attention?

One example of token engineering used for value allocation is based on the Ocean Data Marketplace [60], in which a community-funded treasury disburses funds in the form of fungible tokens to research teams based on criteria defined by the community. At this point, the research team has full autonomy and is responsible for efficiently distributing the funds across different researchers. Researchers can send tokens to the decentralized knowledge market to gain access to datasets, papers, or any other scientific work published by other researchers, who will receive these tokens as compensation for their contribution to the DeSci ecosystem. Every transaction made in this decentralized knowledge market is facilitated through smart contracts, which collect transaction fees that are fed back to the community treasury.

The discussion on science token engineering leads to a natural question: what are the limits of decentralized science? Alternative value flows designed by the practice of token engineering assume new laws are enforced by digital contracts deployed on a transparent blockchain. As such, their ability to enforce particular actions outside the digital world become very limited. In its current form, DeSci is constrained by the affordances available to digital entities; hence the scope of scientific outputs is also constrained to research that has digital elements.

## DeSci Research

Currently, most active DeSci projects primarily focus on building the infrastructure necessary to allow the dynamics described above, while only a fraction use their resources on research <sup>3</sup>. Examples of research-focused DAOs include VitaDAO and Molecule which, at the time of writing, have funded active projects in longevity research. Other active groups include the Governance Research Institute, Governauts, and Metagov DAO, researching governance, smart contracts, and other aspects of coordination in the Web3 space, all of which are closely related to the development of the DeSci infrastructure. As such, there is no clear distinction between infrastructure research and development. Furthermore, many DeSci DAOs

---

<sup>3</sup> A non-exhaustive list of the DAOs and groups mentioned is available here, building on [61], and will be updated.

have a clear vision of the type of research they want to conduct, but are currently in the process of getting funding or curating an initial community of active contributors.

There are many open questions relating to creating and sustaining impactful communities of practice in a Web3 DeSci setting. Technological changes to the process of science in DeSci may be exciting and impactful, but cannot circumvent the fundamental challenge of carrying out high-quality research. Even with hypothetically-successful systems in place for proposing projects, knowledge management, financial support, and decentralized publishing, individual researchers must still be motivated, committed, and engaged to produce useful work. Creating and sustaining impactful DeSci communities is challenged by the scale of information available for review or meta-analysis projects, the size of the design space of possible experiments, the multiple individual ways of working, and the difficulty in producing quality research. Without long-term mentorship, training, and support for scientists at all career stages, it is difficult to imagine that merely increasing the extent or type of system decentralization would result in a successful scientific ecosystem.

## DeSci Publishing and Review

DeSci seeks to use Web3 tools to improve the transparency and incentives within academic publishing and to improve the accessibility of published products. Efforts in DeSci around publishing are, in most cases, responding to the occasionally-misaligned incentives within current academic publishing, such as the lack of incentive for reviewers to provide unbiased and quality reviews, the incentive for journals to publish revenue-generating papers (even at the expense of rejecting epistemically-profitable, but less monetarily-profitable, research), the disincentive to share negative results, etc. [62]. With respect to transparency and incentives, Web3 DeSci projects are experimenting with using tokens (i.e. cryptocurrencies), markets, reputation metrics, and forum-like platforms. Increasing accessibility involves using decentralized storage networks to store published products (e.g. papers, reviews, datasets, code). This section hints at the direction of a DeSci approach to publishing and peer review with a discussion of current DeSci projects dealing with academic publishing and how blockchain-based technologies could potentially fit into the picture.

Since 2017, a number of authors have proposed blockchain-based systems for peer review [36,63–68]. Various works such as [36,67], [68] and [66] propose the use of blockchains to make the peer review process transparent and incentivize good behavior through reputation points recorded on the blockchain, though they differ

in details, with [36,67] and [66] focusing on the entire pipeline even including publication. Below, we discuss several of these systems.

Avital [64] focuses on the motivation for a blockchain-based peer review market, where authors hire reviewers, and all participants interact through smart contracts on blockchains. Ants-Review builds on this idea of a blockchain-based peer review market [63] and outlines an implementation that uses Ethereum [69] and the InterPlanetary File System (IPFS) [34,35]. In Ants-Review [63], an issuer (i.e. author) creates an AntReview for a paper. The AntReview specifies the paper's issuers, file address (on IPFS), requirements, deadline, and approver. The approver is responsible for approving and paying reviewers (using the funds of the AntReview) for their reviews. A reviewer submits a review by uploading it to IPFS and sending the review's IPFS address to the AntReview for the approver's assessment. Anyone can add funds to an AntReview. These funds can be distributed only by the approver and can be sent only to reviewers. If the funds aren't spent before the review's deadline, those who contributed funds are refunded. Note that, in AntsReview, all parties are pseudonymous, associated only with their Ethereum addresses, and that the use of IPFS and Ethereum means everything is open access.

With Reddit-like features for curating, discussing, and incentivizing research, ResearchHub [66] is an even further departure from the current peer review and publishing system. Its users can upload, comment on, summarize, upvote, and downvote papers and posts. Users are rewarded RSC, an Ethereum-based token, for these activities. Papers and posts are aggregated within "Hubs" which function like journals or subreddits. Allowing ideas to circulate in the open, ResearchHub shuns the traditional peer review model of waiting until a paper is mostly or entirely finished before having a small number of reviewers comment on it. ResearchHub relies on the community and the incentives created by the RSC token to create a system in which the better a post or paper is, the more it is seen and its author(s) rewarded. ResearchHub also plans to create a feature allowing researchers to raise funds from others in RSC. The platform differs from other projects in the DeSci space in that it does not currently rely much on peer-to-peer, Web3 protocols, such as Ethereum and IPFS.

At this point, to better grasp the opportunities that Web3 protocols have for publishing and reviewing in science, we might benefit from a brief look at the protocols themselves. Affordances for increasing transparency and accessibility are primarily in decentralized storage networks, such as IPFS and Arweave [70]. IPFS, the InterPlanetary File System, addresses files with content addressing based on cryptographic hashes. In practice, this allows two things. First, a file can be stored on any IPFS node in the world and can still be accessed by anyone who has the file's address. Second, an address can never correspond to more than one file. These properties mean that any file uploaded to IPFS is available to anyone with an internet



connection (this increases transparency and accessibility), and that a file cannot be changed without its address also changing (this increases transparency). Arweave increases transparency and accessibility by providing permanent and open storage. It does this with a blockchain-inspired data structure called a blockweave and a proof for the network's consensus mechanism called Proof of Access in which, to receive a reward, a miner must prove it is storing some randomly determined previous block of data. There are other storage protocols used to store papers and related data, but these examples sufficiently illustrate the common benefits of these systems.

## DeSci Funding and Finances

A more traditional method of funding a DeSci research project is through grant applications. In this case, the process starts with a research investigator submitting a proposal for a particular research endeavor, which is reviewed by the individuals or funding agencies who allocate the awards based on their own criteria determining merit. In the case of DeSci and DAO-based organizations, funding can be provided to the DAO within their internally-defined structures and further distributed within the DAO in order to fulfill the initiative of the grant. Alternatively, stakeholders or investors may provide funding due to their initially genuine research interest or other incentives, and ultimately become detached from the original research motive, thus allowing the DAO to distribute resources to fulfill emergent agendas.

One alternative approach for funding DeSci is the creation and maintenance of markets for projects (bounties) or the outcomes of projects such as data or intellectual property. When an organization/individual is interested in conducting research based on interests, a systematically-managed decentralized exchange for task performance can play a pivotal role. The marketplace could have research proposals by organizations/individuals and researchers as well. A base level requirement for any market-supported project would be to produce some monetary or other form of value investors agree on. Key to the synergetic researcher-investor dynamic would be finding an alignment across multiple aspects of the project such as focus, constraints, cost, timelines. The machine learning platform Kaggle implements methods similar to this, where only researchers that perform best according to the project-defined metrics are rewarded.

Once a project has begun, it can benefit from mutual researcher-investor direction and redirection. Regulation of such direction is imperative; DeSci could build on top of existing blockchain infrastructure with a few modifications. For instance, a structure known as a DAICO (DAO+ICO [71]) proposes to lock all the proceeds of an ICO (Initial Coin Offering) into a decentralized autonomous organization (DAO) smart contract and put the governance over that DAO in the hands of the investors. This locking needs to be consensual from both ends and needs to be recalibrated as

projects change. A roadmap with the specific, measurable, achievable, relevant and time-bound targets can be agreed upon; the investor gets the required output and the researchers get a well-defined source of funding. Other than direct funding, people interested in the project can involve themselves passively, forming a niche supporting or keeping a watchful eye on the project. Various options for this type of implicit contribution can help build a community, with social outlets such as newsletters and citation value networks. These communities can contribute collectively in terms of contributions to reduce the reliance on initial major investors and even foster the project in the case that investors pull out, hence decentralizing the investor influence. One particular project in this regard, Science Fund [72] (Figure 1), allows donors to contribute to science funding pools opened for a specific topic with an amount of their choosing. In return donors receive a NFT receipt (a "Science Funding Token", or SFT).

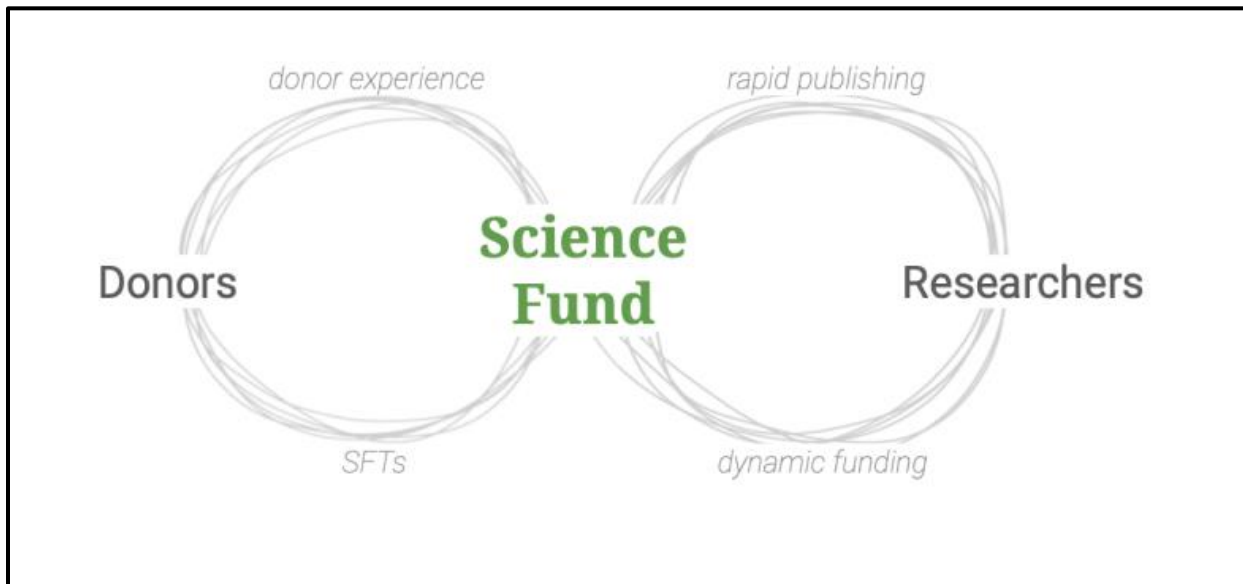


Figure 1. Dynamic research funding model of Science Fund [72].\

These mechanics allow for the following affordances:

- **Curated Funding Pools:** Each donation is added to a chosen funding pool. Researchers can apply through an open protocol with their track record from early published groundwork. As new results emerge from the funded work, the NFT's metadata is updated.
- **Impact Tracing:** Donors have a much more significant participation and visibility into the projects they helped fund. Their SFTs are a personal key to unlock a continuously-evolving experience data trail about their donation's impact.

- **Future Participation:** The NFT receipts become the key to future interactions with the scientists they supported, with evolving forms of participatory governance in the field(s) they contributed to.
- **Funding allocation:** When researchers apply to one of the topic-specific funding pools, they are not coerced to reveal all their invaluable insight upfront. They simply need to share their most influential scientific work (any already-published scientific artifact they created, such as a publication, preprint, or even just a cool dataset) to a given pool. Selected scientists' works then form the basis of a growing evidence trail that is matched to the donor's minted NFT. The NFT receipts become the key to future interactions among scientists and donors, and open a universe of endless new possibilities of mutual exchange and future interactions.

## DeSci Data and Knowledge Markets

Web3 tools enable markets for data storage and compute services that could be used by DeSci advocates to develop a "knowledge market" or decentralized data commons [73]. Agents in this potential market exchange data and other services such as data analysis and model training. Note that, because of their digital nature, such markets welcome opportunities to integrate human and machine agents, allowing for a significant increase in automation. While the base layer here is a data marketplace, it is plausible that future developments could emerge layered on top of this, due to the incentives and informational integrity provided by the market. This section covers the base layer of a data marketplace and how it encourages certain beneficial developments. In particular, data marketplaces have the potential to facilitate data archival, interoperable data standards, data discoverability, and increased automation, and they provide an easier way for industry to build on and fund work in academia.

The base layer of a DeSci knowledge market is first a data market. Multiple blockchain-based data marketplaces have emerged [74–76] since the Ethereum mainnet launch in 2015 [69]. Blockchains provide a number of benefits over traditional data markets, principally the ability to create datatokens, a construct pioneered by Ocean [74]. A data token is, just like any other blockchain token, a digital asset. However, unlike regular tokens, datatokens can be used to access data services. Consuming a data service involves either, 1) being sent the associated dataset, or 2) having the data provider run a compute job on the dataset. This datatoken construct is significant because it allows data services to be purchased on

decentralized (blockchain-based) markets. Anyone can publish a dataset as a type of token, and anyone with sufficient funds can purchase one such token to consume the data service. The datatoken construct is even more valuable within the context of another mechanism developed by Ocean, namely, one that incentivizes the curation of quality data: Users can lock-up funds to signal support for a particular token (called staking) and earn a portion of the transaction fees associated with purchases of the token. Rational actors will seek out and stake on high quality datasets to receive a higher portion of the total fees, thereby curating quality data.

While this base layer is only a curated data market, it indirectly incentivizes the development of additional components relevant to a DeSci knowledge market, such as an increase in archival of scientific data. With a marketplace, data collectors have the opportunity to be financially rewarded and otherwise recognized for simply storing and selling their data. With the transparency of blockchain-based markets, it would be easy to keep track of how many times a researcher's or lab's datasets have been requested, allowing for the emergence of reputational metrics associated with data. It has been observed that currently, "there is little incentive to invest time in archiving or repackaging data sets" because "investing time in a project beyond its usefulness for publication is counterproductive, given the high expectations for producing research publications" [77]. Filling this incentive gap, a data marketplace could provide another avenue of being rewarded and recognized for the valuable contribution of collecting quality data, and could therefore lead to increased data archival and availability.

Interoperable data standards are incentivized by a data market. The benefits of interoperable data standards can be seen in, for example, the neuroscience community, with its adoption of BIDS, "a standard for organizing and describing MRI datasets" [78]. This standard allows neuroscientists to more effectively collaborate within and across labs. It has reduced the amount of time they spend reorganizing datasets, leaving them more time for productive scientific work such as data analysis. In a data marketplace, where data suppliers are in competition and where the data users prefer certain formats over others, the data suppliers can be relied upon to provide data in those preferred formats, given sufficient competition. Even if a dataset is originally published in an esoteric data structure, someone can profit by purchasing it, organizing it into a known format, and reselling it. In a market where data purchasers may buy data from varied sources to improve their models, they are likely to seek out datasets that are compatible with each other. That is, they will seek out data structured in interoperable formats. The result is a data marketplace in which suppliers are actively organizing data to suit the needs of purchasers, converging on interoperable data formats.

Another potential emergent benefit of the data marketplace is increased data discoverability. Currently, finding scientific data is cumbersome [79]. Researchers

often rely on social connections and fragmented data repositories. A data marketplace helps to address this issue through the competition between data suppliers who can gain from increasing the discoverability of their data. The more discoverable a dataset is, the more likely it is to be purchased. Data suppliers can increase discoverability by, for example, adding relevant tags and other metadata. Tags can help data search engines sort datasets. Metadata might even include a cryptographic signature from the researcher(s) who collected the data, attesting to the source and quality of the data. Additional organization and discoverability of datasets might come in the form of a knowledge graph, in which datasets can be related to each other, to researchers, to papers, to code that has been used to analyze it, etc. Note again that the data marketplace can be read and operated on by both humans and machines; much of the sorting of datasets in the marketplace might be done by automations created by data suppliers. In the end, the suppliers within a data marketplace have a strong incentive to make their datasets easy to find, so they will likely adopt whatever methods best accomplish this.

Moreover, a data marketplace where data are curated, highly available, structured in known and interoperable formats, meaningfully-tagged, and discoverable, provides good conditions for an increase in automated scientific discovery and, more specifically, better conditions for meeting the Nobel Turing Challenge introduced by Hiroaki Kitano [80] and further refined by The Alan Turing Institute [81]. This challenge involves creating "AI systems capable of making Nobel-quality scientific discoveries highly autonomously at a level comparable, and possibly superior, to the best human scientists by 2050" [81]. Smaller challenges within this grand challenge include ensuring data are shared, linked, and machine-readable. The data marketplace described above could help address these and possibly other challenges, such as simply financially incentivizing researchers to make their valuable datasets available.

Finally, a data marketplace affords profitable interactions between academic labs and organizations in industry. Labs can sell data to companies that need data and have AI expertise but that might not have the expertise or equipment to collect such data. This not only allows companies to more easily find and use scientific data, but by buying such data, companies would also be funding the scientists who collected them. This opens up a market in which data with both epistemic and pragmatic value can be easily priced and discovered. It also makes it easier for companies and the economy at large to communicate their practical data needs to researchers; this could enhance the current system in which a study's perceived pragmatic value is a function of how well its value is communicated through a grant proposal. A data marketplace could lead to a greater flow of information among academic and non-academic (e.g. citizen, government, non-profit, profit-oriented) organizations, with respect to the value of various kinds of data, allowing easier and potentially better funding of data with pragmatic value (though these benefits of funding may not

extend to data which has only epistemic value). A well-designed data marketplace enables fair price discovery of digital knowledge assets and encourages productive applications of meticulously-collected, high-quality scientific datasets outside of basic research, boosting research applicability and downstream innovation.

## Teamwork Modeling with Systems Engineering

To decrease the complexity and uncertainty that individuals face within large-scale DeSci projects, it is useful to introduce the systems engineering approach [82,83]. This process of sensemaking for individuals and teams can be thought of as the use of a map, enabled with modern features like personalization and dynamic rendering [84]. For better communication and team attention management, it is important to have consensus on the team objectives, project lifecycle management, practices in use, and metrics for tracking the performance and relevant behavior of all team members. Targets and goals of the projects are set by the needs of external stakeholders and individuals fulfilling specific project roles. Governance for decisions, assignments, and resources are based on unequivocal agreement according to the team members' affiliation and authority to manage each other's labor. We consider such a group of individuals as an organization or organizational unit in the case of specific tasks to resolve.

Akin to the use of Systems Engineering in the building of complex infrastructure (e.g. bridges, planes), we propose that people involved in DeSci projects document the roles involved in these team dynamics, as well as their concerns and preferences. This allows DeSci activity to be structured according to an established lifecycle with needed practices [85,86]. Here, we are not concerned with the domain-specific practices of a project lifecycle, but with general practices that are relevant for DeSci development, such as social modeling, economics, governance, and development in a broad sense. During any DeSci project, we can associate a human or other entities as actors who are fulfilling the roles that need to be performed. These entities fulfilling project roles engage in practices that entail completing tasks or cases. All the relevant objects and open cases must be resolved during the project for it to be a success (for instance, knowledge artifact production or fundraising). These cases need to be documented with reference to the roles, practices, and actions required for successfully achieving the target goal. This case/issue tracker functionality is an attention management tool for the people involved in the project. This process allows for productive, auditable, and coherent work on projects.

Like CeSci, DeSci faces challenges and failure modes across multiple domains, some of which were briefly described above. At this incipient and formative stage, what may provide utility for DeSci would be a useful unifying framework to allow for system conceptualization, design, and implementation. In the following sections, we



describe a possible starting point for the integration of DeSci and CeSci systems, through the combined application of Active Inference and Systems Engineering [83].

## BOLTS of the CeSci to DeSci Integration

The difficulties of successful implementation of novel systems, even where they are simple or viable to be implemented in parallel rather than as replacements to extant systems, lend themselves to underestimation. Successful, ubiquitous systems often feel obvious in terms of their value proposition and their reasons for adoption. For example, the product "bar code" would appear to be ubiquitous and obvious in terms of its benefits, but despite its conception in the mid-20th Century, by 2004 only around 80 to 90% of the United States' top 500 companies had adopted them [87]. In terms of its roughly 50-year climb to this broad adoption, it took a quarter of a Century just to find adoption within a single market segment after a small collection of large institutions began courting potential standards [87,88]. The difficulties of implementation, adoption, and integration of novel systems relevant to our purposes can be summarized through a qualitative use of the Active Inference Conflict Model, which helps to frame conflict across business, operations, legal, technical, and social (BOLTS) surfaces, especially where it relates to information differentials [89]. Below, the bar code is used as an example that characterizes how the BOLTS surfaces apply to complex patterns of adoption and operation:

- **Business.** Integration with and replacement of extant receipt and payment systems, few of which would have been compatible with any bar code standard, represented a significant difficulty. Additionally there are the financial and non-financial costs of barcode adoption that an organization would have to consider.
- **Operations.** Integration with and replacement of extant inventory systems, few of which would have been compatible with any bar code standard, represented a significant difficulty.
- **Legal.** For some products, technical standards and instantiations of product IDs would have to map to regulatory compliance systems, contracts, and be available for use in advanced records systems.
- **Technical.** While the idea of the bar code was conceived in the mid-20th Century, it took decades to confirm widely adopted standards. Adoption at any given point in



the climb to broad adoption meant the risk of implementing a pattern that might become obsolete.

- **Social.** People didn't necessarily trust machines, and were concerned with the potential for pricing errors at the time-of-scan. Companies were also concerned with how bar-codes would affect the shelf-appeal and artistic design of products.

This "bar code" example helps to illustrate how the implementation, adoption, and integration of new systems is roughly proportionate with the abstract and real integration conflicts with extant systems. While modern, shared technological infrastructure has allowed for more expedient implementation of new systems, conflicts across BOLTS processes still exist and may be greatly exacerbated when systems are intended to be used by institutions and organizations, which have their own formally-defined BOLTS protocols, in addition to individuals where BOLTS protocols are likely to be encoded in norms, habit, and narrative [89].

While DeSci may be implemented in parallel to CeSci, the overwhelming majority of research funding is channeled through large organizations, foundations, and especially government agencies. Each of these and their respective channels are beholden to a complex, complicated, and interconnected web of BOLTS standards, norms, and controls, which are a struggle to navigate. While it is tempting to make do with what funds might be available, a bridge between CeSci and DeSci is not just about access to funds at scale but also about access to tangible resources, influence on institutional funding and agendas, ability to collaborate, and avoiding scarcity in funding. In addition, solving conflicts relevant to a DeSci-CeSci bridge may lead to greater accountability, reliability, and impact of DeSci information products, and therefore improve the ability for potential investors to trust in the new marketplace of ideas. Areas of tension and interoperation between DeSci and CeSci are summarized through the use of BOLTS below:

### Business

While government agencies and non-government organizations are not operating under the return on investment model found in for-profit businesses, they still consider the dollar cost of impact, opportunity costs, and likelihood of impact. Agencies have to consider how to structure "portfolios" of funded projects and initiatives and do so based on their current mission, the missions of their stakeholders, their sources of funding, and the track-records of the organizations and people they fund. In order to make DeSci compatible with or even

superior to CeSci projects and organizations in this context, DeSci systems would have to enable a high level of situational awareness for funders in terms of both progress and potential for impact, while offering advantages in terms of cost and comparability. This may not be immediately achieved – like most new services and systems, DeSci may start off with inefficient processes, and will become cheaper and more efficient in producing informational products of higher quality over time.

### **Operations**

Accountability and use of funds is taken very seriously in many of the CeSci funding channels, especially where funding flows through multiple hierarchies. Even where funding flows through less steps, such as in NGOs, demonstrating success in order to raise funds constitutes its own stream of work requiring a substantial allocation of attention and funding. Documentation and records management in operations serves to fulfill the "evidential function", pro forma evidence of action and use [90]. These records should be systematic and need to meet relevant standards for reliability, integrity, compliance, and comprehensiveness [91]. It is here that DeSci offers challenges and opportunities relative to traditional systems [20]. Where CeSci organizations rely heavily on manual processes in order to meet documentation requirements, DeSci is predisposed toward generating records by merit of reliance on formalized, computational protocols. However, these advantages will only be realized if there is a directed effort to ensure automatically generated records are adapted to map to the myriad CeSci requirements.

### **Legal**

Many of the business-, technical-, and operations-related documentation requirements extend to legal use. It is in this domain that records serve their "warranty function", clarifying intents, deliverables, requirements, and other agreements [90]. Project documentation and both personal and organizational identity become key obstacles to DeSci's compatibility with CeSci, especially where teams are emergent or may change in structure over time. Extensive work is required in this domain in order to overcome legal compatibility challenges, given that no amount of advantages of efficiency, cost, quality, or impact

can allow DeSci to circumvent problems related to CeSci funders being able to exercise their options related to legal recourse.

### **Technical**

Technical aspects of CeSci funding blend with operations and legal aspects, as would be the case with most bureaucratic systems. The synthesis between these two sections (operations and legal) is where the potential bridge between DeSci and CeSci is currently most precarious, given that an inability to meet legal requirements will eliminate this bridge entirely. However, it is also where the bridge has the potential to outperform traditionally-funded organizations, as meeting operating and legal requirements computationally would, as previously suggested, offer a number of advantages.

### **Social**

The proverbial "pipeline" of government and institutional funding to research and development is monolithic and internicine, and at times quite obscure [92]. Those involved in the channeling of funds have standard operating procedures encoded both in terms of business, operations, legal, and technical surfaces, as well as in social norms and in narratives. The preference for DeSci to offer more accountability or other efficiencies for research is compatible with the preferences and requirements of those who currently occupy positions which channel funds in CeSci. Case studies, post mortems, and high levels of visibility in both successes and failures, as well as test-runs in incubators or regulatory sandboxes, could assist in communicating the benefits of DeSci to potential stakeholders. Further, finding ways to allow DeSci systems to be flexible enough to map directly to CeSci standards for accountability could also facilitate DeSci integration. For example, even though DeSci allows for and can encourage purely horizontal teams, one can start with teams which can: assign a higher authority and offer them power of direct oversight, allow them to measure and assess compliance and performance, and be responsible for reporting and conveying that information to funders (even where this could be automated) [93] to help to introduce CeSci standards to DeSci without requiring excessive accommodation.

## BOLTS and Reliability for Science Systems

It is important to provide reliability in all of these BOLTS dimensions, as new technologies and systems can easily fail if there are repeated bad experiences or unexpected incompatibilities. Even if the technical and legal aspects of DeSci funding out-perform CeSci funding, failure of DeSci to provide a smooth integration from the social and business perspectives could delay or prevent wide adoption. For example, traditional funding agencies often come in at middle-to-late stages of research projects, which is why so much preliminary work is required for a NSF/NIH grant. It is often said grants are not “for ideas”, but rather they are for seriously-developed avenues of specific tractable research. In DeSci, some more speculative vehicles might exist (e.g. a “Learner’s Fund”), but we cannot expect these new vehicles to back-propagate the risk-tolerance of DeSci into legally- and organizationally-bound CeSci funding agencies.

A priority of both CeSci and DeSci is to ensure reliability of outcomes and high levels of accountability, in complex scientific commons. If a team has been funded, disintegration or failure to produce work doesn’t just mean a negative outcome for the team, it means a negative outcome for the funder, which has the potential for network impacts on trust within the entire market. This means that teams within the DeSci space would not only need to vastly outperform CeSci in terms of cost, time-to-impact, and efficiency, but would also need to effectively be perceived as high-reliability organizations in order for the market to survive. Given that DeSci teams are likely to be emergent, they may not be able to rely on intimate trust between one another, and instead will have to rely heavily on shared Ontology, Narrative, Formal documents, and Tools [83]. The team working environment should be designed to create the best possible conditions for the cognitive security of the individuals on teams, such that they stay on task, maintain consistency in terms of goals and intents, and understand their impact on the team and the future of the market as a whole. This type of system design also means high cognitive security on the part of the funder, which means high visibility and high quality narrative information management tools at their disposal [38], the ability to do retroactive analysis of track-records of individuals and collections of individuals, the ability to rapidly compare projects, and the ability to negotiate cost as a basis to reduce risk based on current information. Additionally, transparency ensures that when things do go wrong, it reduces the negative sentiment of the funder on the network, and might be mitigated by insurance options (e.g. community stake in research) or in-house recourse options (e.g. sanctions and penalties on negligent and bad-faith actors)

While we detailed the BOLTS conflicts related to funding processes as DeSci practices are integrated into existing CeSci practices, this framework can be applied to several other aspects of this integration such as the publication of knowledge artifacts, scientific review of knowledge artifacts, communicating research results at conferences, analyzing data, training and mentoring new researchers, and developing scientific software. CeSci and other fields already have successful aspects which should be preserved as DeSci is integrated, such as specific mentorship, individual, and peer development programs. As a framework that furthers the holistic design of transdisciplinary systems for science, we turn to Active Inference for additional insight into DeSci which will be outlined in the following section.

## Active Inference, Systems Engineering, and Science

In this section we justify the use of Active Inference in systems modeling, introduce an entity description for science modeling consistent with Active Inference and Systems Engineering models, and revisit patterns in CeSci and DeSci.

### What is Active Inference?

Active Inference is a framework for the integrated modeling of entity perception, cognition, action, and impact in the niche. Active Inference draws upon the formalism of the free energy principle, which provides a single statistical imperative underpinning the assembly of organic and inorganic matter: the minimization of informational free energy [94–96]. This principle applies to systems at every scale. The process of free energy minimization has been positioned as the driver of autopoietic processes in living systems ranging from cells to cities [97,98]. Its core meaning is that it allows a teleonomic, nearly-agentive account of self-organization. For example, a set of interacting systems may be likely to spend a lot of time in an aggregated form, reflected by persistence in a low free energy region of their coupled dynamics corresponding to an aggregated state [95]. The minimisation of free energy at the system's scale entails the minimisation of variational free energy for each of its constituting particles, which can be interpreted as the optimisation of a Bayesian model entailed by the particle's structure (more specifically, the difference between model Complexity and Accuracy) [99]. Because of this correspondence, we can claim that entities aggregate in virtue of reducing their uncertainty about future sensorimotor states [100,101].

The related framework of Active Inference leverages this formalism to model one or multiple interacting entities as involved in continual perception and action angled

towards reducing the expected uncertainty about their sensorimotor flow [102,103]. They can do so either by altering internal states so as to infer the causes of sensory states, or by altering external states so as to cause expected states. Both processes participate in a common control architecture based on predictive processing, which can be interpreted as a hierarchy of "algorithms" matching input to output over nested time scales [104]. In this hierarchical model, previously-acquired beliefs about controlling behavior in different contexts are internally encoded, and result in adaptive arbitration between epistemic and pragmatic value choices (i.e. exploration vs exploitation, [105]). Thus, the framework of Active Inference allows us to model the observations, beliefs, and predictions (i.e. expected states) of agents as statistical dependencies, or Bayesian probability distributions.

Much of the focus in the Active Inference literature has focused on organisms, especially in the context of human behavioral neuroscience [106]. However, Active Inference has also been applied to various other informational and cyberphysical settings such as human communication [107], remote teams [83], human conflict [89], trust interactions with robotics [108], and epistemic communities [25]. Here we apply Active Inference and Systems Engineering to the case of modeling centralized and decentralized science.

## Active Inference and Science

Evidence of collective intelligence, understood as the capability of human societies to solve problems, surrounds us. In most cases, it is instantiated in cultural artifacts or cognitive gadgets [109] that follow from a process of cumulative cultural evolution dating from the branching of the Homo genus [110,111]. We are surrounded by living fossils of this cumulative process, from smartphones to hospitals, from net fishing to general agreements about the recognition of private property. Our capability to understand and navigate the world does not simply follow from information in our minds, but it is (to some extent) entailed by the very structure of the ecological niche and the social constraints that shape our behavior. This means that entities are, to some extent, autonomous or agentic regarding their own cognition [29]. Entities engaged in Science as a cognitive endeavor are no exception.

Like other instances of collective intelligence involving stigmergy [112,113], Science is a cumulative endeavor that relies on previously-acquired beliefs and established communities of practice. Scientists and other cognitive entities perceive stigmergic cues from the niche (including epistemic resources, as well as other social entities), update their cognitive models (learning), and act to modify the niche in the form of research outputs (Figure 2). As scientists follow new leads or choose to ignore them, communities form coalesce and disperse, and scientific theories are created, developed, and forgotten. Frequently, scientific endeavors are undertaken with

multiple, potentially competing, goals in mind. For example, scientists may undertake their work with an aim to achieve some or all of the following outcomes: earn a paycheck, improve their reputation, improve the reputation of their company or institution, conduct empirical research without bias, solve a particular problem, or discover something novel, simple, or beautiful.

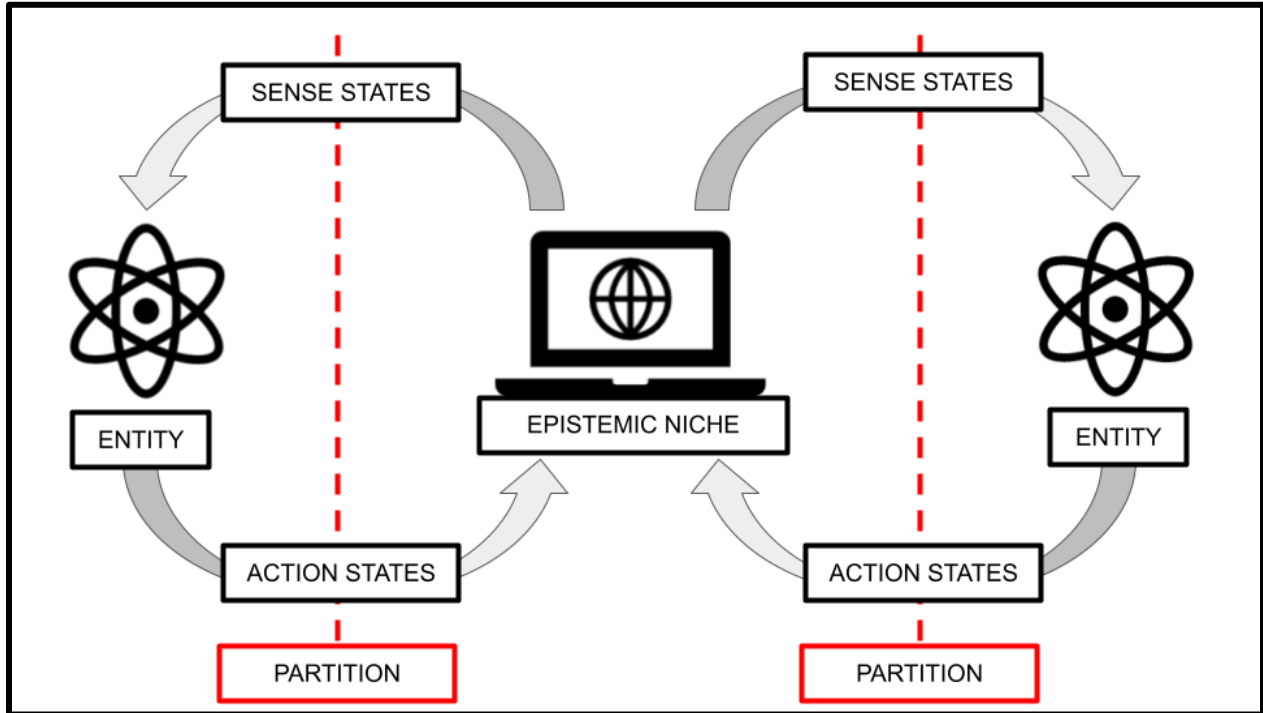


Figure 2. Two active entities interacting via their shared epistemic niche. For more detail on the entity perception-cognition-action loop in Active Inference, and the Markov Blanket formalism comprising the partition shown here, see [95].

Active Inference is relevant for model-based understanding of scientific systems because it offers a composable multiscale ontology and integrative framework for entity and system behavior [98,114]. Thus the usage of Active Inference can help us understand the entanglement of multiple processes occurring at nested scales of behavior. In Active Inference, actions across scales are modeled as a systematic attempt by cognitive entities (i.e. scientists and institutions) to reduce uncertainty about aligning their future perceptions with their preferences, so as to ensure their persistence of their defining organization. For example, the development of administrative systems and agriculture was described within Active Inference formalism as a product of City-States' attempt to understand and control their reality [29], analogously to the process of morphogenesis as a process of cells manifesting a body plan which reduces their uncertainty about their cellular niche [100].

Importantly, the Active Inference account is not limited to conservative (homeostatic) behavior, but also to more complex phenomena such as memory,



anticipation, abductive logic, and autopoiesis. For example, Active Inference has been used to model the systematic alteration of one's niche [104,115,116], as well as the creation of new tools and their incorporation into one's cognitive identity [100]. Therefore, it very much provides grip on innovation, by contextualizing subjective systems of meaning in relationship to the concrete (biological and sociocultural) organizations which underlie cognition [117].

In the context of action selection related to research (e.g. experimental design, decisions about scientific communication), Active Inference provides a division between two kinds of motivation in research: epistemic value (informational gain related to e.g. new research findings or knowledge) and pragmatic value (reward associated with increases in wealth, longevity, or status). Any innovation in institutional or cognitive processes will affect the broader system both epistemically, by defining the social norms which entail scientific activity, its cognitive underpinnings, and the kind of cues and actions scientific activity involves; as well as pragmatically, by reinforcing the specific behaviors which help the relevant (institutional or human) agent to reduce (subjective) uncertainty about future sensorimotor states. When viewed through this lens, the scientific process balances action selection based upon the both epistemic and pragmatic value, and this process plays out across multiple scales of nested organization. For example, at different stages of their day/career, the individual human researcher prioritizes more exploratory/learning activities, while at other times prioritizes more exploitative/productive activities. At the scale of an individual, institution, or even scientific field, the epistemic activities are reflected by basic research, while the pragmatic activities can be considered as translational or application-oriented research, or engineering.

As metascience and sociology of science typically focus on the motivations and systems of incentive on the scale of individuals and institutions, we consider the multiscale nature of Active Inference to provide a relevant angle for future research. This work reflects early steps towards leveraging this framework into a more detailed understanding of how scientific entities interact, and how contemporary changes associated with the challenges and opportunities of DeSci could improve on the status quo.

# Active Entity Ontology for Science (AEOS)

Here we describe the attributes and structural basis of a descriptive entity-oriented ontology for Science, consistent with the principles of Active Inference and Systems Engineering. This project has the working title and acronym Active Entity Ontology for Science (AEOS). Previously, among many other use cases, ontologies have been used for schematic mapping of cryptographic systems and knowledge ecosystems [118]. Additionally, approaches for complex systems modeling based upon entities and their affordances have precedence in economics [119] and other areas. We build on this work to create a useful and versionable resource which stands to be developed and extended by researchers and practitioners in the coming years. This work is meant to be a starting point for the construction of active entity-oriented ontologies of science. See [coda.io/@active-inference-lab/active-entity-ontology-for-science-aeos](https://coda.io/@active-inference-lab/active-entity-ontology-for-science-aeos) for an interactive site that provides an updated version of AEOS. Table 1 provides example entity interaction motifs for CeSci and DeSci ecosystem models. These motifs are presented as initial drafts on the specification and visualization of subsystems/modules that are important for larger networks of interacting entities.

| Area of Concern                   | Example CeSci Motif                                                                                                                                                                               | Example DeSci Motif                                                                                                                         |
|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Funding</b>                    | Team Publishes Knowledge Artifact (Grant), then Requests Funding From Funding Agency, which Scientifically Assess the Grant.                                                                      | Team Publishes Knowledge Artifact (Grant), then Requests Funding From a DAO which Scientifically Assesses the Grant                         |
|                                   | Funding Agency Provides Opportunity for Researcher and Provides Constraints for Researcher and Sends Fungible Asset to the Team.                                                                  | The DAO Engages in governance regarding the Grant and Sends Fungible Asset to the Team.                                                     |
| <b>Communication</b>              | Academic University Communicates Asynchronously To Lab via Computer System                                                                                                                        | Human Communicates Synchronously To Human via the Computer System                                                                           |
| <b>Cryptographic relationship</b> | Research investigator uses Cryptographic Identity on Computer System to Provides Funding To Team and Provides Updates to Funder like Research Agency (Government)                                 | Research investigator Team transfers Intellectual Property as Non-Fungible Asset to a DAO Cryptographic Identity that is a Smart Contract . |
| <b>Scientific Review</b>          | A Research investigator Sends Fungible Asset to Journal/Publisher to Publish Knowledge Artifact like a Paper. The Journal/Publisher Requests Scientific Review from other Research investigators. | After Publishing a preprint Paper, the Research investigator receives a Fungible Asset as a function of the use of the Knowledge Artifact   |
| <b>Research &amp; Analysis</b>    | Research investigators in Lab Team Scientifically Assess Primary Knowledge Artifact then Publishes Knowledge Artifact                                                                             | Research investigators in DAO Community of Practice Scientifically Assess Primary Knowledge Artifact then Publishes Knowledge Artifact      |
| <b>Publishing</b>                 | After Scientific Review, a Research investigator Publishes Knowledge Artifact at Journal/Publisher.                                                                                               | A Research investigator Publishes Knowledge Artifact like a Paper on a DAO-produced knowledge market.                                       |

Table 1. Some example motifs of Research and Funding in CeSci (Centralized Science) and DeSci (Decentralized Science). See the AEOS site for an interactive version of this Table, where all entities and affordances can be unpacked and explored.

## AEOS Entity partitioning

Any AEOS entity consists of 3 kinds of states that are partitioned off from the entity's environment: Internal states (states inside the entity that are hidden from the outside, for example those involved in cognition), Sense states (incoming statistical dependencies), and Active states (outgoing statistical dependencies). Together these three states (Internal, Active, and Sense) constitute the "particular" or autonomous states of the entity [99]. One of the advantages of Active Inference as an integrative framework is this: used instrumentally, it can model various phenomena and ensembles across scales from the quantum [120] to the social and planetary [121,122], including heterogeneous types of entities which enact the scientific ecosystems of today and tomorrow.

Using the Markov blanket partitioning of variables on a Bayesian graph: Entities consist of specified Internal states, reflecting the (observer's model of) the Entity's generative model of perception, cognition, and action. External states are those that are inferred or acted on by Entities. Internal states are partitioned from external states via a Markov blanket, which is further divided into nodes with incoming statistical dependencies (Sense states) and outgoing statistical dependencies (Action states) [95,98,99]. Active states are reached via an internal process of policy selection, which incorporates information about what affordances (capacities for action) exist and the preferences that the entities would like to reduce uncertainty about the realization of (for example a preference for comfortable temperatures or financial success).

Entity models of this type have advantages in that they can be compared across entities and niche surroundings [115], and enable variational Bayesian inference and message passing algorithms [123]. In this case of online DeSci, a type of digital message passage algorithm is occurring on a bipartite graph composed of active and informational entities in the AEOS. This may be represented as a factor graph, which presents with an interesting avenue for future research and application [123–125].

## AEOS Entity Classes and Types

The fundamental object or kind of thing in the AEOS is the Active Entity, which in online settings is almost always interacting with Informational Entities [83]. Drawing upon the Active Inference framing of a system, this entity is a system of interest that is partitioned from its niche surroundings based upon its capacity for persistence, thus making it a "thing" [126,127].

In AEOS, there are two Classes of Entity (Figure 3): Active epistemic entities (which select action policies and enact affordances), and Informational entities (which contain information and are acted upon). There are multiple Entity Types within each Class. In the digital/online setting, Active Entities are always intermediated by Informational entities (e.g. People can talk online through modification of a shared epistemic niche reflected by a computer system). Informational entities are always intermediated by Active entities (e.g. two databases can only communicate with each other through a computer system). "Adaptive" and "mere" Active Inference entity notion is via [128].

Both classes of entities (Active and Informational) inherit the same type of Active Inference partitioning scheme described above. See the AEOS site for an updated list of the diverse types of entities currently modeled. Some examples of Active entities in AEOS include Human, Computer System, Organization, Team, DAO, and Academic University, Community. Of special note are the Organizational entities, composed of groups of humans, computer systems, or other organizations (e.g. a university made up of departments, made up of labs, made up of people), along with their informational resources. As with any other Active Entity, Organizational entities engage in outgoing actions from a set of discrete or continuous options (affordances), and also can be the recipient/target of action from other entities.

Entities of the Informational class are those which are being modeled for their informational content, and are solely acted upon. Some types of Informational entities may also display agentic qualities when used by Organizational and Computer System entities, for example a software program. The Informational entities are used as epistemic resources in the niche of other Agent-class entities. Some examples of Informational entities in AEOS include Knowledge Artifact, Fungible Asset, Non-Fungible Asset, Dataset, Metadata, Paper, Intellectual Property, Code, Grant, Blockchain, and Project Catechism

## AEOS Policy Selection, Areas of Concern, and Roles

Active entities engage in policy selection based upon their internal generative model (Figure 4). The behavioral policies that active entities enact are selected from a set of possible action sequences for that entity: their affordances. In the AEOS, the affordances are determined by the type of entity, its roles, and its context – for example a human entity with administrator role may have the affordance to modify a file in a certain situation. The specific affordances that an entity engages in, in a specific setting, reflect the role (formal/assigned or informal) that the entity has (Figure 5). This role-based approach towards organization draws on Systems Engineering [83], and allows the flexibility to both describe current systems as well as design new systems. For example if a human is providing grant funding to a team,

that human entity has the role "Funding provider". With respect to the affordance of providing funding to a research team, this role could also have been fulfilled by a funding agency or a DAO.

Below in Figure 6-8 are several initial graphical examples of applications of the AEOS. The natural language sentences at the bottom of each image, reflect CeSci and DeSci motifs to be found within each Area of Concern (Communication, Scientific Review, Funding). Words in blue reflect AEOS entities and active relationships (affordances) that link the entities. These examples show the capacity of AEOS to model various key motifs and patterns found in both CeSci and DeSci systems. One possible advantage of a system such as AEOS is that scientific motifs can be described in natural language, have an interactive visual representation, and also provide an avenue for formal modeling. We are currently exploring the integration of AEOS models with complex system modeling software such as cadCAD [129] or darcspice [130], in order to enable description, analysis, and simulations of epistemic ecosystems.

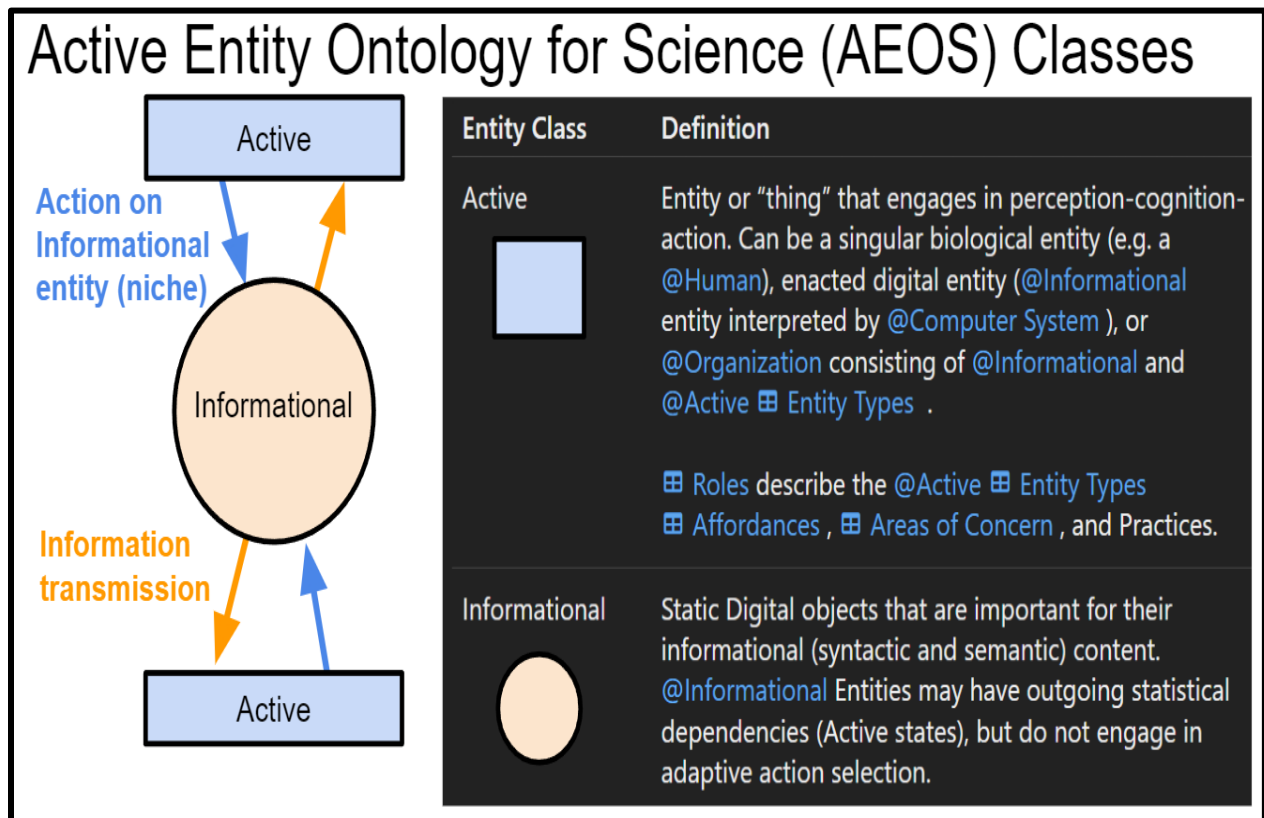


Figure 3. Entity Classes in Active Entity Ontology for Science (AEOS). Active Entities (blue boxes) are any kind of organization with agency at any scale, for example humans, teams, organizations, institutions, and DAOs. Informational Entities (orange circle) is any type of computational data or epistemic resource, for example a dataset, grant, article, or fungible/non-fungible asset.

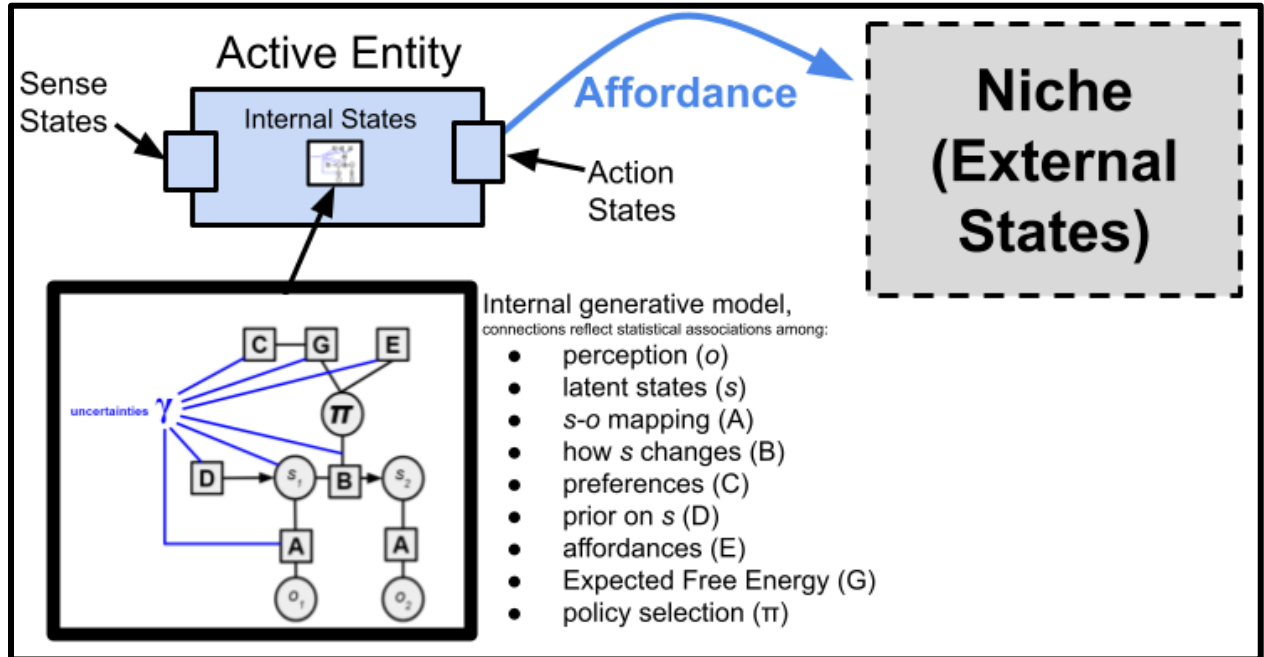


Figure 4. Active entity action selection. Using the Active Inference partitioning [106,127], the Active Entity at top left can be modeled as having incoming Sense States and outgoing Action States which interact in an affordance-based fashion with the niche. The internal generative model of the entity, here modeled as a partially observable Markov Decision Process, describes the process of action selection given an incoming stream of observations.

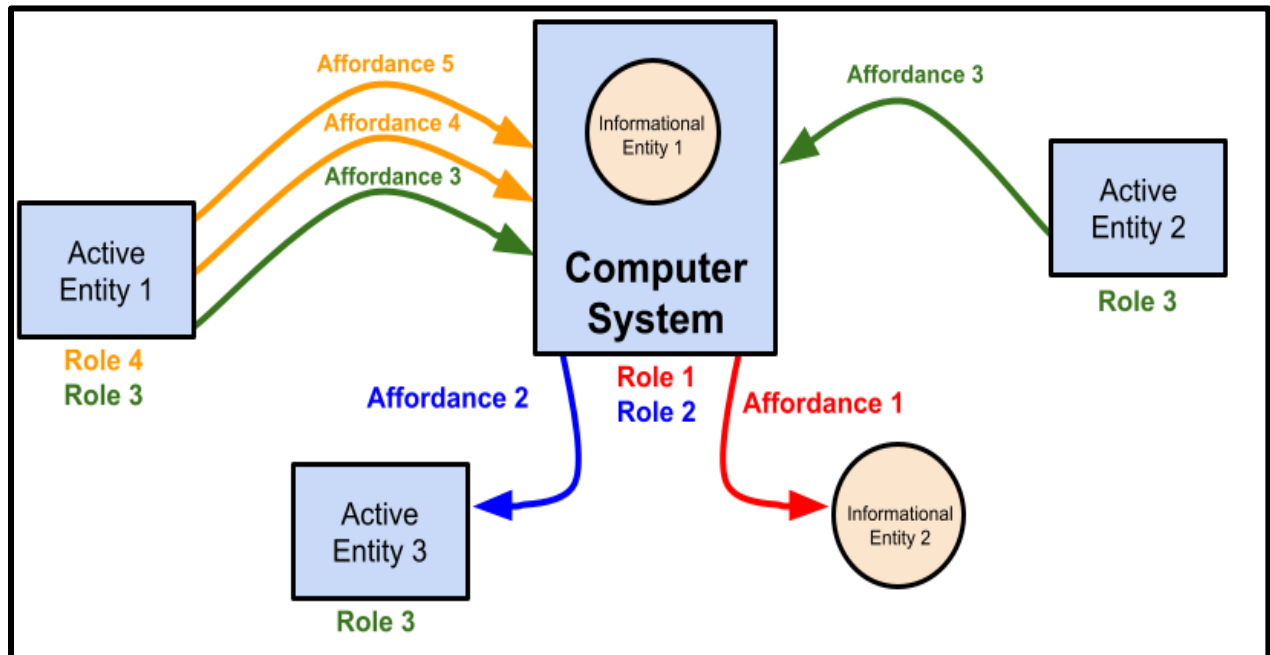


Figure 5. Entity Roles. Each Entity has Affordances (capacities for action) and can have Roles in the organizational context (assigned performances to be done) In the online context, all Active entities are intermediated by computer systems, and implicitly informational entities (shared epistemic niche).



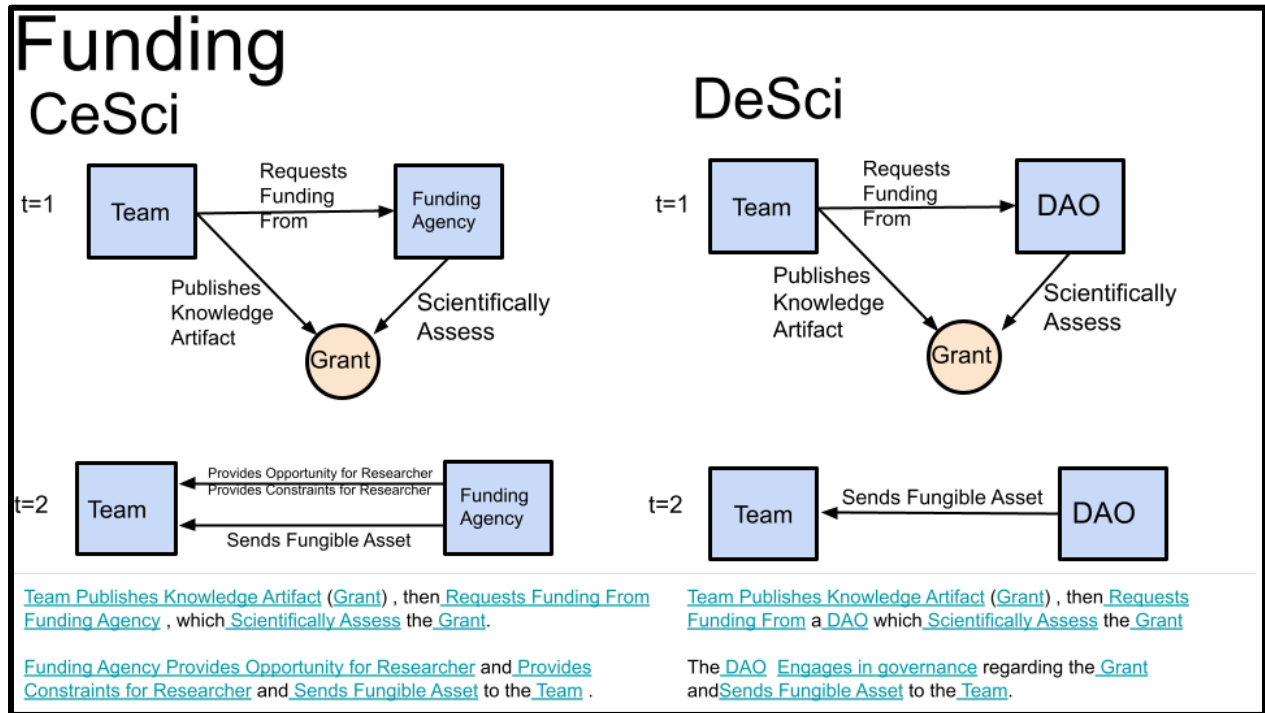


Figure 6. Example motifs of Funding in CeSci (left) and DeSci (right). Blue boxes are Active Entities in AEOS, orange circles are Informational Entities. In this example, the initial timestep  $t=1$  is noted on the top of each side, and the following timestep  $t=2$  is below.

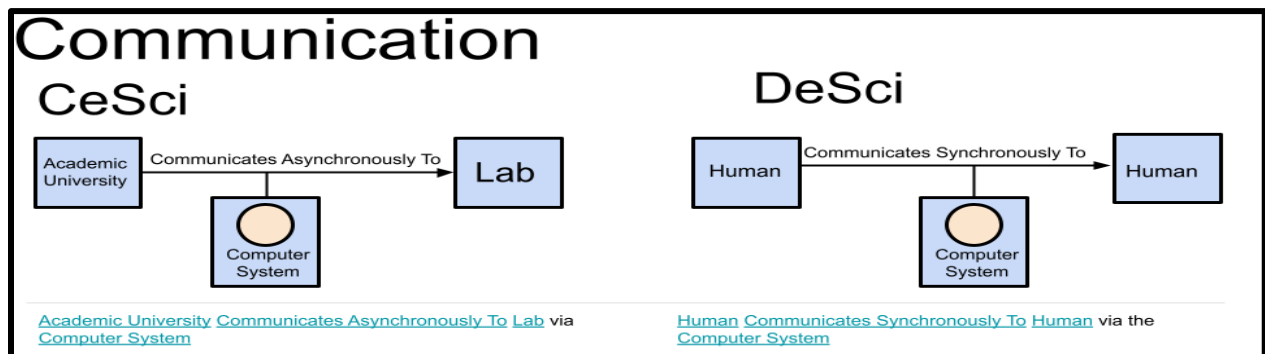


Figure 7. Example motifs of Communication in CeSci (left) and DeSci (right).

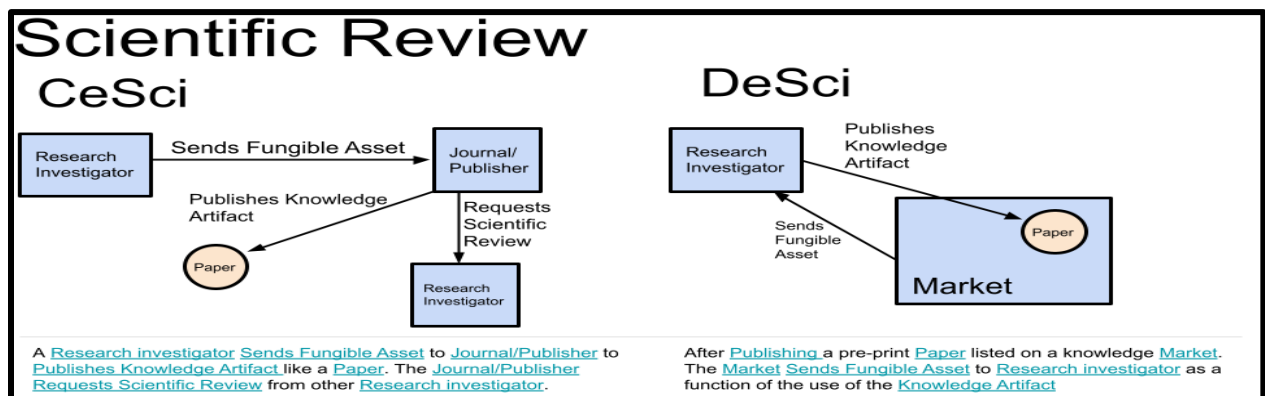


Figure 8. Example motifs of Scientific Review in CeSci (left) and DeSci (right).



## Discussion

This paper evaluated the opportunity to implement decentralized science in Web3 technologies from the perspective of Active Inference.

We first highlighted the distinction between two approaches to scientific research, respectively Centralized Science (CeSci), where research is driven from the top-down perspective of centralizing institutions, and Decentralized Science (DeSci), where it is driven from the bottom up through situated sensemaking by communities of practices. We argued the second stance can facilitate the governance of scientific knowledge (understood as an epistemic common good) by providing both the incentive and the opportunity to build a detailed understanding of various systems and problems in an integrated language.

We consequently reviewed potential avenues to implement DeSci in the era of Web3. We underlied a major limitation to marketization and tokenization, one of the main mechanisms of Web3 technologies, namely that creating real world value for the token renders it vulnerable to speculative investment and hoarding. We discussed the potential roles and benefits of modern affordances, including online teams, blockchain, tokenization, and smart contracts. Finally, we discussed several practical aspects of integrating DeSci concepts and practices into CeSci, such as business, operations, legal, technical, and social (BOLTS) aspects of this transition.

Finally, we discussed the role of the Active Inference framework in describing scientific ecosystems. As a formalism relating the basic activity of dynamical systems to epistemic inference, we argued it is a natural fit to describe science as a socio-institutional system. We outlined an Active Entity Ontology for Science (AEOS), intended to define key Entity Classes, Entity Types, Areas of Concern, Roles, Affordances, and Action-Perception States involved in conducting scientific work, whether CeSci or DeSci, with examples illustrating particular embodiments of this ontology. We hope that further development of research within the scope of AEOS would yield fruitful progress in our ability to design and evaluate possible forms of scientific organization.

Some specific next steps include:

- Implementation of AEOS using complex systems modeling frameworks such as cadCAD [129], which have already been applied to relevant cyberphysical systems in Web3.
- Exploration of Active Inference ontologies and relationships with category theory [120] and other formalisms.

- Creation of Graphical User Interface for DeSci (DeSciCAD), to enable the real-time use of AEOS in science ecosystem design and operation. For example, a user could select entity types, assign them roles, connect the entities with affordance-based edges, and engage in simulation and parameter optimization. The front end of such a design interface would ideally be accessible and graphical, on the backend there could be a composable Active Inference simulation.
- Exploring system designs in AEOS that scaffold and incentivize different kinds of research such as basic and theoretical, translational and applied, and quantitative and qualitative work. One specific application would be to use AEOS in behavioral modeling of incentive gaps, such as getting people to publish data even if they don't publish the paper, or finding ways to accept modular contributions, so that in the case of retraction or deciding not to publish, pieces can still be made available and citable.

We hope that this work has provided a useful step towards the interactive modeling of scientific knowledge as a process of multiscale active inference.

## Chapter VII

---

# The Synthetic Intelligence Guild

## A Social Technology for a Digital Bazaar

Alexander Tucker

### Abstract

---

This paper proposes that the revival of the ancient social-technology of trade guilds may serve as the foundation for the restoration of mutual trust in digital information exchange environments. The paper develops an initial architecture for establishing *the Synthetic Intelligence Guild*, a collegial architecture for practitioners, experts, and apprentices of the information craft to collaboratively develop and maintain standards, insurance, arbitration, and other mechanisms for promoting a reliable, diverse and prosperous information market. These mechanisms are facilitated by the re-establishment of trust and social capital enabled by the guild's reputation system. The primary purpose and benefit of this framework is providing a foundation for developing and maintaining a reliable and efficient information exchange platform (i.e., the bazaar) for trading tools, services, and expertise, as well as structured and unstructured data. As such, the guild is a low-cost and high-reward social technology for cultivating the emerging craft, art and trade of the information industry.

---

# Situational Assessment

*"When a system is far from equilibrium, small islands of coherence have the capacity to shift the entire system."*

- Ilya Prigogine

The accelerating rate of technological change has rendered our world increasingly globalised, complex and interconnected - wherein tiny perturbations can shift the trajectory for the whole civilizational system. This innovation was made possible by the onset of the information age. Never before has such an incomprehensible amount of information, education and entertainment been available and readily accessible to so many: from Bohr, Bach and Basquiat to Gauguin, Gödel and Ginsberg – nearly all the recorded artefacts of human civilization are available at our fingertips. Yet, this abundance of information is a double-edged sword. The unprecedented mass of data artefacts produced in the combinatorial explosion of peer-to-peer communications has led to mounting volatility, uncertainty, complexity and ambiguity. Institutions once designed for information scarcity are now faced with overabundance. Good faith attempts to develop informed opinions are met with a deluge of contradictory information, sock-puppets, subversive conspiracies, PR, weaponized information, malware, advertising, micro-targeted skinner boxes and supernormal stimuli.<sup>1</sup> Perverse incentives to maximise digital programmatic advertising revenue optimises for limbic-hijacking and propagates flawed thinkers and threat actors alike — resulting in a technologically enhanced race to the bottom.

Although our current industrial institutions have been incredibly efficient at serving material desires and increasing standards of living by nearly all relevant metrics - life expectancy, gross national income, curable disease, acute malnutrition, extreme

---

<sup>1</sup> Examples of human supernormal stimuli include opiates, junk-food, pornography - see also, cuckoo brood parasites.

poverty, clean water and sanitation, etc.<sup>2</sup> The fact remains that the limitations of our current information institutions remain the primary bottleneck for continued scientific, technological and social progress. Our contemporary industrial information model is effectively locked into a strategy of top-down centralised planning which cannot process, adapt or integrate the complex dynamics of exponential technologies and accelerating information environments.<sup>3</sup> This is because a finite number of specialised experts relying on a fixed set of bureaucratic procedures cannot cope with the increasing growth of complexity in digital environments. The rate of data production has greatly outpaced even the sharpest projections and most esteemed experts. This is plainly demonstrable: even the simplest decisions, like designing a seating plan for a 60 person fundraising event has more possibilities (60!) than the number of atoms in the universe ( $10^{80}$ ) - consider the complexity emerging from the potential interactions of  $3.95 \times 10^9$  internet users.<sup>4</sup> This banal example comes nowhere close to the real world dynamic and nonlinear interdependencies between human and machine systems: cell towers, satellites, retail transactions, IoT devices, routers, aerial photography, weather stations, smartphones, etc. In short, institutions, analysts and researchers attempting to perceive, discern and categorise the accelerating evolutionary outgrowths of our collective interactions, despite immense skill, motivation and expertise, will find that “it takes all the running you can do, to keep in the same place.”<sup>5</sup>

In addition to our collective inability to make sense of reality, we are currently beset by a pantheon of global catastrophic risks - biological, technological, financial, meteorological, and sociological.<sup>6</sup> Although there is a growing awareness of these complex problems, they remain notoriously difficult to diagnose. This is because the

---

<sup>2</sup> See declining extreme poverty rates, child mortality, life expectancy, literacy, hygiene & sanitation, access to clean water, better nutrition, medicinal advances and income per capita. See Our World in Data and work by Steven Pinker.

<sup>3</sup> See Ashby's Principle of Requisite Variety: “any system that governs another, larger complex system must have a degree of complexity comparable to the system it is governing.”

<sup>4</sup> Naturally, the complexity of the problems faced by policy and decision makers are far more daunting - this is plainly illustrated in the cases of high speed algorithmic trading, machine intelligence, nanotechnology, retail genomics, 3D printed weapons, computational propaganda, etc.

<sup>5</sup> See both Lewis Carroll's Through the Looking Glass & Van Valens's Red Queen Hypothesis.

<sup>6</sup> See Nick Bostrom's Global Catastrophic Risks.

increasing complexity and magnitude of these challenges are effectively impossible to resolve with our conventional means and capabilities.<sup>7</sup>

Unfortunately, the bottleneck is not merely information processing capacity, but also access to information - a wealth of scientific, financial, epidemiological, and personal information is locked away in silos and stovepipes. This is in large part due to the fact that existing industrial information institutions emerged to hoard, monopolise, and exploit scarce information. This was adaptive for a time, however, limiting open, global, and transparent access to accurate and reliable information is maladaptive in an age of mass peer-to-peer communication and exponential technological growth. We must contend with the fact that our information problems cannot be solved by any finite individual, institution or group using any fixed technology or framework in isolation. Any hope of finding solutions to these multifaceted, interconnected, and complex problems will require the coordination, cooperation, and collaboration of the global scientific, business, professional, and engineering communities. However, this is perhaps our biggest hurdle given the fractured state of geopolitics and the zero-sum dynamics of market failures - for many, the situation has degraded to the point where even calls for good faith collaboration are themselves seen as acts of bad faith — a classic “tragedy of the commons” dynamic.<sup>8</sup>

As such, these are not merely technological problems but also legal, economic, and social ones. We require new information institutions for re-establishing trust in information exchange environments. Without solving our coordination problems we stand little chance of collectively developing and exchanging the necessary knowledge for managing global catastrophic risks, augmenting our problem-solving and increasing our predictive capacity in an increasingly uncertain world.

---

<sup>7</sup> See term “wicked problems” - for non-local challenges distributed in both time and space.

<sup>8</sup> An absurd equilibrium wherein the dominant strategy leads the system to converge on a sub-optimal payoff for all players - where any rational actor would prefer to move to a mutually beneficial higher payoff state yet lack the coordination to do so.

# Minimum Requirements

*"In the 21st Century, the most important mind will be the synthesising mind."*

- Murray Gell-Man

It is long overdue that we move from industrial to information age institutions. However, before considering plausible architectures we must broadly sketch an outline of our essential requirements.

As such, our first high-level criteria must be the following:

- We must **prevent Race to the Bottom dynamics**: address, Hobbesian, multipolar, and Thucydidean traps, coordination failures, negative-sum game theoretic dynamics, free rider, principal-agent problems and/or prisoners' dilemmas, i.e., avoid chaos. This entails maintaining an ordered **Information Commons** - wherein the value of cooperating in a group is greater than the cost benefit of self-serving intra-group competition.<sup>9</sup> This requires aligning the local incentives of individual members with the global interests of the shared information commons. A functional example of an information commons can be found in open source software development.<sup>10</sup> In this case, the source code for each project is available for all to see and use.

---

<sup>9</sup> See E.O. Wilson's Multilevel Selection Theory. See also footnote 11 for controversies regarding naive group selection.

<sup>10</sup> See Linus Torvalds & Linux or Larry Sanger & Wikipedia.



Customarily, when revenue is produced, it is in part returned to maintain and contribute to the wider commons on which it depends. In practice, however, despite rising interest and recognition of the value and utility of these commons, it still remains exceedingly difficult for any individual technologist, researcher or citizen to be meaningfully remunerated when working to develop and improve open-source commons. As such, most contribute without any market incentive or institutional protections. An information organisation seeking to fill these gaps by providing incentives for contributing to the commons, offering robust benefit and risk sharing mechanisms, decreasing transaction costs and socially ostracising defectors and cheaters could render a prosperous information commons essential for competitiveness.<sup>11</sup> Examples like this illustrate that new information organisations must serve two overlapping goals: (a) a substantive increase in individual strategic advantage and (b) a culture and community of trust for the aggregate pursuit of preserving and improving a shared information resource available to the benefit of all members.

- We must **prevent Centralised Capture**: maintaining this information commons while protecting against coercion, collusion, conspiracy, capture, autocracy, kleptocracy, oligarchy, censorship and/or subversion by misaligned special interest groups, i.e., avoiding tyranny.

Protecting an information commons without central-planning, i.e. steering clear of both chaos and tyranny is perhaps abstractly useful. However, as discussed, we do not yet have a clear reducible diagnosis of the real world complexity of our problems

---

<sup>11</sup> A naive group selection theory is easily dismissed: as defectors and free-loaders have an obvious advantage over blind cooperators in a group - see Richard Dawkins eloquent explanation in *The Selfish Gene*. As well as work by Stephen Pinker and others. Heated controversies aside, however, it is simultaneously true (without contradiction) that groups of effective cooperators robustly outcompete “groups” of defectors - this is demonstrable under certain payoff conditions. Naturally, the success or failure of such an enterprise is in part contingent upon an efficient mechanism for discovering and excluding defectors. This is easier said than done.

and, as such, remain in serious need of continued education and reinvention. As such, information institutions must maintain an unwavering commitment to systematically think about thinking and continually design new methods for acquiring, processing and using knowledge in the 21st century.<sup>12</sup> We should be wary of any attempt at nailing down fixed and static rule-based models — and leave the door open to revision, error-correction and iterative improvement going forward. In other words, the primary and continued duty of the guild is to learn from and incorporate the highly-specialised knowledge of its members and *not* attempt to establish a centralised Ministry of Truth.<sup>13</sup>

Recognizing the limitations of finite committees and fixed rules - the guild model facilitates a spreading network of trust maintained by members themselves. The success of our information institution is contingent upon a socially enforced system of voluntarily cooperating agents robustly aligning the interests of diverse members with one-another. The maintenance of this social order enables the emergence of an open and trustworthy information commons and marketplace. This market leverages the knowledge, expertise and creativity of a diverse ecosystem of agents to overcome the cognitive limitations of central planning and thereby providing an unprecedented reduction in uncertainty to all participants. More on this architecture below.

Second, we must increase our capacity for Observing, Orienting, Deciding and Acting (OODA):

- **Observing:** Cultivating and maintaining high-levels of shared situational awareness in challenging information environments and for extended periods of time. This includes discerning weak signals within vast noise, identifying opportunities and threats and upregulating and refining unpolished insights. This requires leveraging a wide network of domain specific experts with highly specialised knowledge — because, subject to certain constraints, those closest to the subject matter are those

---

<sup>12</sup> Such an applied epistemology would include: calibrating predictions, forecasting, probabilistic reasoning, avoiding cognitive biases, inside-outside perspectives, setting base rates, mitigating over- and under- confidence, black swans, unknown unknowns, induction and deduction, red-teaming, the scientific method, applied game theory, etc.

<sup>13</sup> See Orwell's 1984, Huxley's Brave New World, Kafka's the Trial or Solzhenitsyn's Gulag Archipelago and many other fiction and non-fiction examples.

best suited to perceive, detect, parse and process subtle signs and patterns that would otherwise have gone undetected. By composing these separate nodes into a unified problem solving system, distributed information networks have the potential to tap into a greater diversity of sources, minds and methods in a scalable and non-linear manner simply unavailable to centralised planners. This vast new collaborative sensory apparatus would increase the scope, scale and sensitivity of monitoring what is happening in societies and around the world and is quintessential for activities that depend upon rapidly exchanging verified information, i.e., early warning systems, disaster-relief, humanitarian organisations, business strategy, and vital and essential scientific discovery, e.g. rapid-response vaccines.

- **Orienting:** Cultivating and maintaining proper orientation is essential for avoiding threats and acquiring the necessary resources for continued existence and the surplus necessary for attaining increasingly complex goals.<sup>14</sup> In this context, orientation is the art and science of giving meaning to information. This is the most critical step in the process and tends to be elaborate, intricate and qualitative. That said, we may be aided by the development of dynamic and interactive quantitative tools for augmenting individual orientation. For example, consider tapping into a cohesive network of domain specific experts and channelling the strongest arguments and refutations and the best moves and countermoves for a given decision and publishing these results in a publicly available analysis tree. These crowd-sourced maps would provide an overview, from a bird's-eye perspective, of the landscape of available recourse and potential trajectories.<sup>15</sup> Charting the range of competing reasonable theories and counterfactuals not only

---

<sup>14</sup> The word 'orientation' is derived from the latin 'orient' meaning 'east' - most likely because the top of early maps of the physical world pointed east towards the rising sun - the sunrise providing a fixed direction around which one could navigate.

<sup>15</sup> Please see, "the cosmic perspective" maintained by Carl Sagan or "the view from nowhere" held by Thomas Nagel.

promotes the creative generation of alternate hypotheses but also invokes a constructively adversarial analysis and contested exploration of the landscape of choices available. Such a cognitively diverse network of constructively adversarial actors is well suited for rapidly and substantively testing the quality of assumptions, information and sources. As such, effective crowd-sourced maps are a relatively simple tool for augmenting an individual's ability to deduce and derive meaning out of vast amounts of information; avoiding threats, seizing opportunities, as well as identifying the range of available recourse and potential ramifications and externalities of possible choices. Tools such as these would be essential in a distributed system for coordinating decision makers and stakeholders - ensuring the cultivation of shared situational awareness and orientation by providing everyone in the system with an overview of the principal priorities, projects and problems. This is the equivalent of a whole network of experts and practitioners thinking aloud, weighing alternatives and deliberating their next move. This landscape serves as an educational as well as navigational stigmergic tool for experts and apprentices to rapidly survey the entire problem space, discover, explore, and evaluate which projects they deem most feasible, viable and meaningful. Thereby allocating the scarce resource of expert attention towards the biggest and most consequential problems and opportunities. Cultivating a system wide focus improves accountability (publically documenting successes and failures), prevents duplication of efforts and errors, identifies hidden assumptions, propagates insights and maps externalities and tradeoffs. This greater oversight improves overall horizon scanning, risk modelling capabilities, and generally increases system wide understanding in an exceedingly complex and unpredictable world.

- **Deciding:** Evaluating which of the available strategies is most favourable depends on careful planning and weighing of alternatives: including foresight (prospectivity) and hindsight (retrospectivity) and modifying behaviour in accordance with patterns that are the most favourable to our desired ends. As discussed above, relevant, accurate

and up-to-date maps are non-intrusive tools that do not impose their will on us - yet we tend to obey them nevertheless - as they naturally present more or less favourable routes to reach our destination. Powerful tools, like reliable maps, coordinate large groups without coercive control. That said, in most cases there will be multiple viable alternative routes that all approximately meet our criteria - even after carefully checking and re-checking our assumptions, evaluating our sources and analysing alternatives. As such, diversifying decisions by hedging, probing and conducting multiple concurrent safe-to-fail experiments will likely lead to the best results. The greater the diversity of these reasonable experiments the higher the probability of success. When all approaches have been ruthlessly evaluated by both substantive critique and rigorous experiment, that which survives may be adopted by the community nearly unanimously and without any need for top-down enforcement. In short, increasing the range of available recourse while simultaneously augmenting individual and collective capacity to evaluate and test these theories based on direct experimental evidence greatly increases the chances of discovering substantive innovation.

- **Acting and/or Disseminating:** Increased capacity to perceive, evaluate and decide results in an augmented ability to make informed decisions and thereby to achieve objectives in a wider context; in this sense, access to information places an upper bound on good decision-making. As such, it is crucial to not only possess the relevant insights but also to mobilise and orchestrate the relevant means and capacity into timely and effective responses to tangible opportunities and threats in the environment. This includes but is not limited to, delivering the right information to the right person at the right time. It is quintessential that we avoid information bottlenecks, silos and stovepipes by increasing peer-to-peer communications. Empowering the creation of consumer facing information products enables the development of increasingly bespoke solutions. In short, a networked peer-to-peer information system composed of specialised experts and producers with a direct line of communication to information consumers vastly

increases the quantity, quality and granularity of options generated. This increased access to a diversity of information products within a constructively adversarial system ensures robust options to survive and a diversified research, experiment and development portfolio optimises the chances of success given uncertainty.<sup>16</sup>

It is painfully clear that the challenges we face are neither simple nor linear and that we must proportionally increase our capacity to evaluate and respond to chaotic complex systems in a scalable and non-linear manner.<sup>17</sup> The methods outlined above are necessary but not sufficient and are already employed by many high-performing institutions. However, in contrast to existing practice, the guild model employs this perception-action coupling in a widely distributed and decentralised fashion in an attempt to mitigate the requisite variety limitations noted above. In other words, at scale a distributed information network cultivating shared situational awareness is a plausible solution to information overload and irreducible complexity problems.

As such, the information commons and constructively adversarial environment could establish and manifest the following:

### **Decentralised Collective Intelligence**

The power of collaborative intelligence is a function of the interactions of members, namely, their capacity to synchronise, build coalitions and form natural hierarchical groupings which derive benefit from more knowledge and expertise than any agent can possess individually.<sup>18</sup> The strength of these

---

<sup>16</sup> To note, the final “acting” step serves as a guide for future “perceiving” steps - thereby completing the loop.

<sup>17</sup> In short, some questions do not lend themselves to shortcuts, divisibility and tractability - seemingly, the only way to determine the result is to run the program and see what happens. Consider Brooks’ principle of indivisibility: it takes 1 woman 9 months to make a child - however, 9 women cannot make a child in 1 month. Likewise, the assumed difficulty of factoring large integers is the basis of many cryptography algorithms. Outside of simple formal systems, however, this irreducibility is seemingly quite common: weather systems, brains, the three-body problem, turbulent flow dynamics, etc. See undecidability in computational complexity theory, e.g. NP-completeness and exponentially increasing running time. See also the principle of computational irreducibility.

<sup>18</sup> Consider the swarm behaviour exhibited by a colony of honeybees tasked with evaluating the quality of potential nest locations: Scouting bees communicate favourable nesting sites,

coalitions is compounded by the variety of perspectives, sources, and methods deployed as well as the capacity to share insight between them. As such, a diverse information sharing network for facilitating international, interagency, and interdisciplinary cooperation, coordination, collaboration is the cornerstone of effective collective intelligence. The deeper and more integrated the chains of experts across academia, industry and technology, the greater the comparative advantage of specialisation, and consequently the greater the problem solving capacity and the macro-efficiency of the system as a whole.

The digital guild model operates, not as another bounded artificial institution, but as a loose collective of fully autonomous individuals and communities already practising the emerging information craft and trade. For example, a digital-first institution is not bound by geographical location, and as such, small, scattered, autonomous groups can consult, coordinate, and act jointly across greater distances and across more issue areas than ever before. These institutions can thereby facilitate broader international and interdisciplinary coalitions and partnerships across diverse perspectives, disciplines, cultures and industries. A diverse multi-agent ecosystem of functionally autonomous agents trading, collecting, analysing, evaluating and responding to complex information challenges in their local environment is able to discern increasingly subtle and complex patterns while avoiding inattentional blindness, siloed thought and limited bandwidth of central planning.

Increased cooperation, coordination, and collaboration between a diverse range of participants greatly enhances information processing capabilities of both individuals and groups. The relation between the collective and the individual is reciprocally enriching - relying on consent over coercion.

---

the better the site the more scouts are attracted to it, once the number of scouts at a location exceeds the sufficient quorum threshold a decision is made and the swarm moves into their new home. Acknowledging the limitations of analogies, this threshold satisfying behaviour parallels the accumulative process of neuronal firing in decision making and can therefore be loosely categorised as collective intelligence.



Every member who joins will be empowered through specialisation and delegation and retains the right to exit if the group does not deliver on its operational promise. The greater the benefit that group membership provides; the greater variety of perspectives, information sources and expertise are brought to bear on the complex problems at hand. Collective enterprise facilitates an interconnected and interdependent web of collaborators transferring knowledge, maintaining an interoperable information commons — aligning self-interest, efficiency and optimization with the good of the group and the information commons.

The members are bound to one another through a networked reputation system incentivizing cooperatively adversarial information verification and thereby minimising the need for centralised enforcement. A well calibrated multi-agent system aggregates, evaluates and disseminates local interactions throughout the system. This process of corrective feedback loops, accounting for a diverse range of competing variables, leads to a self-organising, self-regulating and self-orienting meta-system capable of maintaining order in rapidly changing environments. The system conducts experiments, reinforces successes and discards failures. Rather than relying on a single perfect principle, the system is leveraged on hedged probability: tolerating a certain amount of failure as a natural component of the exploration, learning and experimentation necessary for progress. In short, collective intelligence harvests the diversity of the system - leveraging the evolutionary power of variation, selective retention and replication as its primary generator function.

Decentralised cognition may help mitigate the harms of the information-complexity problem because it is Adaptive, Rapid, Antifragile and Innovative:

- **Adaptive:** Flexible design implies the ability to adjust to changing environments. The guild model deploys multiple independent probes simultaneously enabling the freedom to move seamlessly between strategies as dictated by the infinite variety of circumstances — greatly reducing the likelihood of surprise and avoiding single-points of failure. Maintaining a diversified portfolio allows grand strategy to evolve, change and adapt with

the shifting and complex challenges at hand. This capacity for replanning, context-switching and pivoting when confronted with unforeseen circumstances is critical in a world of unprecedented technological, social and economic shifts. Integrating multiple disciplines, perspectives and models - each preparing for multiple possible futures - increases the chances of success, unlocks unconventional and efficient solutions and enables dynamic and fluid adjustments to unpredictable environments.

- **Rapid:** Agile and responsive design implies the ability to react to changes in the environment in a timely manner. This implies the ability to mobilise the appropriate means, resources, expertise and information to respond decisively, appropriately, and quickly to an ever changing fitness-landscape of opportunities and threats. The guild model is a hybrid between decentralised and centralised architectures - in a sense, it is a decentralised and democratic meta-system composed of a series of autonomous centralised and monarchic sub-systems.<sup>19</sup> These subsystems (i.e. qualified experts and individual institutions) are free to act rapidly and decisively on local knowledge and specialised expertise without conflict, permission, bureaucracy or obstruction. As such, maintaining a dynamic and diverse multilateral network of experts, practitioners, and other stakeholders which can be rapidly activated to form remote teams in a flexible manner greatly reduces uncertainty and the dangers of information overload.
- **Antifragile:** Cultivating a robust, resilient and highly redundant networked ecosystem of builders and thinkers contributing to a portfolio of independent projects competing in an efficient marketplace retains the powerful ability to benefit from adversity.<sup>20</sup> To grossly

---

<sup>19</sup> For an intuitive, albeit non-democratic, argument for why ships (or city states) should have professional captains (or philosopher kings), consider Plato's Ship of the State allegory from the Republic (Book 6, 488a–489d).

<sup>20</sup> See, Nassim Taleb's Antifragile.

oversimplify, this is because experiments, adversarial markets and turbulent environments cause brittle projects to fail quickly while sturdy projects hold up to scrutiny and withstand the test of time. This process of selection allows subsystems to go extinct while the metasystem grows increasingly stable. Ideally, the goal is not to prevent, fact-check or censor malicious or flawed information but rather to facilitate an information architecture which is not unharmed by propaganda, grey-zones and muddled information but rather gains from repeat exposure.<sup>21</sup> Antifragile meta-systems evolve in rapidly changing environments and across a broad variety of contexts due to being modular and redundant, having no single point of absolute failure and are therefore built to withstand major unpredictable perturbations. In short: robustness is the ability to maintain effectiveness across a range of tasks, situations, and challenging conditions; resilience is the ability to recover from or adjust to misfortune, damage, or destabilisation in the environment; while antifragility is the ability to gain from disorder.

- **Innovative:** The ability to generate new methods and hypotheses and update existing approaches in light of new evidence does not easily lend itself to rigid systematisation. However, a distributed adversarial collaborative model is extraordinarily generative in this domain because it synchronises a network of highly skilled tradespeople from diverse backgrounds simultaneously engaging different elements of the problem in diverse ways. This network of autonomous experts allows the meta-system to probe many different probabilities, upregulating success and downregulating failure. These agents are coordinated via feedback loops to effectively allocate resources and attention: withdrawing from unproductive strategies and continuing to search for even greater improvements

---

<sup>21</sup> Consider how an adaptive immune system retains an immunological memory in response to viruses and pathogens and therefore gains an enhanced response to future encounters - such acquired immunity due to previous exposure is the principle behind prophylactic vaccines.

after a goal is achieved. Iterative play allows the meta-system to calibrate an efficient balance between explore and exploit strategies.<sup>22</sup> In short, evolution emerges at the edge of chaos - every failure is a feedback opportunity to recalibrate the boundary and maintain balance. Paying for redundancy allows local failures without breaking the meta-system. In addition to creative generation, a truly innovative organisation implies the ability to incorporate these new approaches and to reinvent outdated processes once they have been confirmed by experiment - balancing between creatively exploring the environment and exploiting existing discoveries for adaptive advantage. The shared information commons aims to facilitate, cultivate and incubate new models and tools, as well as encourage the cross-pollination of ideas. Simultaneously, however, once generated these diverse hypotheses, mechanisms, designs and strategies must be rigorously and ruthlessly tested in a series of parallel and independent experiments. This process of bold variation and selective retention is intrinsic to the transformative nature of the scientific method, biological evolution and efficient markets.

As discussed above, the current information landscape is too complex for centralised organisations of finite size and fixed rules. As such, any information guild meeting these requirements must iteratively augment their capacity to collect, analyse and disseminate successes and address failures at an ever increasing rate — this requires employing alternative decentralised and crowd-sourced strategies in addition to hierarchical models so as to compensate for the accelerating complexities of the information environment.

One effective method for aligning local and global incentives towards adversarial collaborative intelligence is the following:

---

<sup>22</sup> See Bayesian optimization.

## Market Mechanisms

An information bazaar employs novel market mechanisms towards achieving decentralised collective intelligence by aligning incentives of disparate groups, effectively allocating and distributing resources and supporting exploration in parallel domains. Markets are self-corrective because market participants derive benefit from exploiting market inefficiencies and oversights - making it increasingly difficult to identify unpriced failures from publicly available information. In this way, markets have institutional mechanisms for error-correction and self-regulation.

Efficiently calibrated markets, within a framework of regulatory constraints, incentivize a broad variety of market participants to compete for scarce resources - this competition functions like internal checks and balances. As such, differences, misalignments, and disagreements between economic agents is a feature of markets and not a bug. Historically, competitive markets have shown a remarkable (albeit imperfect) capacity for coordinating actors of diverse, often antithetical perspectives towards a single purpose. Actors attempting to pursue local reward functions, in aggregate, merge into a meta-stable system which can achieve goals far beyond the capacity of any individual or single group.<sup>23</sup> In this manner, efficient free markets mimic homeostatic ecosystems in their capacity to self-regulate and efficiently exploit resources without centralised control. In this sense, markets meet the above outlined criterion for an adaptive, innovative and antifragile system - encouraging the formation of cooperative coalitions and natural hierarchies for generating hypotheses and rigorously testing them in a competitive environment. Efficient markets function like an optimization algorithm resulting from the convergence of multiple independent and competing perspectives exploring a wide range of strategies simultaneously; which has (subject to selection pressures) the capacity to search the space of possible strategies for solutions in a robust and efficient fashion. This search increases the probability of rapidly generating alternative novel and plausible solutions (theories) and rigorously evaluates them (through

---

<sup>23</sup> See Leonard Read's I, Pencil.

peer-review and experiment). Increasing and diversifying the amount of concurrent low-risk experiments is an effective way of dealing with volatility, uncertainty, chaos and ambiguity. Each strategic approach reorganises the available assets (in this case relevant data, knowledge, experience, and expertise) to be creatively employed to meet the needs of a variety of situations leading to the chance of absolute failure to fail drastically.

As such, science, evolution and markets are all examples of emergent, scalable, and robust systems composed of a variety of entities pursuing a variety of goals that nevertheless can engage in cooperative problem-solving and attain relative mastery over complex tasks and unfamiliar situations. A well calibrated information marketplace could similarly select for powerful and robust information products through the process of variation (knowledge production), replication (dissemination), and selection (criticism and experiment).

What emerges from within the trusted guild network will be a self-regulating information marketplace mechanism for the exchange and bartering of ideas, tools and methods. Multiple parallel solutions and strategies which are all approximately aligned with one another can compete within a structure of checks and balances; and, through diversity, hedge against game terminating mistakes. By empowering guild members to stay informed and vote with their feet - members pursuing local (micro) self-interest incentives will nevertheless result in institutional evolution at the global (macro) scale.

# Proposal

*“The power of intelligence stems from our vast diversity, not from any single, perfect principle.”*

- Marvin Minsky

What follows is an initial sketch of a reasonably plausible guild-bazaar architecture reflecting and incorporating the preceding concepts and requirements without assuming the incorruptibility of any given component.

This model advocates constructing collegial information guilds as a social technology for increasing trust, collaboration and cooperative endeavours between extant communities of practice across business, operating, legal, technical and social domains. The guild will be supported by a capture resistant, self-regulating and secure information exchange platform, i.e., the digital bazaar, through which the guild’s work can ensure genuine and candid distribution by digitally extending trusted networks beyond the Dunbar number. The guild is a framework of social norms, i.e. an honour system within institutional scaffolding inspired by and modelled on the ancient craft fellowships and trade guilds. Historically associations of artisans and traders have succeeded in establishing shared standards, protocols and risk sharing mechanisms dedicated to cooperation, development and trust between members and preserving common interests. Defending supply chains, ensuring quality controls, administering independent arbitration and dispute resolution. A central pillar of the guild ecosystem was maintaining a viable marketplace for members of the craft to exchange and trade knowledge, wares, goods, and commodities.

This voluntary risk sharing community is responsible for maintaining a transparent, predictable, secure and trustworthy forum for its members. This includes incorporating and maintaining open information commons and a prosperous marketplace — the digital bazaar. The bazaar is primarily a mechanism for guild members to exchange verified information on a basis of trusted guild networks. Continued membership and access to the market is dependent on maintaining a high



degree of trust ensured by the distributed guild reputation system - thereby upholding a high degree of trust in the products and services offered by the bazaar.

Principally, by providing a verified information exchange platform the guild greatly reduces information-, transaction- and settlement-costs and mitigates many market failures. But most importantly the guild provides a basis of trust. The guild will maintain the shared public-goods infrastructure necessary to harmonise reliable, stable, and safe information commons — and increase the value of the shared information resources. The guild-bazaar, enabled by this infrastructure, is a network of autonomous agents making requests of one another in order to take advantage of specialised knowledge, delegate complex problems to domain experts, and benefit from the aggregate intelligence of the meta-system. The guild is composed of, and accountable to, the members and is therefore predicated upon providing value, cohesion, security and utility to the commons. The guild-bazaar is a resilient architecture for facilitating the emergence of an overall meta-stable system providing reliable enforcement and predictable standards.

### **Guild**

First, the guild model is the preferred institutional infrastructure for designing, engineering, deploying and maintaining an efficient source of verifiable information. Members of the guild are chosen by the established members and their status maintained based on capture-resistant networked reputation. Members are required to support and protect the commons, just as the commons is required to support and protect its members. The reputation of the guild is dependent upon its members being trustworthy associates and reciprocally the reputation of the individual member is enhanced by being a member of a reputable and trusted guild. Thereby, the members and the guild form a mutually trusted symbiotic relationship. And, information produced and distributed by the guild can be considered as verified information insofar as the reputation and existence of the guild depends upon living up to its own quality standards. The guild is entrusted with protecting and preserving this information commons by facilitating free and open trade between guild members within the walled-garden of the bazaar. This may include, offering dispute resolution services, providing standards and recommendations, ensuring the conservation of a diverse and prosperous information trade, as well as maintaining the integrity of the trading platform itself. The guild is a deconfliction entity representing it's members and the prosperity of the information trade in the bazaar - as such, the

guild employs the necessary constraints for enabling a functioning market and reserves the threat of ostracism to a member from the bazaar for substantially disrupting the ecosystem. That said, appeal to guild-ostracism should be a last resort - ideally members should be responsible for resolving disputes internally based on pre-arranged contractual agreements, precise terms of association, and the selection of an independent arbitrator. In other words, the guild is the custodian for the bazaar - and the bazaar is a clearinghouse for contracts made independently between members - freely choosing applicable standards and resolution methods for themselves. In addition to maintaining the trading platform, the guild will provide an informal forum for community experts and apprentices alike to meet, discuss, and cross-pollinate ideas, projects and solutions — without the fear of having ideas, hypotheses, proofs, etc. misappropriated. The guild provides a collegial commons for sharing valuable insights, providing mutual-assistance and the cooperative pursuit of projects. Additionally, the guild shelters the culture, legal framework, communication and exchange infrastructure. The guild provides the social, legal and economic basis for facilitating international and interdisciplinary cooperation and collaboration in information environments by fostering a culture of highly skilled, good faith, evidence based, and non-naive communication based on collegial values and a shared sense of purpose.

### **Bazaar**

Second, the information bazaar utilises markets as a tool providing both price and reputation as metrics for increasing trust in information environments. A well calibrated market ecosystem is a capture resistant tool for incentivizing a diverse range of actors to openly exchange information, software and services with one-another in a scalable manner. This multi-sided marketplace takes advantage of network effects and could therefore potentially meet or exceed the current information complexity bottleneck. As argued, a competitive market mimicking biological systems could radically outperform aspects of centralised bureaucratic institutions. An information marketplace leverages decentralised collective intelligence to rapidly explore the vast space of possible solutions and exploit low hanging fruit. This marketplace may

initially be inefficient, however, given its nonlinear nature, a network of moderate scale could likely provide bespoke information products at least an order of magnitude more efficiently than current approaches.

Consider an analogy to Object Oriented Programming wherein autonomous and modular API's can be composed into larger meta-structures capable of increasingly powerful and intelligent behaviour. A programmer can make use of a specialised API while remaining ignorant of the nuances and complexities of its inner workings; as such, the programmer doesn't have to reinvent the wheel or hack together sub-par code but instead can make use of existing and well proven sub-programs. Analogously, the guild-bazaar coordinates and composes experts and institutions with specialised knowledge into an overall meta-system which has greater problem solving capacity than the sum of its parts. In this way, the guild brings localised, specialised and artisanal knowledge to bear on increasingly difficult problems as the guild grows and evolves. A network of autonomous entities with specialised knowledge making requests of entities with different specialisations decreases the amount of complexity that each individual agent is responsible for. Akin to how object oriented programming creates increasingly intelligent systems by coordinating specialised sub-systems, the guild fosters an evolving intelligence by coordinating the capacities of its specialised members. The guild is a network of experts and institutions coordinating voluntarily to solve problems which require vastly more knowledge and intelligence than any one agent could possess – thereby aiming to resolve the red-queen problem outlined above.

The bazaar is a trading-floor open to guild members to freely exchange their expertise, code, research, and data - along with the seller's reputation, reviews, and rating, the price mechanism signals to the other members the value of the information resource. The price aggregates a massive amount of information into a single metric which helps buyers make the appropriate decision. The price of information products aggregates information about comparative advantage and indicates to the market participants how to most efficiently allocate their time. Efficiently allocating the scarce resource of our attention to critical problems and highly demanded

activities that are of highest value to the market. Market participants are modular and composable, enabling complex webs of cooperation: a savvy arbitrageur may purchase datasets wholesale from multiple information providers, aggregate them using a rented proprietary algorithm, run the result through another pattern-recognition software (running on rented cloud-space) and sell the result to a researcher for reputation and profit. As the market grows it expands its capacity to adapt and innovate, becoming increasingly antifragile to change and greatly exceeding the information exchange and information processing capacity than any individual or institution working alone. The collective intelligence model brings nuanced, localised and specialised knowledge to bear on wicked problems thereby increasing collective understanding of our increasingly complex globalised information ecosystem. Extracting insights from information is an art and a science - this craft should be accompanied by a collegial, international and interdisciplinary association of information experts facilitating prosperous trade in the information sector - promoting principles of trust, transparency and integrity among all members of the industry. This implies synchronising an ever increasing range of information sources, expertise, perspectives and specialisations into a coalition environment which synchronises effects over a growing number of domains.

### **Governance**

The guild-bazaar is a hybrid governance architecture taking advantage of both centralization and decentralisation - the model combines an autonomous marketplace (the Bazaar) with a digital open society (the Guild).<sup>24</sup> The guild members provide a coherent overall vision, goal and strategy and the marketplace settles decisions in a fast, unbureaucratic and distributed fashion. This model facilitates both vertical and horizontal information dissemination - in an effort to optimise overall information flow - both quantity and quality. As discussed, complex systems cannot rely solely on hierarchical top-down command and control and must distribute decision making to

---

<sup>24</sup> Governance comes to us from the Latin "Gubernare" and the Greek "Kubernetes," meaning "helmsman" or "steersman," connoting the related art and science of navigation.

as many stakeholders as possible. As such, the bazaar relies on crowdsourcing, aggregating the knowledge and expertise of a diverse range of experts and specialists.<sup>25</sup> Simultaneously, the guild maintains an open information commons aggregating ideas and analysis from the marketplace in a publicly available analysis tree - showing possible moves and countermoves - arguments and refutations for different potential trajectories for the guild-bazaar.

Rather than having a single individual or committee responsible for understanding and acting upon the vast data stream; the guild-bazaar calls upon highly-specialised autonomous agents to evaluate domain specific problems. The cardinal claim being that an adversarial ecosystem of competing interests is an optimal foundation for building a transparent and open information system in which the best ideas, programs, and data-sets are spread rapidly across the network. Consequently, developers, researchers, and experts will be in higher demand and therefore receive higher compensation for their services - generating increasingly bespoke and personalised information products. The value of information products will be judged on a case by case basis by the users evaluating its perceived fidelity and utility. It will be the guild's responsibility to ensure the continued alignment of the open market with an open society - avoiding capture from special interests, advertisers and propagandists.

The guild model brings together a broad range of interdisciplinary experts, specialists and communities of practice as a basis for providing, regulating and maintaining an efficient marketplace, fostering an open and equitable information trade, promoting reliable trade standards, collecting and disseminating accurate and reliable information, as well as providing a forum for its members to negotiate trade agreements and to resolve disputes. The guild is a risk sharing network of members who hold each other accountable and honest through a decentralised and independently verifiable collegial honour-system. Thereby shifting the focus from

---

<sup>25</sup> For examples of effective crowdsourcing see protein folding, neuronal mapping, search for exoplanets, MTurk, Kasparov vs. World, etc.

verifying data to verifying institutions, organisations and individuals. Participation in the network not only augments the strategic sovereignty of its members but fosters a basis of mutual trust, de-escalating tensions and encouraging cross-cultural and interdisciplinary collaboration in the information domain. This will encourage an international, interdisciplinary and interdepartmental information network of scientists, engineers, professionals, researchers and firms. Upon this foundation of trust enables building and maintaining an efficient and mutually beneficial information marketplace — the bazaar.

Since ancient times experts, artisans, crafts-people and traders have formed guilds to share tools of the trade, methods of art, trade-secrets, formulas, processes, designs, instruments and patterns - as well as regulating markets for communal benefit. In modern times, this practice has shown extraordinary efficacy in the open-source communities, business engineering teams, tinkerer, maker and hacker-spaces for sharing knowledge, code, and information related to a given technical area or topic of interest. Reestablishing trade guilds will also present opportunities to educate and certify, implement insurance agreements, and explore increasingly counterintuitive potentials. (see Appendix C). These mechanisms are among the many potential emergent consequences of implementing a guild model and nurturing professionalism and collegiality around the art, science, trade and craft of the information industry in the 21st century.

# Architecture

*"Amateurs talk strategy.  
Professionals talk logistics."*

- General Omar Bradley

The previous section has attempted to illustrate why the guild-bazaar is a favourable architecture to establish trust in information environments through iterative reputation feedback in an adversarial environment.<sup>26</sup> This section attempts to explore how this model could evolve from current forms, methods, tools and resources available at our disposal. The aim here is to set in motion the simplest possible self-replicating system based on simple rules.<sup>27</sup> First, we must develop and maintain a guild culture, an *Esprit de Corps*, based on shared principles, goals and foundational values **(§1)**. Second, we should consider establishing a formal institutional coordination mechanism to synchronise the actions of the community and make efficient use of finite resources **(§2)**. Third, we should consider building and implementing a secure digital marketplace platform which optimises the objectives above by designing a capture resistant information organisation **(§3)**.

Currently, law and (computer) code are the primary tools for creating constraints which enable coordination of diverse interests into a reciprocally beneficial ecosystem. That said, the enabling legal and technical constraints (§2 and §3 below) are derived from the normative guild structure (§1) – as such, if these formalizations begin to hinder or counteract §1 they must be disregarded and replaced. As such,

---

<sup>26</sup> See Adam Smith's Invisible Hand.

<sup>27</sup> See work by Von Neumann, Turing, Conway and Wolfram.



the rules, codes and standards of the guild are a continuous re-negotiation between guild members—concurrently and collectively driving towards coherence.<sup>28</sup>

## The Social Structure

### The Social Structure

The end purpose of collegial society based on trust, reciprocity, respect and the pursuit of truth is self justifying. As such, the guild social kernel (§1) along with an institutional scaffolding (§2) provides a strong foundation for developing, deploying and maintaining the technical infrastructure underlying the bazaar (§3). The guild is tasked with establishing a trusted community capable and willing to bootstrap the initial two-sided market and setting off a self-reinforcing feedback loop of supply and demand.

#### Espirit de Corps

The guild's formal honour system is built upon a foundation of norms and principles (without platitudes) for facilitating trust, community and incentive alignment on a shared basis of values and goals. These may draw upon abstract ideals and will serve as a north star and direction for any future legal or technological guild developments. Norms are certain self-imposed restrictions which information workers voluntarily undertake to participate in the guild structure. Norms are intended as heuristics and are therefore to be referred to in spirit only and not the letter. Consequently, norms that can be written down cannot be true norms, but here are some examples nonetheless:

#### Pursuit of Truth

This implies a personal responsibility for maintaining intellectual humility, honesty, openness, boldness, scepticism, curiosity, tolerance, accountability and endeavouring upon a good faith advancement of knowledge, wisdom and the increased capacity for discernment and intelligence. In essence, downregulating bad faith communication and upregulating individual

---

<sup>28</sup> For example, the rules of language allow its users to voluntarily communicate and coordinate as long as they (roughly) abide by the appropriate rules established by communities of participants. Without the basic rules there would be no possibility of saying anything whatsoever. That said, there are many languages and most phrases and ideas can be freely and adequately expressed in Korean, Italian or Hindi. In a similar fashion, voluntarily complying (opt-in) with the enabling and self-selected constraints (e.g, the laws, constitutions, and code) of the guild enables members to derive the benefits and fruits of the guilds collective transactions — thus incentivizing voluntary compliance with guild standards.

capacity for independent and impartial verification, audit, experiment and review. This implies using the most comprehensive feasible body of evidence, actively minimising bias and maintaining independent review. In short, withstanding claims made on faith or authority alone - *nullius in verba* or the more modern incantation: *verify, don't trust*.<sup>29</sup>

### **Fellowship**

An embodied sense of cohesion, society and collegiality towards the pursuit of knowledge is inalienable for the survival of open societies and lies and the continuance of the human project.<sup>30</sup> As such, practising rituals of respect and reciprocity are essential to fostering an emergent professionalism of the art and craft of the information trade. This implies not only a commitment to integrity, decency, trustworthiness, honesty and upholding contractual commitments - but a genuine fellowship to foster the interests of the information and knowledge industry.

### **Security**

Facilitating resilience, risk sharing, education and mitigation across the entire security umbrella: physical, informational, cyber, cognitive, socio-economic, material, legal and operational security, etc.<sup>31</sup>

### **Liberty**

Ensuring freedom, sovereignty, independence, and autonomy. This implies freedom to pursue one's own conception of the good and promote one's own ends. Concretely, this implies ownership over information products and the freedom to export or derive value from their contributions. This is entrenched in the composability of the guild system

---

<sup>29</sup> This need not imply that every claim be verified individually, rather, that it can be individually verified and independently replicated by anyone (open data and open source) within a network in an adversarial environment - means that we may raise our certainty that a claim is reasonably founded in evidence in proportion to the trust we give the network.

<sup>30</sup> The word "science" comes to us from the latin "scientia" - meaning "knowledge" or "understanding." Science has its roots in natural philosophy, the word "philosophy" comes from the Greek "philo-" meaning "love" and "sophos" meaning "wisdom" - most often translated as "love of wisdom." Finally, "Sapiens" means "discerning, "wise," and "knowledgeable" and along with the genus "Homo-" means "wise-man." These concepts of Knowledge, wisdom, and discernment are embedded firmly within the human project.

<sup>31</sup> In the information domain security often becomes near synonymous with being educated, informed and capable of autonomously verifying claims.

and reliance on subsystems to be autonomous and self-sustaining - free to organise, grow, interact, adapt, respond and replicate as they see fit. Likewise, the guild must maintain economic, political and editorial independence to avoid self-censorship or misaligned incentives of external actors.

#### **Meritocracy & Equality**

Every member of the guild must be roughly equal in opportunity to all others - the only differentiating factor being their expertise, mastery and performance in a given area. This implies inclusivity, involving relevant experts, and employing a range of people and skills. Naturally, all ideas must be evaluated on their merits alone.

#### **Formal Institutional Entity**

This formal legal entity (e.g., nonprofit, cooperative, consortium, etc.) is the outside world's interface to the guild-bazaar - ensuring the guild-bazaar remains legible and the value proposition comprehensible to existing institutional actors. This formal institution will serve as scaffolding for the guild norms (§1). This step requires answering difficult questions about allocation of decision rights, resource allocation, budgeting, administrative roles, legal takeover defences, delineation of duties and responsibilities, operations, onboarding, information dissemination, and other housekeeping. The primary task of this step is to formalise the above principles into traditional rules, standards, and bylaws of operation and governance of the initial guild entity: in order to direct the political, social, and economic direction of the guild towards engineering a prosperous information bazaar (§3).

#### **Communications Infrastructure**

Enabling system wide orientation. A set of open and transparent mechanisms for coordinating and communicating strategy, policy, management decisions to everyone in the organisation. Clearly disseminating priorities, trajectory, goal orientation, learning, pitfalls and critical decisions through the whole organisation. A centralised open information infrastructure breaks down information silos and rapidly and efficiently distributes critical information to the relevant stakeholders. These central information channels will increase the flow of critical information so as to integrate the specialised groups into a harmonious whole that optimally achieves the desired end.

#### **Central Knowledge Base**

An open online information repository and knowledge base written and maintained by the guild community through an open collaboration, annotation and editing system. This could take the form of a wiki or other information repository for papers, guides, standards, and data-sets

deemed critical and in the common interest to preserve. This is intended to provide easy and open access to critical information resources. This intentional institutional memory is a tool for tracking facts, models, choices, arguments relevant to the organisation's success and failure.

### **Milestones, Budget & Long-Term Timeline**

An open and clear roadmap of concrete operational goals ensuring accountability and measuring successes and failures. Long term strategic goals should be broken down into smaller, manageable, measurable and time-constrained milestones. The achievement of milestones should be a systematic elimination of risks not self-congratulatory progress reports with moving goalposts. High level examples include (a) assembling a team, (b) building a platform, (c) initial launch, (d) filing appropriate articles of association, (e) financing, (f) meeting or exceeding growth metrics, (g) operations & logistics, (h) establishing partnerships, etc.

### **Advanced Decision Support Systems**

A supervisory control and data acquisition dashboard for orienting stakeholders and decision makers. It is essential for accountability, error correction and progress to provide a quantitative, measurable, visual and interactive understanding of the meta-system across space and time by inserting reliable, meaningful and relevant facts and data directly into strategic discussions. For example, consider the mission control centre (MCC) at NASA.

### **Crowd Sourced Standards**

Guild membership is voluntary but required for participation in the market - as such, guild members must comply with a minimum standard of mutual non-interference and the preservation of the commons and the maintenance of an efficient market.<sup>32</sup> Given that members are always free to exit, this framework of enabling constraints must genuinely reflect the users of the system by aligning incentives in a positive-sum manner. That said, no single standard is desirable for all communities, use cases and persons - as such, the guild intends to evolve a range of more technical standards bottom-up from within the relevant domain, profession, speciality or industry; drawing on existing best practices of high performing institutions. These standards should be rigorously tested and refined - and the most robust, effective and reliable of

---

<sup>32</sup> It is important to note that monolithic standards raise the essential question: quis custodiet ipsos custodes?

these standards will gain the support and adoption from a diverse range of participants staking their reputation on the framework's capacity to encourage a prosperous marketplace. Given the rate of technological change standards should likely contain sunset clauses and be actively renewed by participants on a regular basis.

### **Guild Arbitration**

As discussed, a fixed set of formal rules cannot cover every possible future state of events. As such, we must proactively define a “trigger event,” “escape valve,” or “boundary condition” wherein the rules break-down and we can return to the norms outlined above (§1) to resolve the dispute. For example, if in any sense a rule restricts the ongoing pursuit of a prosperous information trade in the bazaar (§3) it must be overturned and replaced. Formalizations and renormalization will be an ever ongoing conversation. As such, if a law is misaligned with a higher purpose, disregard it. This is an invaluable safeguard against capture, gaming, and hacking, because the community can override a given attack. Complete recognition of the provisional nature of law and code as useful tools in the pursuit of §1 – and enables replacements when superior alternatives arise.

### **The Bazaar**

The Bazaar is a trusted and credible computational medium matching buyers and sellers. The platform functions like a clearinghouse for decisions made at the edges - reducing settlement risk, providing liquidity, increasing predictability, diminishing transaction costs, providing incentives for participation, diminishing information costs, providing meaningful feedback to both parties: e.g., reputation scores and price metrics to guide decision making etc.

### **Autonomous Market Structure**

A multi-sided information marketplace mechanism enables the emergence of decentralised networked governance discussed above. Enforcement and legal decisions, as we know them now, may be found to be enormously and needlessly cumbersome by our descendants and it is likely that we will eventually migrate (at least in part) to computational contracts and processes; an automated market process is guaranteed to run according to some pre-agreed code and provide the correct output. No one can prevent or tamper with the execution, and no one can censor and block any users' inputs from being processed. Ideally a trusted decentralised mechanism could end up largely replacing less efficient centralised bureaucratic control mechanisms. This is, however, a way off still and will need to be embarked upon carefully and thoughtfully and not as an initial foray.

### **Mechanism design**

A prosperous information commons must be a design constraint on a synergistic information market: orienting participants towards positive sum and away from zero and negative sum dynamics.<sup>33</sup> Thereby aligning short and long term goals, individual and collective interests, as well as local and global incentives. This requires ensuring that price reflects both global and local costs and benefits.

### **Honour and Reputation System**

In addition to market price, the buyer is informed by reviews, recommendations, and ratings on the product and even memberships, associations, certifications of the seller-themselves provided and maintained by the market network - in order to ensure accountability and security. For example, Reputation may be group transferable, inherited, alienable, fungible, disputable, vouchable, etc.

### **Dispute Resolution**

Following the same logic as guild arbitration, however, rather than relying on guild arbitration it is likely that the contracting parties will select an impartial arbiter of their own. There are certain routine activities which can be settled by computational contracts or automated marketplaces - and there are some things that are best settled by your peers in a neutral forum of your choosing. See, the EU subsidiarity principle.

### **Market Participation**

A trusted digital forum for exchanging information products, software, research data, consulting services, tools, etc. Sellers post their products or services at a certain price for prospective buyers to purchase. As the market grows, demand will arise for increasingly bespoke information products and services. The goal of the marketplace is to increase guild members and information workers capacity to build, bundle, unbundle, configure, extend, modify, choose, or otherwise derive full benefit from the tools, information and knowledge shared. Specialisation and crowdsourcing enables increasingly niche information products, more esoteric data, greater personalization and bespoke information solutions - setting in motion a virtuous self-regulating information flow network controlled by a series of feedback loops.

---

<sup>33</sup> Zero-sum because the payoffs win (+1) and loss (-1) sum to zero. In negative-sum games the total of gains and losses is less than zero.

### **Market Participants**

Market participants may include, but are not limited to, data providers, aggregators, intermediaries, standards organisations, researchers, software licensors, consultants, fiduciaries, accelerators, gig-economy workers, algorithmic agents, researchers, scientists, community organisations, social institutions, craft organisations, royalty collection agencies, corporations, labour and consumer unions, cooperatives, universities, mutual funds, insurance pools, partnerships, publishers, professional societies, NGO's, nonprofits, government agencies, corporations, investors, science journals, businesses, law firms, educational institutions, public sector, healthcare providers, industry, financial sector, insurance providers, transport & logistics firms, telecommunications companies, supply chain management companies, etc.

### **Market Products**

Market products may include, but are not limited to, documents, metadata, products, tools, technologies, standards, imagery, labelled and unlabeled datasets, traffic patterns, training sets, documentation, services (collation, integration, analysis, commentary, reviews, software services, surveys, compute, storage, algorithmic analysis, replication services, linking, structuring, formatting, tagging, sharing, collecting, gathering, classifying, formatting, cataloguing, credibility assessments, programming or architectural expertise). In addition, there is the added potential of derivatives contracts based on these commodities, such as forwards, securities, futures and options, etc.



# Membership

*"People, Ideas, Machines – in that order."*

- Colonel John Boyd

The first step in this process is to develop a community of stakeholders willing to solve hard problems on a basis of shared principals. A small, operationally focused, interdisciplinary team empowered to develop external partnerships, leverage existing infrastructure and platforms and actively engage with stakeholder communities iteratively during development. Primarily, this is an open call to technologists, researchers, economists, lawyers, project managers and anyone who deeply understands the strategic and normative value of secure and reliable information exchange. If you are interested in participating please reach out to the facilitator.

The guild is a non-partisan, non-political and non-sectarian organisation supporting a broad range of evidence based research and development initiatives across a wide variety of disciplines. The social contract is open to participation without discrimination accorded any non-meritocratic characteristics of any kind whatsoever.

## Summary

*“Borrow a corpse to resurrect the soul.”*

- The Thirty-Six Stratagems (三十六計)

The global interconnectivity and interdependence of communications, transportation, supply and financial networks, facilitated by accelerating technological progress, has rendered our already complex world increasingly difficult to understand and even harder to manage. When pathogens a few nanometres across can grind global industry to a halt; when splitting atomic nuclei can set off a chain reactions levelling cities; when a fat-finger trade can set off an avalanche of high-speed algorithmic trading causing trillion dollar crashes, - it becomes abundantly clear that the butterfly effect is not merely a poetic illustration of sensitivity to initial conditions - but a persistent and pernicious reality.

In our situational assessment we examined the limitations of our current approaches to linear processing and information siloing. We illustrated the ever mounting volatility, uncertainty, complexity and ambiguity arising from declining trust in information environments - resulting in a race to the bottom. There remains a growing chasm between the god-like powers of our tools and our seemingly inability to establish the most basic facts about baseline reality. The prospect of instability, uncertainty and social collapse is nothing new - however, in today's interconnected and interdependent world seemingly minor oscillations hold the potential to grow increasingly volatile until they threaten the stability of the global system.<sup>34</sup>

We find ourselves in a transition period wherein rapidly changing conditions dictate the creation and adoption of new information institutions. These institutions must

---

<sup>34</sup> In the grand scheme of history societal collapse appears to be ubiquitous: the Indus Valley Civilization, the Western Roman Empire and the Zhou Dynasty. For more examples see Tainter's Collapse of Complex Societies.

aim to re-establishing trust in information environments - preventing both (i) a race to the bottom and (ii) centralised capture. Institutions for mitigating global risk and opportunities necessitates increasingly integrated intelligence, forecasting and decision making capabilities - including, perception-action coupling at a larger scale and greater granularity. This would require to observe, orient, decide and act (OODA) in a scalable manner. The paper proposes that decentralised intelligence emerging from an efficient market mechanism could process the requisite information resulting in an adaptive, rapid, antifragile and innovative information system.

The paper proposes establishing an information trade institution upon a unifying *esprit de corps* - the guild. The guild aims to foster the interests of the information trade by increasing communication and cooperation between information providers and consumers. This aim is made possible by iteratively evolving a negotiated framework of constraints, a self-regulating control mechanism, for steering clear of both centralised capture and decentralised chaos. These basic standards and dispute resolution mechanisms are a prerequisite for the emergence of a decentralised digital information exchange platform - the bazaar. The intuition being that simple rules can give rise to exceedingly complex and self-organising behaviour.<sup>35</sup>

What emerges from the guild framework is an autonomous market structure calibrated to align incentives of market participants to preserve the information commons. The honour system enables emergent networks of trust, reputation and accountability necessary for specialisation and the division of labour - which in turn greatly increases the scope, scale and sensitivity of discernment far beyond the capacity of any single actor, researcher or institution. The honour system increases collective cognitive capacity by providing an increasingly complex web of integrated, interconnected and bespoke information products. Likewise, the profit motive incentivizes compliance by providing a basis for rational, voluntary and mutually-beneficial collaboration. The price signal allows consumers to gauge the value of an asset given incomplete information - incentivizing multipolar monitoring, redundant cross-checking and overall increased adaptive intelligence in a world of threats, opportunities and uncertainties. In these ways, the system leverages circular feedback loops of communication, exchange and competition to maintain a stable information ecosystem. An emergent marketplace provides a more efficient, robust and innovative architecture for attempting to orient and act in an increasingly complex and interconnected world.

---

<sup>35</sup> See Wolfram and predecessors.

Any attempt to solve the information complexity problem with a fixed and final solution is bound to fail. This is because our palaeolithic brains and industrial institutions cannot account for every emergent contingency and possibility produced by the accelerating and evolving interactions of a complex and non-linear information system. The hybrid guild-bazaar architecture stands in stark contrast to the centralised design of industrial age information institutions - leveraging a robust market network for sharing critical information towards cultivating and maintaining decentralised collective intelligence and shared situational awareness. This principle of collective information verification renders the security and prosperity of one member symbiotic with the security and prosperity of all. Thereby providing a voluntary foundation of mutual-trust upon which we may construct an information society for the collective response to serious information threats.

# Conclusion

*“No plan survives first contact with the enemy”*

- Carl Von Clausewitz

The purpose of this paper has *not* been to propose a fixed and final organisation which eliminates all errors from the outset. Rather, *The Synthetic Intelligence Guild* must be founded on a tradition of substantive criticism so as to check and guard its own activities by lucid, dispassionate, critical and fruitful critique — iteratively correcting errors, filling blindspots and fixing missteps. Because, a system which silences criticism renders itself impotent to regeneration and thus sows the seeds of its own destruction. In this spirit the guild must embody the foundational principles of collaborative evolutionary design from the outset.<sup>36</sup> Holding an unwavering commitment to change, adaptation, and transformation as dictated by new information, evidence and discoveries about the natural world. Riding on the coattails of this perpetually unfolding process of variation, replication and selection towards the ever unfolding emergence of complexity.<sup>37</sup> As the guild grows to

---

<sup>36</sup> An example of collaborative evolutionary design includes the scientific enterprise which has undergone a process of continuous change from the earliest roots of mathematics and natural philosophy. Spanning from ancient Egyptian, Babylonian, Greek, Persian, Islamic, Indian and Chinese traditions - to the Copernican revolution, Newtonian Mechanics, Darwinian evolution and Einsteinian relativity. As the body of scientific knowledge grows - so do the tools, methods, and institutions of science undergo a continuous process of evolution. See Karl Popper's Darwinean analogy for the growth of scientific knowledge.

<sup>37</sup> This need not be a teleological or orthogenetic claim but rather a generic statement about self-organisation: in a system with a large number of initial states and a much smaller number of final states - such as a classical fixed point attractor. Simultaneously, there are also fundamental phenomena which can be described by simple programs but nevertheless generate exceedingly complex, seemingly random and rich consequences (e.g.  $\pi$ , fibonacci sequence, Mandelbrot set, etc). See also, Kolmogorv complexity.

incorporate practitioners, researchers, professionals, engineers, investors, specialists and decision makers across diverse domains its activities will spread, and in this process, it will learn new methods for upregulating cooperation, collaboration and cognition. The guild will grow increasingly nimble in its capacity to self-correct as it gains experience, capacity, and responsibility. Increasing progress will naturally evoke new forms of administration, organisation, research and development to fit the ever-changing circumstances of an unpredictable future. In this way it broadens and expands the range of possibilities, apprehending ever larger obligations and tapping into increasingly non-local and seemingly counterintuitive potentials.

The purpose of this paper is to provide a forward-looking vision of a new information organisation which, for marginal cost and engineering effort, holds the potential to greatly amplify our present information processing and verification capabilities - producing new social, economic, research opportunities to practitioners of the information craft which, if history is any indication, will reverberate through society and provide socioeconomic benefits writ large.

## Chapter VIII

---

# Elements Related to Maturity of Function in Markets

## An Initial Exploration of the Applicability of Market Mechanisms for Solving Challenges in the Information Environment

R.J. Cordes

### Abstract

---

Market mechanisms and market-informed approaches have been suggested to address the current dire state of our global information environment. Markets certainly can be effective mechanisms for decentralized coordination, but it is equally true that under certain conditions they begin to degrade in terms of their ability to convert self-interest into collective utility. As markets mature over time, in terms of the efficacy and appropriateness of function in relation to their current interaction volume, new pressures are placed on certain elements of their operations - generating emergent solutions and restructuring. Some aspects of both these emergent solutions and the market mechanisms which generate them may be of value to analogous challenges in the information environment. In this paper, markets are generalized under an operational definition using fuzzy set theory, and a proposed set of 8 elements related to the development of markets over time are explored, with consideration for market-like structures captured by the operational definition. The usefulness of these 8 elements and next steps for investigation into the applicability of market mechanisms for solving challenges in the information environment are discussed.

---

Elements Related to Maturity of Function in Markets was originally produced as an explainer document for stakeholders based on feedback from contributors to the Verified Information Exchange Environments Program.



# Introduction

---

In response to the current low-trust, volatile state of the global information environment, market and market-informed approaches have been suggested as a remedy. This is done on the basis that, historically, markets have had and solved fundamental challenges such as difficulty in mapping and accommodating buyers and sellers (liquidity), maintaining reliability in terms of exchange outcomes (standards), providing recourse in the case expectations of exchange outcomes are not met (recourse), and creating the institutions necessary to constitute coherent, synthetic intelligence, all of which constitute challenges of trust and authenticity which are analogous to underlying problems in the information environment [1]. Markets can be reasonably described as an effective mechanism for the coordination of activity and allocation of resources at scale [2], however, it also equally true that markets alone, without market design interventions will, at some point in their development, begin to degrade in terms of their ability to provide system- rather than actor-level utility [3] in the way that other social systems engineering informed coordination mechanisms are intended to [4].

In order to consider the use of market mechanisms for solving problems of trust and authenticity in the information environment it is necessary to be able to describe markets in terms of their relative level of maturity, or efficacy and appropriateness of function in relation to their current interaction volume, in order to better understand at what stages and on what foundations they create the conditions for emergence of liquidity, standards, and recourse related solutions. In addition, it is important to be able to understand at what stages they begin to decrease in value as a mechanism. Further, if the intent is to apply market mechanisms in a system which does not appear to be a market as it would traditionally be defined, there is an apparent need to provide a definition of markets, and consequently, of exchange, which encompasses non-traditional markets. This, unfortunately, requires a new approach, as it would appear that finding a definition of market which is sufficient to effectively describe the myriad phenomena and structures which would already be considered markets is a challenge in and of itself [5].

In this paper, there is an operational approach toward defining markets through the use of fuzzy set theory and a precursory definition of exchange. The applicability of this approach in describing myriad market-like structures and phenomena, commercial and otherwise, is then considered. Next, the development of markets is investigated through the use of common “elements” found in traditional markets, with consideration for their analogs in other market-like structures covered under a fuzzy definition of markets. Finally, these elements are briefly discussed in terms of their relationship to information markets.

# Toward a General Definition of Markets

---

## Exchange

Those with some familiarity of modern markets might associate the concept of exchange with “that which occurs in an exchange house”, such as a securities or currency exchange. Those with a familiarity of the history of markets might instead primarily associate exchange with a broader set of trading phenomena, from the birth of ancient East Asian and Mediterranean trading routes to the highly complicated financial systems of today. However, exchange can and has been interpreted as a far more fundamental and general phenomenon. The 19th Century social philosopher Georg Simmel, in his 1900 magnum opus, “Philosophy of Money”, within a section titled “Exchange as a Form of Life”, states:

*“It should be recognized that most relationships between people can be interpreted as forms of exchange. Exchange is the purest and most developed kind of interaction, which shapes human life when it seeks to acquire substance and content. It is often overlooked how much what appears at first a one-sided activity is actually based upon reciprocity...”*

[6]

Nietzsche, in his 1887 book “Genealogy of Morals”, goes further than Simmel, stating:

*“Setting prices, determining values, contriving equivalences, exchanging - these preoccupied the earliest thinking of man to so great an extent that in a certain sense they constitute thinking as such: here it was the oldest kind of astuteness developed... man designated himself as the creature that measures values, evaluates and measures, as the “valuing animal as such”. Buying and selling, together with their psychological appurtenances, are older even than*

*the beginnings of any kind of social forms of organization and alliances: it was rather out of the most rudimentary form of personal legal rights that the budding sense of exchange, contract, guilt, right, obligation, settlement, first transferred itself to the coarsest and most elementary social complexes (in their relations with other similar complexes), together with the custom of comparing, measuring, and calculating power against power."*

[7]

This more fundamental view of exchange and related analyses as underpinning human behavior has since been explored further and become more widely accepted, with biologists and anthropologists recognizing exchange between non-kin as a "hallmark" of human sociality as "ancient as the genus Homo" [8] and including abstract interactions such as providing access to mates, ideas, and future reciprocation within the scope of exchange [8–10]. For example, the choice of early humans to cohabitate becomes an exchange of potential opportunities to observe innovations - the choice to share, as Simmel suggested, becomes an exchange of goods and benefits in the short-term for the potential of future reciprocity in the long-term. Even where the choice to share is explicitly accompanied by an expressed disinterest in reciprocation, an abstract exchange could still be argued to have occurred - for example, when giving to a charity, an individual expends resources and receives abstract benefits related to mental well-being and social standing.

This view of exchange as fundamental to human interaction does not mean exchange is exclusive to human-to-human interaction. Simmel argues that even where one makes short term sacrifices for long term gain, these abstract cases are not analogous to exchange between individuals, but instead *the fundamental essence of what constitutes exchange generally* [6], "the willingness to sacrifice one thing for another" [11]. Further, this view does not mean exchange is exclusive to humans, as animals can participate in exchange with humans for tangible short term benefits [12], analyze and respond to disparity in exchange rate and quality [13],

and even participate in the kinds of abstract exchanges humans do, with or without conscious acknowledgement of the exchange itself or of its value. For example, lycaenid caterpillars providing nectar to ants in exchange for protection [14], or kingfishers feeding non-kin nestlings which may improve chances of mating in the future [15]. In both cases, expectation of return is embodied in action.

In this context, exchange becomes quite general while still holding value for use in a variety of fields. For our purposes, exchange can thus be defined as:

**Exchange.** *To expend or offer a tangible or intangible service, object, option, or benefit with expressed or embodied expectation of return.*

---

## Markets

The definition of exchange offered above begs the question of how to define a market. Unlike the concept of an exchange, which can be defined as a mechanism at all levels of analysis, markets can carry connotations of mechanism, physical and quasi-physical space (e.g. a digital environment), set or collection, or concept at any level of analysis - making them far more difficult to define [5]. Within the scope related to any of these connotations, there is also the difficulty of ensuring scale- and approach-agnostic definitions, for example, being able to simultaneously account for a market as defined by market segmentation (e.g. in marketing and sales at local, national, or international) [16], national level economies [17], markets “as the socio-economic phenomenon which takes place in the marketplace”, or the location of that socio-economic phenomenon, whether it be a building, a cluster of buildings, a province, or a website [5,18]. Further, just like exchange, markets have long been used as abstractions to scope categories and collections of exchanges outside of commerce, in everything from sports regulations [3] to biology [15]. Across all of these approaches and connotations, there appears to be at least one commonality: that each relates to collections of expectations related to exchange, and as such, a mathematical structure known as a “fuzzy set” may be of use in providing a definition which is both sufficiently rigorous and general enough for our purposes.

---

## Set Theory

Set theory is the mathematical logic that is concerned with “sets”, or more simply, collections of elements. The term element used here is very general, such that a set could be a collection of variables, numbers, objects, or even other sets, as well as empty sets in which no elements are contained. Elements can belong to more than one set (leading to intersections between sets), but it is always binary whether they have membership in a given set or not. There is no measure for ambiguity of element inclusion in traditional set theory. However “more often than not, the classes of objects encountered in the real physical world do not have precisely defined criteria of membership... yet, the fact remains that such imprecisely defined ‘classes’ play an important role in human thinking, particularly in the domains of pattern recognition, communication of information, and abstraction” [19].

---

## Fuzzy Sets

The real-world ambiguities of class membership paired with the limitations on representation of class membership in set theory motivated the creation of a new approach: “fuzzy sets” [19]. Fuzzy sets are a generalization of sets which introduces a continuum of *grades* of membership. “Such a set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one” [19]. This allows for object classes outside the context of mathematical formalism and software engineering to be more rigorously defined. A simple illustrative example might be the capability to move from the ‘set of all numbers greater than 5’ to the ‘fuzzy set of all numbers *much* greater than 5’.

Like ordinary set theory, the fuzzy set formalism is scale-agnostic in terms of the system it intends to represent given that sets may be contained within other sets without necessarily instituting hierarchical organization, allowing for the representation of systems and subsystems of any scale. While some aspects of traditional set theory are approach-agnostic given their generality, fuzzy sets are especially

approach-agnostic due to the capacity to represent an element's class membership in terms of either their likelihood of belonging or as their degree of belonging. Most importantly, as previously stated, this membership function of a fuzzy set addresses the ability to represent ambiguity in class membership in myriad systems where the lack of binary membership is either useful or unavoidable. These three properties of fuzzy sets have led to its broad application in areas such as data processing, decision support systems, management and logistics, medicine, graph theory, control theory, topology, operations research, and natural language processing [20,21]. Further, these three properties have made it valuable to both the social sciences and economics, where ambiguity in class membership in terms of human expectations would otherwise create limitations on quantitative approaches [22]. For the same reasons that fuzzy set theory is "particularly well suited as a bridge between natural language and formal models" [20], it is also well suited as a bridge between the social construction of what constitutes a market and the necessary formality of market definition.

For example, a collection (a set) of objects and benefits represented as the aggregate of expectations of potential market participants about the likelihood of being able to engage in exchange of those objects and benefits, constitutes a market definition of a given environment superior to that of simple binary segmentation, especially where markets are in early stages of development (e.g. barter markets [23]). With the use of more specific conditions and formulae for measuring class membership and the use of fuzzy set operators, such as unions and intersections (See "Fuzzy Sets" by Lofti Zadeh [19], for more information on these operations and the differences between their uses in set and fuzzy set theory), these fuzzy sets can in turn become higher resolution and more descriptive. For example, the intersection of:

- a set of objects and benefits represented as the aggregate of expectations of potential buyers (a set of individuals who have expectations about being able to make purchases related to the environment) about

the likelihood of being able to *purchase* those objects and benefits (A), where expectations are valued at greater than 0.75 and

- a set of objects and benefits represented as the aggregate of expectations of potential sellers (a set of individuals who have expectations about being able to make sales related to the environment) about the likelihood of being able to *sell* those objects and benefits (B), where expectations are valued at greater than 0.75,

$$A \cap B = C$$

would represent a formal definition of a given market in terms which might be highly comparable with those we would find in linguistic terms (e.g. a fish market, or a book store) while also being consistent with what we might expect in terms of liquidity associated with any of the objects and benefits. This is not to say that a fuzzy set represents a *model* of the market. Instead, fuzzy sets as-market-definition could be functional outputs from institutional, system, and cognitive models of the market which take into account norms, standards, narrative, and other factors which affect the perception and collective construction of what constitutes a market without complicating the expression of its definition [24,25]. Further, nested sets of categories of products and derivatives of products, or aggregate expectations weighted based on fuzzy sets of participants or institutions sets might be operated on in order to define other aspects of markets and their interconnections, and could consist of a variety of conditions and metrics for membership beyond expectation of opportunity for exchange.

---

## Markets

With the use of the preceding definitions of exchange and fuzzy sets, and the acknowledgement of the need for a definition which captures the spirit of the myriad definitions of markets while remaining scale- and system-agnostic, we can define a market as:

**Market.** *An environment with any composition of systems, infrastructure, or norms, about which agents have real or embodied expectations of exchange regarding a fuzzy set of tangible or intangible services, objects, options, or benefits.*

In simple terms, markets can be seen as *exchange environments* that can act as mechanisms; with the designation-as-market being created from the collection of *fuzzy* expectations about that environment being useful for exchange. These expectations are *fuzzy* because they vary from individual to individual, change based on feedback, and, given that these expectations live in the minds of actors and observers, are very difficult to quantify. However, regardless of how *fuzzy* these expectations are, they are still real. For example, it is rare that one goes to a book store without embodying expectations about potential exchange [26]. As further clarification, under this definition, both the book store and the town in which it resides constitute equally valid examples of a market related to books, with one offering a much higher likelihood of the ability to purchase or sell books than the other - and this is the case whether or not individuals have elected to consciously consider, define, or act on these expectations. This definition also captures those markets wherein expectations are projected on the participants by observers, such as the exchange of food for security between honeydew-producing Homoptera and ants, as behavioral “choices” related to production and provision, which are influenced and embodied as though they were consciously considered [15,27].



# Maturity of Market Mechanisms Over Time

---

---

## Heterodox Economics

The definition of market offered above is useful because it can capture various connotations of markets in common parlance or obscure instances such as biological markets, and because it is compatible with recent systems-oriented approaches to markets. These approaches include pluralist economics [28], complexity economics [29], new institutional economics [24], and, more generally, heterodox economics [30], as well as other approaches which do not sit within a defined field, such as those which simply see markets as a general structure, unspecific to commercial activity, with proximal similarity to other adaptive, evolving structures [3]. These approaches come with their fair share of critiques, however, despite being associated with *dissent against orthodoxy*, the recognition that “economic activities are embedded in culture” [28], that markets constitute complex, adaptive, social systems [3,24,29], and that the impact, design, and analysis of market mechanisms is a meaningful challenge beyond the scope of more traditional economic models [31] are all relatively mainstream views now - with some new work in these spaces being produced by individuals associated with the economic theories that heterodox approaches are intended to replace [32]. McMillan, in his book, “Reinventing the Bazaar: A Natural History of Markets”, offers a useful, albeit unintentional, summary of many of these concepts:

*“Textbook economic theory does not dispel the markets-are-magical notion... The supply and demand diagram, expounded in countless Economics 101 lectures, is a bloodless account of exchange. It leaves unexplained much of what needs to be explained. It tells us what prices can do, but is silent on how they are set. Supply and demand bypasses questions of how buyers and sellers get together, what other dealings they have, how buyers evaluate what they are buying, and how agreements are enforced... A market is a*

*social construction. If it is to work smoothly, it must be well built. The term market design refers to the methods of transacting and the devices that serve to allow transacting to proceed smoothly. Market design consists of the mechanisms that organize buying and selling; channels for the flow of information; state-set laws and regulations that define property rights and sustain contracting; and the market's culture, its self-regulating norms, codes, and conventions governing behavior. While the design does not control what happens in the market-as already noted, free decision-making is key - it shapes and supports the process of transacting."*

[3]

As complex adaptive systems, markets have nonlinear and emergent properties, making their behavior difficult to predict. However, nonlinear and emergent are not synonymous with chaotic, complex adaptive systems have systemic tendencies [29], developmental patterns [3,33], and interaction motifs [34] expressed and moderated by component mechanisms, interactions of component entities, environmental constraints, nested and adjacent systems, and threats [34–37].

In the case of markets, or economic systems generally, constraints may be tangible constraints, such as those generally focused on in traditional economics approaches (e.g. supply, demand, and physical and quasi-physical infrastructure constraints), or intangible constraints, such as those focused on in heterodox economics approaches. These intangible constraints can be placed into two groups, formal and informal:

#### **Formal Constraints.**

- Laws
- Rules
- Rights
- Organizational Relationships and Treaties

#### **Informal Constraints.**

- Taboos
- Customs
- Traditions
- Codes of Conduct
- Interpersonal Relationships

[24]

## Elements of Market Development

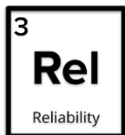
These intangible constraints coevolve with the market and the social systems within which the market is nested [29], allowing for the market to grow in terms of volume of exchange, or more broadly, in terms of interaction volume, and mature in terms of facilitating these growing interaction volumes without collapsing or splintering. This coevolution between interaction volume, intangible constraints, and relevant capabilities is argued here to be related to 8 common elements:



**Emergence.** The emergence of new market dynamics or behaviors based on current state and environmental conditions (e.g. the simultaneous presence of supply and demand).



**Information.** The ability for participants to receive, send, compress, and access information at a rate and level of quality appropriate for the complexity and volume of exchange.



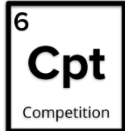
**Reliability.** Expectations regarding the reliability of exchange outcomes and rights are commensurate with reward and risk, and are preferred relative to available alternatives.



**Recourse.** If expectations of an exchange outcome is not met, there is recourse available to reconcile the affected participants' relationship with the market, enforce agreements, and sanction bad-faith and negligent actors.



**Specialization.** The ability for participants to specialize within the market at a depth commensurate with demand and appropriate for the complexity and volume of exchange.



**Competition.** The market has sufficient accessibility and opportunities to encourage new and old participants to innovate and adapt to capture competitive advantage.



**Resiliency.** Risk mitigation and supply buffering commensurate with the probability and potential impact of disruption, and complexity of the systems which facilitate, supply, and enable exchange.



**Sustainability.** The ability to ensure continued operations and mitigate the negative externalities market interaction creates.

---

## Maturity of Mechanism

As a market develops and balances these elements, it becomes more mature, or more effective for handling exchange volume and complexity [3], and acting as mechanisms for collective tasks, such as setting prices, allocating resources, and directing attention [2]. While markets-as-mechanisms for collective tasks is sometimes seen as a libertarian concept, outside the context of political discourse this is a thoroughly apolitical and fundamental aspect of markets; within the context of auction theory, these “elements” might be seen as related to the efficacy of the mechanism for collective price setting and discovery [38], and within the context of collective intelligence or crowdsourcing, as elements related to the appropriateness

of task delineation, task modularity, solution requirements, and feedback requirements given the number of agents in a system [4]. Within the context of work on market history and developmental cycles, the notion of market-as-mechanism sometimes serves as a useful bridge between Keynes and Hayek, as there is a necessary recognition of the usefulness of emergent, self-organizing, adaptive, and problem-solving aspects of markets while also acknowledging the necessity to nudge them through mechanism design [3].

These 8 elements are iteratively developed by actors and organizations of actors within the market [39], with each acting as a necessary foundation for others, continuously in response to new pressures of interaction complexity and volume brought on by further maturity [3]. Below, the development of markets is explored using the 8 elements and the relationships between them.

## Market Development Over Time

---

---

### Emergence



Wherever the primordial soup of demand, available or potential supply, and sufficiently low barriers to interaction exist, haphazard exchange will emerge. This is fundamentally true regardless of formal constraints, whether it is exchange of goods during Mao's agricultural reforms [3] or exchange of tobacco within jails [40]. If haphazard exchange is successful, interaction volume increases to the extent that tangible constraints and availability of information allows.

---

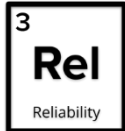
### Information



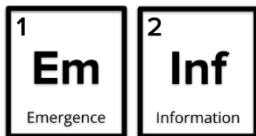
This increase to interaction volume generates new frictions for liquidity and coordination between buyers and sellers which can be expressed as information differentials. Where haphazard exchange is the norm, these information differentials represent significant opportunities for trading advantages and arbitrage - incentivizing work in routing and connection of information and finding more efficient ways to communicate. In biological markets, the limited ability for any particular generation to innovate to improve coordination is offset by evolution of signaling capabilities [15,41]. Markets whose agents are incapable of overcoming

constraints on information flows fail to mature beyond haphazard exchange. Finding solutions to these information-related challenges both facilitates existing demand and invites further growth.

### Reliability



As each interaction reinforces and contributes to expectations of future interactions, growth begins to place pressure on the reliability of both quality and liquidity. At the risk of offering a convenient “just-so” story: in biological markets, solutions to this pressure could be argued to appear in the development of seasonality and periodicity of cooperative behaviors, wherein species shift from haphazard chance encounters as a basis for potential mating contacts to collecting in droves at specific times to improve liquidity and reliability of expectations, thereby reducing the often significant costs of discovery, signaling, and performing. In commerce, we see solutions appear in the form of fairs and periodic markets.



The emergence of periodic markets or submarkets as a mechanism for increasing reliability of liquidity and outcome of exchange is not just a historical phenomena - it still occurs in modern times, both in traditional and rural markets (e.g. in the barter markets of the Songola [23]) and in advanced economies (e.g. conventions, conferences, and product shows). These periodic markets create reliable liquidity while also generating opportunities for setting common norms and building trust between individuals of different groups. These norms are not exclusively cultural - for example, it is in this stage that we see the emergence of common reference-currencies and standards, the transition from haphazard valuation to standards for trade in the form of nested sub-markets, such as fish-for-rice, salt-for-soap, obligations-for-objects, or standardized cloth patterns in barter markets [23,24,42]. While these phenomena are best characterized as solutions to improve reliability of liquidity and exchange outcomes, they also act as emergent foundations for information compression and help to

coordinate demand and the emergence of new actors and types of interaction.

A key case study on the impacts of, reasons for, and solutions to pressures on reliability in markets experiencing rapid increases in interaction volume is the agricultural boom of the 11th and 12th Centuries. Following the collapse of the Roman Empire, there was an extreme decline in economic activity - with some scholars going as far as to say that commercial activity was relatively “nonexistent” [43]. During this period, trade was fairly haphazard and highly constrained for a variety of reasons. Due to the presence of pirates and brigands, even the ability to get goods to a market without them being stolen was relatively uncertain [44]. However, a sudden increase in agricultural productivity effectively reset the economy, with the reliability of availability contributing to the growth of new towns and cities, and most importantly, the emergence of traveling, periodic fairs to which, or with which, merchants could travel to with expectations of reliable liquidity [44–46]. This reliability of expectations of value and timing encouraged the provision of protections by local governments, which in turn increased reliability of expectations of the markets themselves [45].



At this level of maturity, reliability within the market allows for stable connections to other markets, placing new pressures on the development of information flow. In particular, it places pressure on the ability to map between systems of information compression. This is expressed in the numerous “pidgin” languages developed across the world. “Pidgin” is one of many terms for describing languages which serve as intermediaries between two other languages, which emerge where it is “impossible or impracticable for the peoples concern to learn each other's language” [47]. Other terms include creoles, jargons or trade-specific languages, mixed languages, *patois*, and *lingua franca*. These languages often emerge through trading encounters [47,48], and serve as an example of emergent ONFT (ontologies, narrative, formal documents, and tools) [49].

After the establishment of reliability, the market is a fundamentally different object. Whereas in previous stages,

if a market failed to meet new pressures it might have simply stagnated or dissipated - now markets run this risk of “collapse”. The failure for the market to operate as expected now does demonstrable damage to its participating agents by merit of the fact that it is at this stage of maturity because it has become reliable enough for interconnection with other reliable systems - for agents and other markets to embody or plan based on expectations of its continued operation. Further, success will now mean the potential for explosive growth in interaction volume as reliability invites further interconnection with other markets and systems. This generates a feedback loop introducing new pressures on reliability and information flows.

---

## Recourse



In much the same way that increased availability of foraging related communication among eusocial insects invites eavesdropping by predators of ants and those looking for a free meal [50,51], increased communication about value being exchanged creates a cascade of effects which includes both improved coordination between existing participants and a deluge of new agents - some of which will seek to exploit the system.

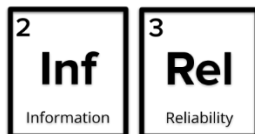
Complex social systems rely on reputation mechanics, whether they are based on gestural accounting in dolphins [52] or on alcohol in iron age France [53]. As groups become larger, approaching limitations of management of interpersonal relationships [54], they begin to rely on a process sometimes referred to as “identity fusion” and homophily, which is developed through ritual, shared customs, bonding, and feedback loops of communal sharing [55,56]. Identity fusion through communal sharing and distribution of surplus develops stable coalitions which help establish reliability in everything from stable trading routes to emergent militias and security retainers [53,55]. Further, the clarity of what constitutes an ingroup and its norms allows for taboos, or the breaking of those norms, to be the foundation for social sanctions and avoidance of ostracization as a means of maintaining good-faith behavior in sharing and trade [44,57,58]. Some argue that the development of “moralizing gods” preceded or even enabled



the development of large trading centers, as they allowed for identity fusion and reliability in expectations of behavior using common signals of trust at a scale far beyond the limitations of the mechanisms of evolutionary psychology discussed above [59–61].



However, when markets grow to the extent that they invite “aliens”, or traders from foreign cultures, the mechanisms discussed are no longer sufficient. For example, following the development of reliability through trading fairs in Europe, interpersonal relationships, memory, and emergent norms soon failed to remain an adequate basis for honest behavior and the trust necessary to continue trade as “transferable reputations for honesty [only serve] as an adequate bond for honest behavior if members of the trading community can be kept informed about each other’s past behavior” [44]. In other words, markets can approach this stage of maturity through emergent standards and measures to improve liquidity and expectations of exchange outcomes, but they risk destabilization and collapse, or “spin-offs, split-ups, split-offs,” [62,63] and other forms of splintering into smaller systems, markets, or trade associations if they do not implement affordances for recourse where standards are not met and affordances for formal sanctions where ostracization is no longer practicable or effective.



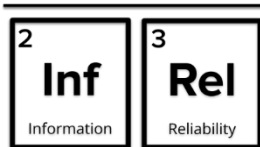
In 11th Century Spain for example, it is at this stage of market maturity we see the introduction of rapidly evolving, voluntary formal constraints on trade, such as third party verification of standards and cooperation with local guilds and magistrates, in response to merchants (especially those alien to the local market) taking advantage of expectations about standards “to the detriment of the customer” [64]. In another example, we can look to the Norden region (Scandinavia and Iceland) in the 9th and 10th Centuries. While this region in this time frame is mostly associated with the famed “Viking Age”, the period of 930 to 1262 is also known as the “Commonwealth Period” in Icelandic history, which is characterized by changes to their governance system and economy. During this period, Norden was home to the development of highly regulated cloth production, checked by dealers, sellers, and regulators in order to

maintain quality standards and ensure the use of cut cloth as reference currency in transactions - resulting in a level of reliability that allowed the standard to be used by, and even replicated in, other markets [42].

In addition to providing alternatives to ostracization in the absence of traditional identity fusion, formal standards and the clear claims and records they result in (e.g. was this norm broken?) also act as an information compression feature to restore reputation management capabilities at larger scales and across distances that would hitherto been highly inefficient or simply not possible [26]. For example, illustrating the effects of this compression across remote multilateral networks, within the 11th Century Jewish Maghribi trading association operating within the Mediterranean, this information compression through formal structure allowed for the remote management of capital, high-throughput information sharing, and the management of insurance policies [65,66]. In numerous other examples, trade and merchant guilds established standards, and did so through a form of local monopoly. While monopolies often carry negative associations, in the earlier stages of market development these monopolies managed by trade associations could be argued to serve a vital role in developing the necessary reliability and organizational coherence for self-regulation processes, such as problem identification and implementation of solutions [67]. Further, analyses on the history of market design have suggested that “guild monopoly rights in [their] home locality may have been instrumental in advancing trade with other localities” [3].

Far and away, the “Champagne fairs” are the most impressive in terms of developing complex methods for improving reliability and recourse. For example, “a merchant could not enter the fair without being in good standing”, and any merchants caught breaking rules of conduct were not allowed to leave until damaged parties were made whole through the fair’s courts [44]. In addition to these stringent controls on input and output, there was a clarity of rules and formal procedure which allowed governing bodies outside the fair to cooperate in extradition, arrest, and the management of other penalties where individuals had left

the fair but reneged on contracts and credit agreements or had failed to meet payment arrangements [46]. Traditionally, law is primarily seen as exclusively within the purview of the state, however, the “spontaneous evolution” of and the voluntarily compliance with the fairs’ codes and regulations, or “soft law” [68] appears to have developed to a far more advanced state than that of the governing bodies of the time [43] - so much so, that while other governing bodies recognized the fairs’ judgment, the fairs did not necessarily recognize those of other governing bodies [46]. They managed to succeed in maintaining voluntary compliance through a “bundling of the services which are valuable to the individual trader with services that are valuable to the community, so that a trader pursuing his individual interest serves the community’s interest as well” [44].



More important than input and output controls on participants, was the input and output standards and controls placed on contracts and their resolution [44,46] - which effectively formalized property rights, an inescapable requirement for the market to mature past an adolescent state [3,45]. Further, the formalization, establishment, and enforcement of protocols related to property rights constituted the development of *lex mercatoria*, or the “law merchant” (i.e. mercantile law), which still stands as the underlying foundation for interjurisdictional (e.g. international) trade [43]. While the development of *lex mercatoria* certainly stands as a monument to human achievement in terms of commercial law and markets, it also serves as an excellent example of the generalization of evolutionary principles to mechanism and protocol regarding shared standards, controls, recourse, and trust signals. The general, flexible structure of underlying protocols allowed for their rapid spread to other markets [44], and allowed for use-case specific adaptation both independently and in parallel to the state in such a way that its fundamental structure is still present in commercial law today [68,69].

---

## Specialization



In addition to reliable and use-case flexible protocols encouraging interjurisdictional trade, they also offer the underlying foundation for specialization of agents and institutions within the market [69]. Reliable institutions “encourage production by fostering saving, investment in human and physical capital, and development and adoption of useful knowledge” [69]. To continue with the use of the example of European trade fairs and other groups of the time, such as the Maghribi and Genoese [70]: this element is instrumental in the “medieval commercial revolution” built on *lex mercatoria* and its cousins. This revolution saw the “invention, diffusion, or earliest perfection of holding companies, of cashless transactions using bills of exchange, of contracts for marine insurance, and of advanced bookkeeping techniques including... double-entry accounting”, allowing for collective de-risking of remote, long-distance, and interjurisdictional trade practices in a way that had never been possible before [45]. These specializations, such as insurance, accounting, and securities generation, went on to constitute their own markets entirely, serving as a nexus between markets, spreading reliability and risk mitigation as a definable benefit to bring to market in and of itself [71].



A market reaching this level of maturity generates a hitherto unseen level of pressure on information flows, due to interaction volumes reaching a point that they can be used to generate new knowledge or benefits via aggregation of *data*. It is at this level of maturity that markets begin to express the level of sophistication we might see in the markets of present day - for example, in the rice-futures market of Tokugawa-era Japan [72,73]. As a result, markets for information *about* markets can now form, creating the strange-loops and self-referential patterns which can imbue a system with “unexpected richness” [74]. This unexpected richness can be expressed in the rapid growth in intermediaries in exchange enabled by sophisticated data and information flows, which can increase the value of assets, reduce risks, and improve liquidity [75], or in the simple provision of information which helps intermediaries and other participants calculate risk, further improving liquidity and enabling advanced derivatives [71]. In markets with the necessary distributed capabilities, we can also see

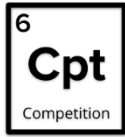
the development of information markets about information markets, such as prediction markets and similar systems [76].

A formalization of the benefit of this reliability, in terms of specialization, can potentially be found in the relationship between reliability and the potential depth and complexity of specialization. For the purposes of this discussion:

- Complexity refers specifically to the number of a specialization's complementary and supplementary connections with, and dependencies on other specializations within a market. For example, a shepherd focused on sheep might be considered to have limited complexity, whereas a blacksmith focusing on armor components might be considered to have relatively high complexity in terms of their specializations' respective positions in the market.
- Depth refers specifically to a specialization's place in a hierarchical order of dependencies which result in the objects or benefits for which the market might be most closely associated. For example, a shepherd focused on sheep might be considered to have a specialization of low depth, given their relative position in a market focused on the production of wool, whereas a miner focusing on iron might be considered to have a relatively high depth in a market focused on the production of armor.

Axiomatically, the more a market's reliability grows, the less risk is associated with investment in specialization depth and complexity. This reduction in risk axiomatically increases the system's tendency toward investment in its components - explaining some of the expressions of this element detailed above. In terms of biological markets, this element is expressed in the evolutionary game theory of niche development, where the long-term stability of the environment facilitates high levels of local specialization [77,78].

---

Competition

At the level of individuals, this investment in depth of specialization provides a much needed function for a system growing not just in terms of interaction volume, but also in number of actors: reducing the impact of crowding and offering new opportunities to newcomers for competitive advantage where incumbent actors have significant advantages.

However, the potential for depth is not infinite and, as suggested above, it is inextricably connected to the level of maturity of the market. Past work on market design suggests that where a market fails to facilitate a depth and complexity proportionate to the number and relative influence of its incumbent agents, and has no the formal constraints to foster competition, it will begin to rapidly degrade in terms of its quality as a mechanism which takes self-interest as an input and outputs collective utility [3]. Given that complex economic systems, in general, have tendency toward expressing power law distributions, and that the incentives for gaining outsized control over interactions in a reliable, high volume market are very high - the ability to embed systemic checks on outsized influence by actors or clusters of actors within a market becomes more and more essential as it matures. These systemic checks, as opposed to simple penalties, are necessary as the incentives may grow to an extent that, despite the potential for sanctions, the game theory begins to favor risky actions such as attempting to gain illicit advantages in recourse affordances (bribes), developing secret coalitions (conspiracy and collusion), and other difficult-to-discover rule-breaking behaviors. Further, incumbent participants developing outsized influence or control over the market does not necessarily require collusion, bad-faith, or even intentionality, as all of the following factors can contribute to barriers to entry which constitute the de facto establishment of monopoly:

- Advanced production techniques by established firms, which are inaccessible to new entities, that offer either cost or quality advantages which are consequently inaccessible
- “Imperfections in the markets” for services, credit, materials, or other benefits and resources, or

“ownership or control by agreement” of these strategic factors by incumbent entities

- “Accumulative preference of buyers” for established incumbents
- Economy of scale or other efficiency advantages offered to large incumbents which are unattainable by newcomers

[79]

---

## Resilience



Whether outsized influence is checked or not, markets at high levels of maturity seem to be tied to another element, which could be referred to as resilience - or, alternatively, a risk motif, which might be referred to as fragility, in the case that resilience fails to be established. Mature systems of all kinds have a tendency to suffer from their own successes in optimization and reliability.

As a system's components...

- become more specialized and optimized for certain tasks, they become less capable of adaptation to changes in the environment
- build on the reliability of other components, they become more dependent on their continued operation and less tolerant to disruption and fault
- contribute to the system's overall complexity through meaningful interconnection with and dependency on other components, new complex threat surfaces emerge [36] and, consequently, the potential for network impacts of faults from individual components and the risk of cascading failure increases

The interconnection between these 3 factors expresses itself, as noted, in a tendency toward either fragility or resiliency. Resiliency in biological systems is often achieved through numerous layers of “safety nets”, with a key example being the adaptation of photosynthesis toward high levels of redundancy rather than efficiency [80]. In human social systems (societies), where the apparent natural tendency is toward developing hierarchical

centralization, there are numerous historical examples of rapidly increasing social and economic complexity followed by rapid and catastrophic collapse [81]. In one example, which illustrates the notion of social systems as-social-markets (as might be found in ethological [52] and social market analyses [53]), and the generalization of the elements listed here to systems which may not traditionally be defined as markets, we can consider the following account related to the c. 2000 B.C. Harappan Civilization of northwestern India:

*"[The Harappan Civilization had] gridded, standardized streets, [seaports, massive granaries], and systems of drainage and refuse disposal... [and] a striking uniformity through time and space in pottery, ornaments, bricks, weapons, implements of bronze and stone, seals, and civic planning... Yet by roughly 1750 B.C. this regional uniformity... had broken down... Street frontages declined, brickwork was less careful, bricks from older buildings were reused in new expedient ones, older buildings were subdivided... expressive art became simpler... groups of unburied corpses were left lying in the streets... [and] Harappan occupation was followed by people who lived among the ruins in flimsy huts"*

[81]

This account is remarkably similar to accounts of the Roman Colosseum in the time after the collapse of the Roman Empire, during which time the historical monument was used by squatters, dismantled in pieces for external construction projects and makeshift "shacks... nestled in and around the building", used as a place to dispose of or bury bodies, and even used as stables for grazing animals [82]. Other examples of the more general pattern of increasing complexity followed by collapse include the Bronze Age Collapse [83] and the collapse of the Hittite, Minoan, Western Chou, and Mycenaean civilizations [81].

In terms of commercial markets specifically, we see this pattern emerge in enough instances to warrant its own category of literature on specific underlying sub-patterns



which lead to collapse, such as information feedback loops and poor cognitive security of participants (manias), interaction volume spikes (e.g. from panics), and complex interconnection [84]. Of course, among the most famous of the examples in terms of collapses in markets as they would be traditionally defined, is the cascading collapse in US capital markets in 2008 - which was the result of complex interconnection between bundles of mortgage derivative products. Often, the only solution in these kinds of situations is to slow things down, to allow good information to catch up to bad (especially in the case of panic) [84] or to pad the system with extra resources in order to manipulate risk analysis and adjust perceptions of the market.

As opposed to collapse at the level of societies, markets have the opportunity for restoration where there are stakeholders that find it worthwhile to invest in restoring functionality. For example, in the case of the East India Company of the 18th Century: following a series of incidents in Bengal the Company was left with an excessive amount of debt and unpaid tax. The interdependence of many of the European banks with and on the company meant that, “when knowledge of this became public, 30 banks collapsed like dominoes across Europe, bringing trade to a standstill” [85,86]. However, being seen as being “too big to fail” they were given a relatively large bailout and allowed to continue operations [86].

---

## Sustainability



It would appear that the global, interconnected markets of today are testing the limits of their resiliency through the disruptions caused by the COVID-19 pandemic and, at the time of writing, the recent breakout of the second Russo-Ukrainian War, as well as through the impact of related sanctions and potential for spill-over conflicts. The emergence of these kinds of risks facilitates a focus on pragmatics in terms of prioritization of concerns - given that ensuring the continued supply of food and fuel during crisis is of paramount importance. However, as these kinds of risks subside, and markets of high levels of maturity are left to operate in relative peace, a new element comes into focus

- progress toward sustainability and the reduction of externalities. Even if a market is able to:

- manage information flows proportionate to its needs,
- maintain reliability in exchange outcomes,
- provide recourse affordances where exchange outcomes are substandard,
- maintain enough reliability to encourage depth of specialization,
- foster competition, and
- maintain resilience in the face of endogenous or exogenous disruptions to operations

it will run into new existential risks if it cannot meaningfully address the negative externalities it generates for the systems around it [3]. For example, collapses in capital markets are often followed by new regulations which threaten the sovereignty of the market [84].

While the need to address sustainability would appear at first glance to be something that only comes with high levels of market maturity, it seems that sustainability appears as an important element at lower levels of maturity where the market is concerned with public goods, or goods which are shared between collections of actors. The resource or environment in which the resource is contained is generally referred to as a *commons*:

*“In a commons, the resource can be small and serve a tiny group (the family refrigerator), it can be community-level (sidewalks, playgrounds, libraries, and so on), or it can extend to international and global levels (deep seas, the atmosphere, the internet, and scientific knowledge). The commons can be well bounded (a community park or library); transboundary (the Danube River, migrating wildlife, the Internet; or without clear boundaries (knowledge, the ozone layer)”*

[87]

Commons, both in terms of common resources and in terms of common rights, existed long before their definition, seemingly as the default “prior to written records” [88].

Among the earliest defined commons in history, is that which is defined in the Brehon, or *breitheamh* (meaning judge), laws of Ireland. Though they were first recorded in writing in the seventh century [89], the roots of the tribal Brehon legal system are suggested to be prehistoric [88], being passed on orally for centuries in traditional “poetic utterance” [90]. Accounts of its use continue into the 17th Century, where it was drawn on to divvy up tillage rights to common land on an annual basis based on the quality of the soil and other factors [91,92].

A full exploration of the notion of a commons is beyond the scope of this work - however, it is worth noting that both purely centralized management and privatization of common property “are both associated with more degradation than resulted from a traditional group-property regime”, and that the following appear to be stable principles which emerge from analysis of successfully managed, sustainable commons:

- “Clearly defined boundaries
- Rules that are well matched to local needs and conditions
- Individuals affected by these rules can participate in their modification
- The right of community members to devise their own rules is respected by external authorities
- A system for self-monitoring members’ behavior has been established
- A graduated system of sanctions is present
- Community members have access to low-cost conflict-resolution mechanisms
- Nested enterprises - the appropriation, provision, monitoring and sanctioning, conflict resolution, and other governance activities are organized in a nested structure with multiple layers of activities.”

[25]

## Discussion

In this paper, elements related to the maturity of function of a market were explored with the use of an operational definition of markets developed using fuzzy set theory. The driving questions of this exploration were:

## Structuring the Information Commons

### VIII – Elements Related to Maturity of Function in Markets

- At what stages do markets, as mechanisms, begin to degrade in terms of producing collective utility through decentralized coordination?
- What is the relationship between interaction volume and challenges to market operations?
- To what extent can market mechanisms be used to solve problems in the information environment?

What was found is that the mechanisms markets used to solve challenges of reliability and liquidity, whether they were in biological, social, or traditional markets, were generally addressing problems in the information environment as a means of addressing those higher order problems. Further, that markets in early stages of development act as a mechanism, not just for decentralized coordination of production and allocation [2], but also for the development of information exchange systems. For example, problems with reliability required semiotics (common symbolic design and reference) and measurement solutions, and problems of dishonest behavior required reputation management systems. Even in the case of reducing the impact of crashes, solutions often come in the form of psychological adjustment as either first order (market slowdowns to give participants more time to consider decisions) or second order (adding resources to a system to adjust risk analysis of participants) effects of policy. This is to say that the influence of supply and demand are limited in their influence by information affordances and capabilities of the participants - as one could say of a tree falling in a forest with no one around to hear, if there is no information available about availability of supply, does it drive prices?

While more work is certainly necessary to further investigate and formalize the detailed elements and the stages of maturity in markets over time, these development of these elements of emergence, information, reliability, recourse, specialization, competition, resilience, and sustainability do appear to be useful in understanding the level of market maturity and the emergence of related challenges, solutions, and adaptations. However, of more importance is the relationship they reveal between information and the maturity of the market. The primary intent of this exploration was to understand the elements related to maturity in market mechanisms and how they relate to solving problems of trust and authenticity in exchange (e.g. liquidity, standards, and recourse), in the interest of seeing how market mechanisms might be applied to solving problems in the information environment. At any level of maturity, markets are revealed to be universally accompanied by information processing, sharing, and management components facilitated by (or in the form of) controls and standards as a means of enabling continued operations and further development. While these mechanisms of controls

and standards settings are of obvious value to the modern information environment, more exploration is needed to consider the practical aspects of applying them and, given their natural emergence in past markets, to consider why appropriate controls have not yet appeared at scale in the information environment [26]. Perhaps most important, and in need of more specific investigation, is the evolving role of trust within these information components.

## Funding and Acknowledgements

R.J. Cordes is funded through the NSF Convergence Accelerator Trust and Authenticity in Communication Systems Program (NSF 21-572), under award ID #49100421C0036 and is supported in research efforts through a Nonresident Fellowship with the Atlantic Council on appointment to the GeoTech Center.

Thank you to interviewees and to those who provided key information resources.

Thank you to Sam Young for key contributions to the discussion of fuzzy set theory.

Thank you to Daniel A. Friedman for providing key citations and helping to refine the discussion of biological markets and eusocial communication.

Thank you to Scott David for the numerous clarifications regarding commercial law prior to writing.

# Estimating Return on Impact of Misinformation Intervention

## An Initial Exploration of the Use of the Business Case for Estimating Return on Investment of Intervention and Incentivizing Information Sharing

Sahil Shah, R.J. Cordes,  
Pat Scannell, Alex Ruiz,  
& Scott David

### Abstract

---

Market mechanisms and market-informed approaches have been suggested to address the current dire state of our global information environment. Markets certainly can be effective mechanisms for decentralized coordination, but it is equally true that under certain conditions they begin to degrade in terms of their ability to convert self-interest into collective utility. As markets mature over time, in terms of the efficacy and appropriateness of function in relation to their current interaction volume, new pressures are placed on certain elements of their operations - generating emergent solutions and restructuring. Some aspects of both these emergent solutions and the market mechanisms which generate them may be of value to analogous challenges in the information environment. In this paper, markets are generalized under an operational definition using fuzzy set theory, and a proposed set of 8 elements related to the development of markets over time are explored, with consideration for market-like structures captured by the operational definition. The usefulness of these 8 elements and next steps for investigation into the applicability of market mechanisms for solving challenges in the information environment are discussed.

---

# Introduction

While Common Vulnerabilities and Exposures (CVE) databasing is now standard practice in cybersecurity, enabling common defence across sectors through a form of crowdsourcing. Such a federated approach has not yet become standard practice for psychosocial matters such as: online harassment, mis and disinformation, and narrative influence events - some of which are capable of mass disruption, such as changing the outcomes of democratic elections and referenda.

Recent work, such as the pattern collection within the recent book “The Narrative Campaign Field Guide” [1], has indicated that CVE-like databasing of such psychosocial matters would be a valid and valuable approach for responding to misinformation, but more work is needed to understand what tools would be necessary to facilitate this, and in what use-cases it might be most valuable. There have been reasonable efforts elsewhere directed toward CVE-styled detection and sharing of vulnerabilities and exploits in the information environment [1–3], but limited efforts appear to have been made on sharing of responses and informing potential developers of tools for facilitating and incentivizing CVE-related crowdsourcing solutions. Incentive alignment for good-faith contributions is a necessary component of any crowdsourcing solution, and in order to design for incentive alignment, there must be a means of approximating the potential return on choices. Key challenges remain, however, in that there is limited available reference work addressing the return on the investment or broader questions of costs and other impacts regarding the nature and extent of effort given to particular responses to various forms of information vulnerabilities.

This white paper aims to explore the potential for actors to collaborate in sharing practises and responding to mis- and disinformation, how incentives can be identified and how they can be communicated, and how to compare practises through estimations of return on investment, given costs and relative impact. First, we offer some background on what work has been done in adjacent domains on valuing information and relevant impacts. We then explore the potential to quantify the costs and impacts of response and how business cases can be created to invest in response capability, estimate impacts, and extrapolate estimates to non-commercial use cases. Finally, we offer recommendations for developers of related tools and for future work.



## Past Work on Measuring Costs of Information Impacts

Among the most scientifically rigorous attempts to quantify the costs of misinformation-related impacts are those found in game theory [4] and biology. In biology, the causes and consequences of misinformation have been modelled using both generalised toy-problems built on game-theoretic models [5] and in highly specific areas, such as those related to pheromone and biotic noise eavesdropping phenomena where, for example, insects attempt to conceal their activities from other species [6,7]. Attempts to make use of cost-analysis related to misinformation in real world scenarios outside of biology appear to build on similar game-theoretic frameworks [8].

In the case of game theory in toy-problems and biology, rigour is enabled through very tightly constrained scope. Outside of these spaces, attempts to quantify impacts of influence strategies, such as those related to marketing and branding [9], by merit of their intent, are limited due to difficulties in observability. This being the case, post-mortems and *a posteriori*, experience-driven approaches are the standard, using integration of human heuristics and “tradecraft” in order to cope with the complexities of implementing and adapting campaigns [1]. These difficulties are not specific to measuring information impacts - measurement of phenomena within any complex system is notoriously difficult, to such an extent that complex systems are often characterised by the difficulty in tethering phenomena to second order effects or to predict impacts even with complete observability of starting states.

In terms of the analysis of and implementation of interventions within complex social systems, the use of frameworks which consider the use of “attractors”, or centres of gravity within narratives and demographics have been suggested as a sufficient foundation for approaching data collection and qualitative interpretation [1,10,11]. In terms of quantitative analysis, as evidenced by the usefulness of the neural net, there is a value in restraining analysis and audit attempts to the discrete inputs and outputs that are relevant to outcome and objectives, as opposed to attempting to analyse the intermediate black box or “hidden layer” itself, which may provide illusory results even with full observability.

# Determining Information Impacts

Here we consider approaches for defining and estimating the impacts and the value of impacts of information exposures and interventions.

## Defining Information Impacts

There are myriad methods to define the impacts of information depending on the specific use-case and level of abstraction applied. For example, evaluating information impact, and even the definitions of “information impact” itself, may differ wildly between those proffered for different purposes, for example for analyses of narrative influence campaigns by foreign state actors attempting to influence elections versus those proffered within the context of pure information science. Even at similar levels of abstraction of what constitutes abstraction, definitions offered within information science may differ from those offered within the cognitive sciences, or within integrative frameworks which overlap with these domains. However, regardless of the model one decides to apply (e.g., Shannon’s model of information exchange, or the more recent active inference cognitive modelling framework) to evaluate abstract information impacts, the discrete impacts of a narrative influence campaign, or the relationship between information and its impacts generally, the following framework for describing categories of information impacts, adapted from prior work in information science, is sufficiently general to be of value:

**Impact on Behaviour.** A marked change in behaviour or probability of the behaviour in the agent.

**Impact on Knowledge.** A change or magnitude of change in the knowledge of a recipient.

**Impact on Search.** A marked change in how the agent searches for information.

**Impact on the Social System.** A marked change in behaviour or structure at the level of social unit or organisation.

[12]

Each of these impacts is interconnected and overlaps with one another - however, each impact is of sufficient importance and interest to justify their separation, as has been noted work elsewhere [12]. This categorization of impacts maps to both the active inference action-perception loop and to the OODA (observe, orient, decide, act) decision making model (see Figure 1), and this direct mapping to a

cognitive model helps to set practical boundaries on analysis for succeeding sections here.

For example, we can initially rule out analysis of the impacts of information on knowledge *itself* (i.e. in memory), as there would be highly limited observability - though with sufficient argument-mining and claims related data pipelines informed by humans-in-the-loop, aggregate modelling of hidden cognitive state for real world applications may become more practicable [11]. Further, we can state that where behaviour and search pattern are well-defined and observable, we can endeavour to measure impact, without the need to measure or observe the intermediate processor which consumes sense data and generates output behaviour. So long as scope, and focus on discrete inputs (exposures to well-defined information packages or interventions) and outputs (behaviour and search pattern) are maintained, coarse-grained analysis at the level of groups may be possible, as evidenced by the formalisation of advertising [13].

Structuring the Information Commons  
IX – Estimating Return on Impact of Misinformation Intervention

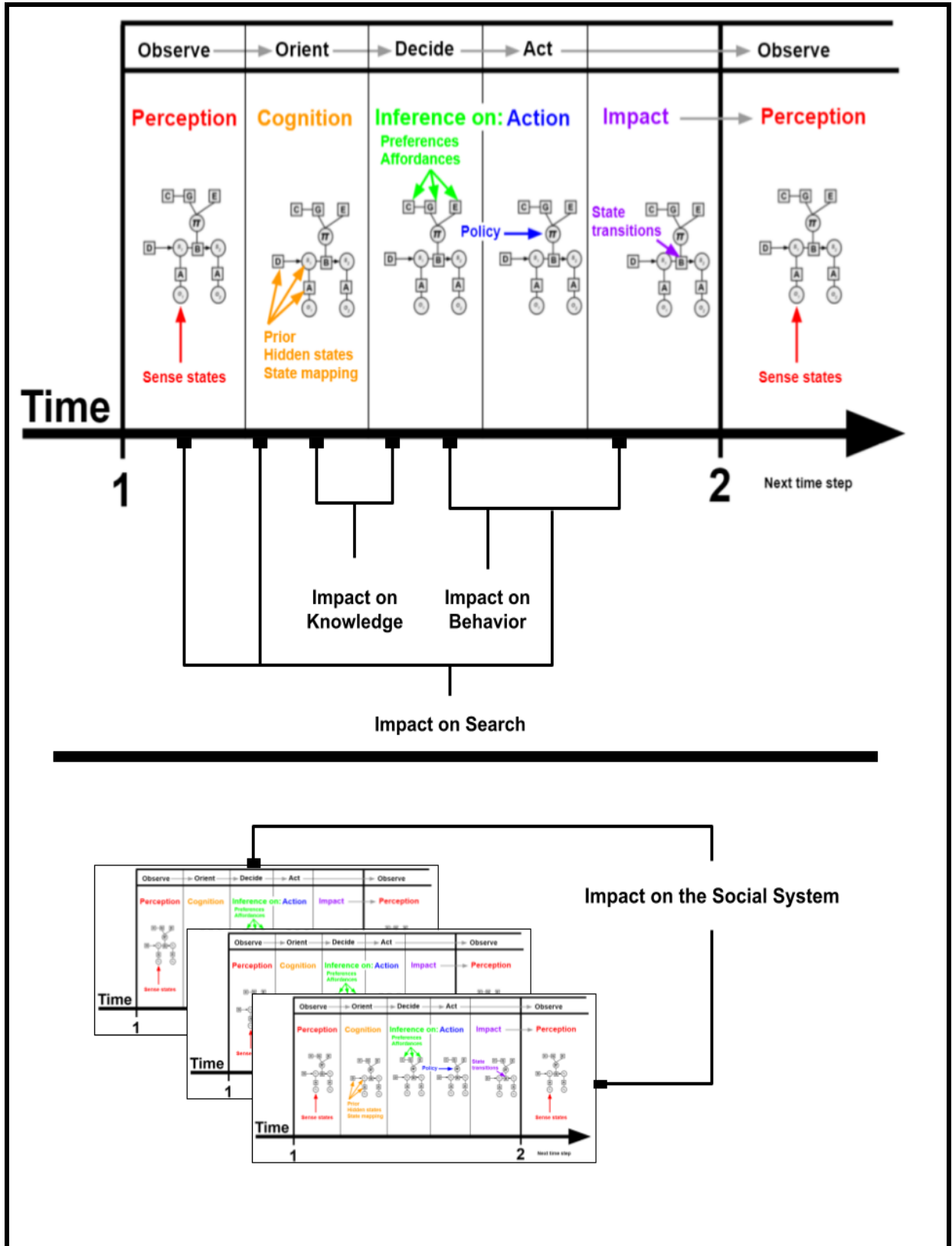


Figure 1. Framing Information Impacts using the Active Inference Action-Perception Loop. Modified from [14].

## Defining a Foundation for Determining Value of Impact

Here we propose a set of terms and discuss relevant limitations and assumptions in order to provide a foundation for determining impact of information exposures and interventions.

The following definitions are used throughout this discussion:

**Collaborative Space.** A collaborative space is here defined as the abstract space wherein 2 or more actors work together to maintain or adjust the status quo of a particular meaning or belief when it is targeted by or is expected to be targeted by some information threat or is subject to some information vulnerability.

**Belief.** A belief here is defined as an expressed or implied claim about the world by an individual or collection of individuals.

**Lens.** A lens here is defined as a cluster of supporting or refuting beliefs, symbols, memes, and other objects relevant to the narrative surrounding some central belief held by an individual or collection of individuals.

The assumptions which underpin and warrant these definitions are as follows:

**The Information Environment has No Clear Boundaries.** Every human being that is interacting with the internet is affecting that environment in ways that may be aligned or misaligned with the intents of any other actor attempting to manipulate that environment. The use of the term collaborative space is a means to constrain the actors, information, and outcomes of interest.

**Narrative is a Black Box.** Narrative is difficult to articulate and analyse [1]. More concrete alternatives in cognitive modelling are not yet ready for use in real world settings, and may not be for some time. However, a lens represents a package which is useful for qualitative aspects of analysis of information impacts and related decision making, and the use of a central belief as a centre of gravity represents an opportunity to scope analyses.

With these assumptions in mind, we assert that a belief held by an individual can be an asset to a company in much the same way that a brand can be, the value of which can be approximated using similar models [15,16]. In some ways, this approach may simply be an alternative to valuation some relevant aspects of the brand itself. Moreover, changes from one belief to another (or changes in certainty about those beliefs) in individuals can have calculable profits or losses in the form of value for

companies, and the return on advertising expenditure can be calculated with consideration for losses to the value of the brand incurred by accidentally poor messaging.

In a **highly simplified** example, the per annum value of a particular belief,  $B$ , of an individual to a company might be valued at the probability of a commercial action or set of commercial actions associated with that belief,  $A_p$ , multiplied by the average per annum dollar value,  $V$ , of that action or set of actions to the company, minus the relevant ad expenditure,  $E$  (see Equation 1); and the P/L of the transition from one belief to another might be simply valued as a function of the difference in the value between those two beliefs (see Equation 2).

$$f(B) = A_p \cdot V - E \quad (1)$$

$$g(B_1 \rightarrow B_2) = f(B_2) - f(B_1) \quad (2)$$

Real-world implementation would require a significant amount of work for a number of reasons, such as:

### Determining Status of Belief

As noted earlier, cognitive status cannot be confirmed - the only available option is to make inferences about the status of belief from relevant data. Large amounts of data may be necessary in order to infer status of belief from interactions with artefacts and claims associated with a lens.

### Social and Cognitive Complexity

As noted earlier, analysis of effect of isolated inputs or phenomenon in complex systems is highly challenging. Cognitive and social systems present a myriad of factors which exacerbates this difficulty. For example, it may be assumed that human behaviours are consistent with beliefs, but this is not always the case. Moreover, beliefs have numerous determinants and in any given time period, individuals will be consuming information affecting beliefs which cannot be captured by the observer. Even within a controlled

environment, measuring the impact of specific information in context with a lens would be extremely challenging.

Individuals or customers will also have different thresholds for actions given the same belief. For example, one person may be willing to purchase a product if it is reported as only over 80% effective, whereas another would be willing to do this at 50%. Direct interviews and self-reports will be vulnerable to post-hoc reasoning and justification, and further, people make decisions through biased information and processes which they may not even be aware of.

### **Accounting for Cost and Potential for Impact**

Attempts to influence narrative come in many forms. Even where heavily patterned or categorised, techniques to influence narrative and belief need to be tailored in order to be effective. Techniques would need to be patterned, and numerous case studies would have to be assembled in order to develop composite functions for cost and estimate of impact given parameterized implementation.

### **Accounting for Risk**

Attempts to influence narrative always carry risk of “blowback” and unintended negative consequences. Further, different techniques and use-cases come with their own specific, parameterised risks. In order to acknowledge this risk in calculation of cost estimates, numerous case studies of failed implementations would need to be assembled and analysed.

### **Accounting for Opportunity Cost**

Given varied parameters within patterns of implementation and pattern- and use case-specific risks, comparison of techniques will be challenging. In order to acknowledge opportunity cost, use-cases, techniques, and risks would all have to be well categorised by pattern, and well documented in order to allow development of heuristics and formal methods for classification of new situations and comparison.

In the face of these challenges, there is the potential to use methods found in auction theory as a means of alleviating, or at least externalising, some of the most complex aspects of measurement, such as determining status of belief or the value of the shift in belief. Auctions can be used as an instrument for purposes other than the

exchange of goods; for example, they can be used as an alternative to other systems in order to increase information transparency [17]. Important to our immediate purposes, is that they can be used as a crowdsourced, information processing mechanism to approximate the value of hidden states [18].

Consider, for example, that if a bidding mechanism incentivises a bidder to express their true valuation of the object of the auction within their bid, then the mechanism is considered to be *incentive compatible* within this context [19], and can therefore be implemented to reveal information about the market. Of course, even with incentive compatibility - the bid value can be considered the valuation of the object by the bidder within a profit-seeking context, as opposed to some objective value of the object itself - as the bidder is expected to seek *profit* from the interaction by bidding underneath what the object might be actually worth to them.

Provided that a small crowd of organisations interested in changing a particular belief can be convened, and that organisations which would provide narrative influence as a service (e.g. small advertising firms) could be convened as well, then auctions might be used to consider the valuation of impact on a belief or the status of belief within demographics through an expression by a willingness to pay, and the costs of particular responses through an expression of willingness to work.

The Vickrey auction is one type of auction which is designed to reveal information about pricing [20,21]. Using a mechanism referred to as a second-price procedure, the auction allows for sealed, single-bid, asynchronous auctions that are isomorphic to synchronous, progressive-bid counterparts. The bidder with the highest price wins the auction, but pays the second-highest bid - allowing a bidder to express their highest and most honest valuation of the object, while simultaneously allowing them to profit from the bid.

Vickrey auctions come with 2 primary drawbacks, vulnerability to impacts of cheating and the potential for reluctance to submit an honest bid. The potential for impacts of cheating are obviously not exclusive to Vickrey and other sealed-bid auctions, however, sealed-bid auctions, as opposed to progressive bid auctions, do not allow bidders the opportunity to exit if the bid behaviour indicates they are being victimised. Furthermore, actual cheating is not necessary to have an outsized effect on the auction quality - the assumption of a positive probability of cheating is enough [21]. In terms of reluctance to submit an honest bid, a Vickrey auction only produces quality results where honest bids are submitted, and behavioural, cultural, and strategic reasons can affect the bid submitted. For example, the rules of bid may be counterintuitive to bidders; given that a bid is generally understood to be the price you would expect to pay in the case of its success, it may lead some bidders to underbid despite the bidding mechanism's incentive alignment. In addition, if the bids values will be revealed after the auction - there can be strategic concerns associated with giving the true value or true pricing.



## Quantifying Cost and Impact of Misinformation and Intervention

If a belief can have value to a company, a belief can also represent an abstract liability. *Misinformation* can be defined **within this context** as information exposures which have impacts on beliefs defined in the collaborative space, such as reinforcement of undesired beliefs or undermining of desired beliefs. Misinformation threats can cause all sorts of economic losses for the firm, such as:

- Reduced sales (e.g., due to rumours about product safety or ethics)
- Fall in price customers are willing to pay (e.g., due to expectations of product longevity or quality)
- Increased cost of production (e.g., employee morale or shortage of applicants)
- Increased cost of capital (e.g., rumours which scare off investors)

A given information threat could theoretically be analysed as to how it might affect a particular shared belief, meaning, or lens, and how that change impacts each determinant of cost and revenue within a composite loss function, retrospectively or in real time. This loss function can then be used to compare and triage information threats, and weigh the cost and risks of various patterned interventions in order to consider response options. As an alternative to complicated cost functions, auction theory may be of use here as well - through the convening of service providers to bid. More work is certainly necessary to consider both the construction of cost functions and the structure of auctions.

## Estimating Impact and ROI in Non-Commercial Information Operations

As suggested above, beliefs unrelated to commercial activity do not have a calculable common reference value, which is necessary for acknowledgement of opportunity cost, accounting for risk, and estimating the impact of particular techniques. However, if techniques, use-cases, risks, responses, and even categories of beliefs themselves are heavily patterned and documented, then ROI for non-commercial related implementations and analyses might be given analogous reference values through extrapolation from their commercial counterparts. In many cases, analogous techniques will not have to be mapped in order for extrapolation. A number of attempts have already been made to pattern narrative influence techniques and risks outside of commercial contexts, for example, the DISARM framework [2] and the Narrative Campaign Field Guide [1], and a large number of these patterns can be used in both commercial or non-commercial contexts.

# Developing the Business Case

Regardless of whether or not the frameworks above are used, the convening of organisations exposed to similar information risks for the purposes of information sharing is a necessity in the development of collective tradecraft. To this end, the development of incentives related to a business case, particularly for smaller businesses who would have more to gain from collaboration and information sharing, may be among the most viable options. This is the setting in which trade associations typically form, and in which guilds have formed in the past, to foster information sharing, cultivate standards of practice and legally-recognized collective “duties of care” for the industry, and other information risk mitigations based on the recognition of the risk reducing value of information sharing. In emerging commercial contexts and nascent markets, where collective attention to shared market and risk metrics is insufficiently developed, achieving levels of analytical cohesion that are prerequisites to collaboration is more difficult. Forming such a business case would require attention to a number of factors:

## **Needs Analysis and Market Making**

Without a proper understanding of how relevant organisations understand their information risks and opportunities, it will not be possible to build a market for information and information services exchange.

## **Tools for Initialising Collaborative Spaces**

Proper understanding of the needs and a common set of information risks and opportunities may allow for organisations to convene, but to ensure they can collaborate, they will need tools to assist in the rapid definition of collaborative spaces in order to build markets (e.g. auctions) around the impact, maintenance, or analysis of particular beliefs. Once defined, collaborative spaces can be tethered to cost and impact estimates to incentivise collaboration between users with similar needs, risks, and objectives who can de-risk together in ways they could not do alone. Actors will then be drawn into the equivalent of an information “neighbourhood watch,” and will have the incentive to collaborate in responding to information threats and sharing best practises when the meaning which is targeted by the information threat is of value to all of the actors involved. For example, beliefs related to airport safety would affect all airports and airlines, as well as a number of government agencies. If modelled as a game, the Nash

equilibrium and strong dominant strategy for all actors would be to collaborate.

### **Narrative Information Management Tools**

Strong patterning and categorisation of information threats, risks, use-cases, and practises associated with impact, maintenance, and analysis of beliefs and lenses, as well as for beliefs and lenses themselves, would be necessary in order to abstract particular situations or beliefs such that collaborative spaces can allow for larger numbers of actors. In order to collaboratively develop collective tradecraft, documentation, and patterns, there is a need for common, usable information management tools, standards, and protocols [22]. Further, proper documentation and collection of case studies may allow for the inclusion of historical data in estimation of impact data, though this will come with its own challenges.

### **Governance Protocols**

Organisations are unlikely to share in proprietary or sensitive information without security assurances. Data trusts and protocols for sharing with selective disclosure might be used in order to allow for information sharing.

### **Clear Value Proposition**

The value of participation would have to be communicated effectively. Access to new markets and the ability to monetise extant data may be viable options in the short term, in the absence of the cost and impact estimates (which could only come from repeated interactions and commitments to information sharing).

### **How to Integrate Non-Commercial Community**

One of the most difficult challenges to overcome, may be the integration of the non-commercial communities into extant commercial convenings and workflows. Work would have to be done to understand specific areas of overlap, and where those working on non-commercial use cases can bring specific value to their commercial counterparts.

# Conclusion

This white paper explored the potential to estimate cost and impact of information intervention and misinformation, opportunities and their respective challenges. We have also discussed the causal mechanisms through which a more generally defined misinformation can cause economic damages, both through linear and non-linear impacts, and offered some paths and potential methodologies toward quantifying costs, incentivizing and convening organisations for collaboration, and extrapolating estimations of impacts in commercial cases, to their non-commercial counterparts. Although a thorough exploration of the potential to estimate the society-level impacts of misinformation was outside the scope of this document, the foundation to do so was considered while evaluating the potential for cost and impact estimation at the level of the firm. However, this work was exploratory, not exhaustive, and more work is certainly needed. To this end, we offer the following recommendations both for future research and for those building tools and methodology relevant to or adjacent to that work:

## **Review of the Literature**

A more exhaustive literature review is necessary. This review should be performed by an interdisciplinary team with members coming from industry, government and policy space, and academia, as, given lack of conformity to ontology across sectors, there is likely a large amount of literature which would be missed by researchers in any particular field.

## **Collection of Historical Case Studies**

While work has already been done to collect relevant case studies, we are not aware of any exhaustive catalogue or collation of case studies relevant to misinformation or information interventions generally

## **Explore Potential Collaborative Spaces**

More work is necessary to discover what industries may contain organisations which are well suited as models for or beneficiaries of common collaborative spaces. For example, where industries might see misinformation costs as especially high. As an alternative, collaborative spaces might be defined, and evaluated for how many organisations, regardless of industry, would find it to be relevant to their interests.

### **New Tools for Literature Analysis**

As noted, there is a lack of ontological conformity across the relevant literature, new tools are required to perform searches which move beyond keywords to semantics and relevant subject matter. For example, advertising spending and political campaign spending could be very closely related depending on perspective and community focus. Unfortunately, there is also lack of ontological conformity even within the relevant fields, which means computational ontology may not be an option - instead we recommend that research questions be developed for advancing search methodology which uses crowdsourcing or humans-in-the-loop.

### **Tools for Integrating Analysis Pipelines**

There are a wide variety of tools, analysis techniques, organisations, “data lakes,” and systems relevant to social listening - but there are few means to integrate respective information pipelines with selective disclosure. There is a need for methods to map and translate data between communities while allowing for the ability to monetise, bundle, and restructure. Most difficult among the requirements, is the need to avoid both central repositories and digital twins, as they come with significant challenges regarding governance, storage, and computational expense.

### **Tools for Inter-Community**

#### **Narrative Information Management**

In addition to the need for new tools for integrating analysis pipelines, there is also a need for tools which allow for effective documentation and evidence collection between communities with effective recourse for handling disagreements. These tools would need to be built to integrate with other systems, and provide affordances for assisting in directing member attention to opportunities and needs within various communities.

### **Experiment Design and Wargaming**

In order to buttress any attempts at estimation, meta-analyses of past empirical research related to information exposure should be conducted or collated and mapped to patterns of practises, risks, and use cases. Further, work on wargames related to narrative influence appears to be limited - given the potential for serious games and wargames to be used for

estimation of the real world phenomena and their mechanisms through extrapolation, it is advised that more work be done to explore potential in this space in order to take advantage of the high level of environmental control they can provide.

### **Research on Applied Cognitive Modelling**

There has been a great deal of research on cognitive modelling for decision making in myriad contexts, however, cognitive modelling associated with decision making using narrative and rhetorical compressions in relation to financial costs, was not found during initial exploration of the non-proprietary literature. It is recommended that we find ways to bridge the gaps between political science, economics, advertising, and cognitive modelling to produce non-proprietary research in this area.

## **Funding and Acknowledgements**

R.J. Cordes is funded through the NSF Convergence Accelerator Trust and Authenticity in Communication Systems Program (NSF 21-572), under award ID #49100421C0036 and is supported in research efforts through a Nonresident Fellowship with the Atlantic Council on appointment to the GeoTech Center.

# Tracking Public Sensemaking through Rhetorical Annotation of Image Memes

Mridula Mascarenhas, R.J. Cordes, V. Bleu Knight,  
Sarah Murphy, & Daniel Friedman

## Abstract

---

Political polarization and declining trust in institutions are driving societal destabilization and radicalization. Recently there has been increased interest in online misinformation intervention and deterrence, for example through the use of machine learning on language use. We argue that addressing crises in the information environment will require a sharper situational awareness and a deeper understanding of how beliefs emerge and crystallize, as well as greater connectivity in the work of teams and organizations in order to reduce the effects of bias and partisanship in collection and analysis. Image memes play an increasingly important role in public sensemaking and discourse and in the emergence of public beliefs. Despite their significance, image memes have proven to be a very difficult category of artifact to collect, classify, and analyze in aggregate. In this white paper, the function and form of image memes are discussed, the challenges of performing image meme collection and analysis within the context of emergent, interdisciplinary teams are detailed, and requirements and recommendations for alleviating these challenges are offered.

---



# A Case for Rhetorical Annotation of Image Memes

A growing loss of faith in normative institutions and official accounts of historical or current events is violently destabilizing social bonds. One of the drivers of this destabilization is a growing divide between those who still grant credibility to institutional narratives and those who have gravitated via social media toward counter-institutional narratives. The rise of counter-discursive online communities has produced shared identities constructed through shared narratives. Recent years have seen the emergence, among social media users around the world, of the “truth-teller” or “digital warrior” identities constituted by the consumption and sharing of counter-institutional narratives [1].

Specifically, counter-narratives are being seeded and maintained through image memes. There are now numerous operational and formal definitions of image memes, sometimes referred to as “internet memes” [2,3]. The newer definitions serve to disambiguate the image-embellished-with-text, ubiquitous on social media platforms, from the more general “meme” originally proposed by Dawkins [4] and further developed by Blackmore [5], Dennett [6], Heylighen [7], and others, referring to any “cultural component passed from one individual to another by non-genetic means, or imitation” [3].

In this article, we focus on the “image meme” format (as seen in Figures 1, 3, 4, 6, and 7), operationally defined as a shareable, digital image that contains either non-textual visual symbols or text or a combination of both. While other media formats (e.g. GIF, audio, video) are also ubiquitous, we focus on the image meme in particular to articulate scalable computational systems for rhetorical annotation and analysis which could enrich current analysis methods such as sentiment, semantic, narrative approaches.

While image memes used to be regarded as ephemeral detritus of the Internet intended primarily to induce humor, the powerful role they can play in the formation of public belief and sentiment is becoming increasingly evident [8,9]. Amidst the voices calling for urgent study of the memetic construction of public belief [10], we emphasize the need to examine image memes for their communicative (rhetorical) function as quasi-arguments in public discourse [11].

We recommend applying the structural framework of the Toulmin model of argument analysis to trace the public argumentation performed by image meme circulation [11]. We advocate the application of an argument framework because image meme content is being widely used to advance claims that appear reasonable, despite minimal evidence presented within the meme. Such memetic content has

demonstrated strong potential to shift public belief and spur public action [12]. The Toulminian model identifies three fundamental components of an argument - claim, evidence, and warrant [13]. The claim refers to the proposition that the audience is being asked to accept. The evidence refers to information that supports the claim. The warrant, often not articulated, refers to each assumption that connects evidence to claim.



Figure 1. Example Image Meme.

For example, rather than dismissing the meme above as intended purely for humor, an audience enculturated to reject institutional narratives recognizes that the meme advances the claim that space agencies such as NASA are engaged in long-term deceit. The meme does this by presenting an argument that can be outlined using the Toulmin model as below.

- **Evidence**
  - **IF** a security camera offers only a grainy image
- **Claim**
  - **THEN** space agencies are lying about images from space telescopes.
- **Warrants** (unstated but implied claims which are already established as true for the audience)
  - **BECAUSE**
    - We should be able to see someone on a security camera much more clearly than celestial objects in space
    - Space agencies have a history of presenting fiction as fact (e.g. faked moon landing, doctored images)

The advantage to annotating memes using an argument framework is that we begin to identify how certain ideas which are not contained within the image meme itself (i.e., NASA is a fake organization) and sentiments (i.e., suspicion toward NASA) propagate across publics. Argument analysis reveals the power that memes have to shape public belief by simulating appeals to logic, even spuriously, by functioning as quasi-arguments. Memes function as arguments when they invite an audience to accept a claim, based on evidentiary information that is sometimes contained in the claim and sometimes implicit. Nevertheless, for an audience to accept the claim, the meme also relies on implicit warrants. Interventional approaches to addressing memetic circulation of misinformation, therefore, need to target, expose, and challenge spurious evidence and hidden warrants that specific memes rely on, in order to induce skepticism toward the memetic form as a mechanism for sound reasoning. Fact-checking articles designed to debunk memes have limited success in dismantling the power of memetic argument, in part because they assume a different rhetorical form. Counter-memes that identify and challenge the argument components of original memes could present a more targeted strategy. However, we advocate strong caution in the use of this approach. Counter-memes should not be used to advance novel competing claims but instead highly purposefully to dismantle already circulating spurious memetic arguments, since the objective of intervention is to challenge reliance on memes for public sense-making.

A rhetorical argument-based approach to analyzing image memes can advance our understanding of their persuasive influence beyond the current practices of syntactic tagging of memes, for example by text recognition [14] or classification of memes into categories based upon visual similarity. Previously, we have argued that a rhetorical approach fills in the gaps endemic to tagging practices by enriching analysis of image memes with rich semantic information embedded in the parsimonious combination of the meme components [15]. While in recent years, small-scale rhetorical analyses of image memes have been published [16], these methods have not been implemented widely.

Much attention has been paid to specific large-scale shifts in public beliefs, such as vaccine rejection [12], COVID denial, and rejection of various global election results. However, we argue that what deserves more attention in academic, political, and security analyses is a focus on the underlying common thread running through these far-ranging belief and attitude shifts, namely the mechanism of change: the creation and maintenance of shared beliefs, and the provocation of shared emotions through memetic persuasion [15,17–21].

Despite legitimate concerns about the degradation of public information due to the infusion of spurious counter-institutional content (e.g., “fake news” and misinformation), we argue that viewing the information crisis as a competition between truth and falsity obscures the nature of the digital information crisis we are facing and, worse still, hampers efforts to restore trust and rework social

consensus. Framing the information crisis as a battle between true and false information has not proven effective in regaining the trust of those disaffected by mainstream channels of information. A simplistic true/false dichotomy ignores the complexity inherent in counter-institutional narratives and furthermore prevents us from studying the rhetorical conditions that enable the subversion of mainstream narratives by competing ones. Deploying this dichotomy through strategies such as fact-checking pop-ups that overlay memes on Facebook can actually undermine the ability of good-faith actors to either correct or contribute to competing narratives. Those who seek to address our information crisis will need to do more than target and neutralize alleged sources of misinformation - they will need both a deeper understanding of how beliefs emerge and crystallize, and a sharper situational awareness.

The scale and rapidity with which emerging political and social events are being co-opted into counter-narratives is possible because of the extreme parsimony and virality of memetic argument. Since the split between institutional and counter-institutional narratives has solidified as a social schema, emerging events create a vacuum into which counter-institutional content can be introduced. This content, in image memetic or other forms, has the capacity to nudge social actors into rejecting institutional narratives about events and can reinforce the rejection of political, corporate, and social institutions. Therefore, information systems for large scale analysis of memes as persuasive artifacts are urgently needed. Such information systems have the potential to provide early indications and explanations of shifts in public belief and attitude, which can then be measured with more sensitive and reliable tools. While there are tools which offer operational situational awareness related to sentiment in text-based artifacts [18], image memes have proven to be a very difficult artifact to collect, classify, and analyze in aggregate and there are no standardized practices or appropriate tools for this process. Even outside the context of analysis, no viable and accessible methods exist for systematic search or collection of image memes.

Accordingly, in this article, we build on previous work that proposed a computational framework, combining rhetorical analysis with an ecosystem approach (see Figure 2), to trace the ebb and flow of narratives through memetic circulation across digital publics [15]. We first provide background on the rhetorical form and function of the image meme. We then offer a set of vignettes to communicate the challenges practitioners face in collection and analysis of image memes, and to explore what tooling and related capabilities would alleviate these challenges. Finally, we provide recommendations for developers who are working on related technologies and those who may be interested in providing the necessary infrastructure for an ecosystem-approach to enable situational awareness in the (mis)information environment.

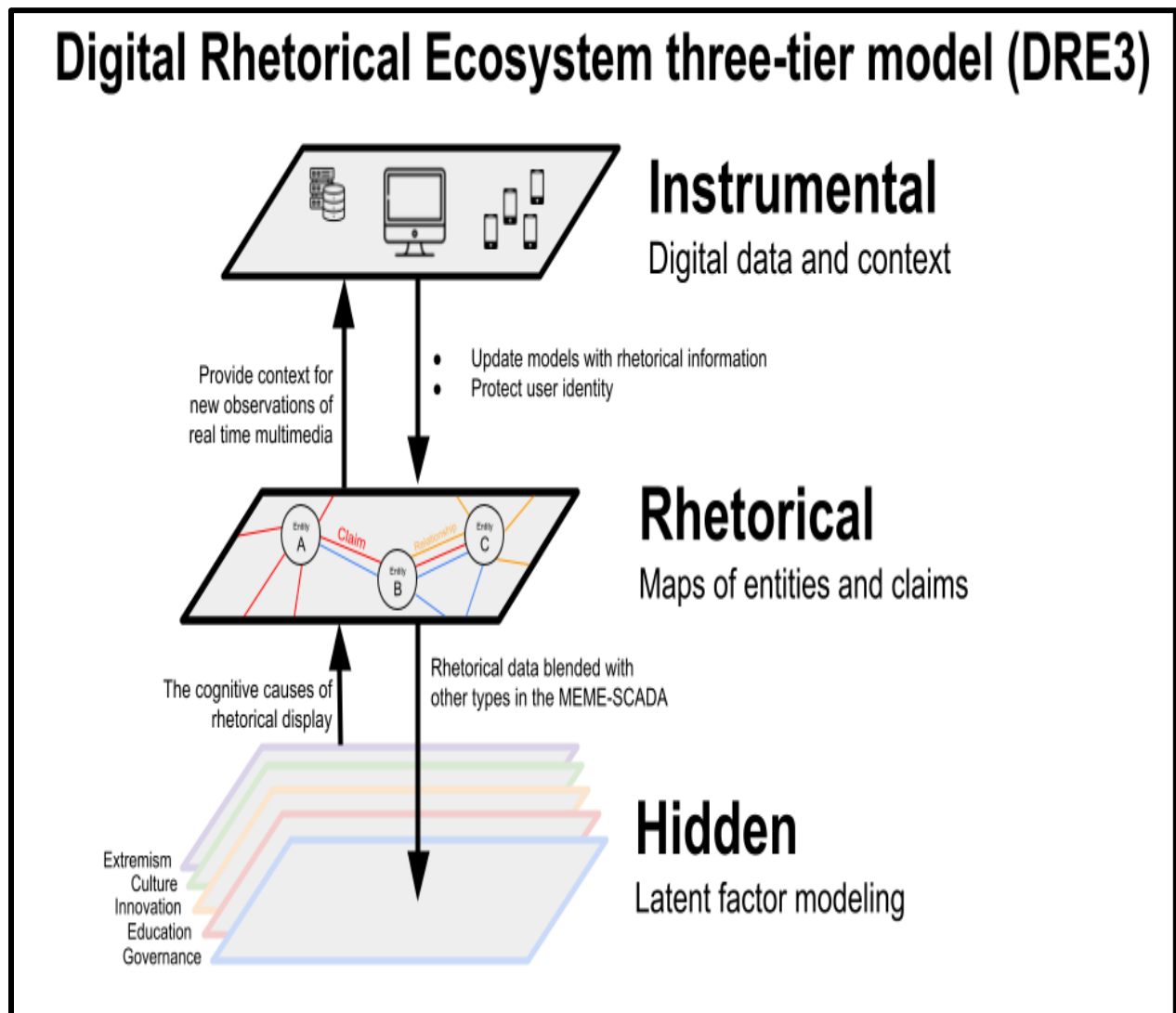


Figure 2. Digital Rhetorical Ecosystem three-tier model (DRE3), from [15].

# Elaboration of the Rhetorical Form and Function of Image Memes

We can consider image memes in terms of two complementary features - form and function. The form of the image meme is established by the rectangular box frame which circumscribes one or more rhetorical elements, demarcating the meme as a discrete communication unit on platforms like Facebook, Instagram, and Twitter. While image memes perform a variety of rhetorical functions [22,23], we restrict our attention to image memes that play a particular rhetorical role by participating in public argumentation by advancing claims [24].

Although image memes can be circulated to drive any narrative online, they have marked success in the disruption of official narratives across the political spectrum [9,12]. Their truncated or compressed form is well-suited to inject targeted challenges to mainstream claims. The parsimonious form of the image meme provides great capacity for semantic encoding to advance persuasive claims while diminishing burdens of proof and elaboration that other rhetorical artifacts, like news articles, would require (or be expected to provide). Various image meme formats exist, such as text-only, image-only (no textual elements), screenshot, and image-text juxtaposition. These varied formats, and combinations among them, can create polysemic affordances [25]; that is, they create the possibility of extracting multiple and multi-layered interpretations within a range of meanings. The strategic ambiguity inherent in memetic artifacts allows for rich semantic encoding. At the same time, the structural features of the memetic form (i.e., the containment of its content in a box, and the text/image syntax) strategically constrain meaning-making by setting up the key elements of an argument and cutting off counter-arguments. Below, in Figure 3, we illustrate the argument development contained in one sample image-text meme.

Figure 3 constructs an argument with the simple juxtaposition of two lines of text above and below a stock photo. The choice of the photo combined with the double textual framing relies on the contextual knowledge of discursive communities to decode the argument. While the explicit memetic content is sparse, its signifying layers are rich, thus allowing the meme to advance a clear and persuasive claim.

The primary claim distilled from this image-text meme is that the official masking policy to combat the virus is not to be trusted. The rhetorical power of the meme draws from its strategy of juxtaposing two official narratives that appear to be mutually exclusive - that is, if the virus is virulent enough to escape the strict safety protocols of a world-class laboratory [evidence], then ordinary masks should be ineffective [claim]. The implied warrant in this case is that both statements cannot



**TELL ME MORE ABOUT HOW A VIRUS CAN  
ESCAPE FROM A LEVEL 4 BIO-LAB**



**BUT CAN'T GET PAST A MASK WITH  
LITTLE DUCKIES ON IT...**

The two lines of text interspersed with the image interpellate an audience into the persona of Condescending Wonka, questioning not only the official COVID-19 narratives but also the intelligence of those who have not yet figured out the contradiction. The meme positions the audience that agrees with its claim on one side against lying officials and people that trust official narratives on the other. The rhetorical deftness of this particular image text meme lies in its ability to swoop an audience, in the course of a single engagement with the meme, into both the line of reasoning set up by the meme and into an interpellated audience identity. Even as viewers might be encountering the meme's reasoning for the first time, having followed the reasoning and accepted it, they come to embody the persona of the one questioning the official narrative and distinguishing themselves from those who don't. The semantic decoding effort demanded by the meme works to enhance the credibility of the meme's claim by interpellating audiences as truth-discoverers. By

advancing claims, memes not only shape public beliefs but also constitute powerful rhetorical audiences, knitting together discursive communities that share memes and bond over decoding and accepting memetic claims.

The boundedness of the image meme above (i.e. its containment with the rectangular box frame) and the parsimony of the rhetorical elements within the meme inhibit central processing and encourage peripheral processing of the meme's claim [27]. The particular rhetorical form of the meme thwarts further questioning into possible reasons why the two supposedly contradictory claims may, in fact, not contradict each other. The success of the meme's argument relies on the implicit warrant that the virus's escape from a protected lab and the possibility of a mask protecting the wearer from the virus are mutually exclusive. The possibility that initial spread was virulent because the virus encountered an unsuspecting maskless population is elided by the memetic structure. Likewise the claim that masks only mitigate but do not necessarily prevent infection is also obscured by the certainty evoked in the meme's juxtaposition of claims. Image memes often simultaneously function as assertive yet weak arguments. Their weakness lies in the fact that their parsimonious form limits elaboration, specifically hiding warrants. However, the parsimony is also responsible for obscuring the weakness of memes. The limited information, visually bounded by the meme's rectangular box, seals a particular conclusion while deflecting attention from warrants that could challenge the meme's claims.

Given the rhetorical power of meme circulation, as elaborated above, to shape public belief, opinion, and sentiment at this time, we urge large-scale collection and analysis of memes that circulate via social media. The ethics, legality, and implications of collecting social media information are a complex and fraught issue, beyond the direct scope of this article, though we discuss some related aspects. Importantly, we have argued previously that personally-identifying information is not necessarily required in collection and analysis of memes [15], elevating privacy protection to a key concern.

Collection and analysis of image memes at scale pose numerous unprecedented challenges that current practices are ill-equipped to meet. This effort will require teams that are curated and structured for efficient and optimal analytical outputs. In the next section, we outline the obstacles that collection and analysis teams are likely to face and, accordingly, recommend practices for structuring such teams, their processes, and outcomes.<sup>1</sup>

---

<sup>1</sup> Outcomes might include situation reports, research products, actionable intelligence, or stewardship of artifact collections for posterity (e.g., the world's largest meme museum).



## Limitations on Artifact Collection by Emergent Teams

Presently, the process for collecting image memes and similar “artifacts” [28,29] for analysis of social narratives, sometimes referred to generally as “artifact collection”, depends almost entirely on the community or communities involved. There are limited best practices, no use-case specific tools, and collection is performed in a wide variety of contexts, sitting at the intersection of myriad fields, including advertising, rhetorical analysis, information operations, misinformation response and intervention, and cognitive security [18]. Given both the plethora of approaches and stakeholders, and the complexity of the phenomena they seek to address, traditional organizations find it difficult to meet objectives in the absence of cooperation with other groups or reconfiguration [18,30]. As such, it is often the case that interorganizational or interdepartmental teams emerge to perform collection and analysis. Here we consider the state of the art for performing analysis of image memes at scale by emergent, interdisciplinary, interorganizational teams seeking to understand the patterns of public discourse around current events. Specific examples below illustrate how memetic analysis can reveal widely-shared public beliefs and opinions on the ongoing Russo-Ukrainian War.

Teams that intend to engage in image meme analysis may follow myriad paths in pursuit of their goals. Below, we offer an overview of the archetypal phases encountered by three common approaches to image meme analysis undertaken by emergent teams, based on the experiences of the authors. The three approaches are listed in order of increasing refinement of methodology and intensity of resource use (time and computational). The three approaches, referred to as **Haphazard Collection**, **Methodological Collection**, and **Automated Collection and Analysis**, are each accompanied by a description of recommended capabilities and affordances which would alleviate their respective pain points (summarized in Tables 1, 2, and 3).

### Haphazard Collection

An emergent team seeking to perform memetic analysis will generally begin its journey through haphazard collection and sharing of memes over some common channel. In the worst case, sharing occurs over email. However, even in the best case, sharing often occurs over an asynchronous chat platform with affordances for setting specific communications channels (i.e., Keybase, Signal, Discord, or Slack). Given that no current tools offer low-code out-of-the-box capabilities for assisting in multi-modal media collection beyond offering storage, the onus is on the organizers and facilitators to expend extraordinary effort to set standards, maintain norms, and motivate members, in order to transform haphazard, general discussion into an artifact collection pipeline. Paradoxically, the social enforcement of such standards and rules too early can be demotivating and reduce enthusiasm for contribution,

effectively requiring a forward-thinking group to choose between losing most information from early collection efforts by not strictly enforcing input standards or risking demoralization and reduced engagement by doing so. It is here that many emergent teams will experience mission drift, a slow process of disintegration, or illusory progress in the form of discussion and collection without definable utility or outcomes.

A team that fails to formalize a methodology will rely on haphazard collection by default and may go through the following stages and challenges (summarized in table 1):

#### **Initial Enthusiasm**

As image memes flow in, the team calibrates a common situational awareness of the information environment through discussion, links to the locations where images were found, and informal references to the events and entities which the images reference explicitly or implicitly. This common situational awareness paired with the social bonding over a shared sense of purpose can create a broad enthusiasm resulting in bursts in collection activity that attracts new members. However, this initial enthusiasm requires organizers and those most committed to the work to now have to focus their attention on onboarding, administrative “housekeeping”, and moderating discussion in order to keep the group’s focus on collection.

#### **Internal Disruption**

Relevant to discussion of this stage is a design principle in engineering, referred to as “separation of concern,” which is described as the adequate isolation of concerns, documentation, and objectives of each system component, such that the component can be error tested and distinguished from the other components with which it interacts [31]. While at the implementation level this design principle specifically refers to functions and blocks of code, at a conceptual level it has been recognized for use outside of engineering domains in guiding design granularity and modularity to improve operational reliability, collaborative productivity, or process visibility [31–34].

Teams often attempt to implement a separation of concerns through the use of multiple communications channels. However, the lack of collection-specific affordances in the chat platforms in-use means there is often an intermixing of

unstructured discussion, structured comments, and collection activity, as well as non-mission focused activity such as professional networking and sharing of unrelated materials, all in the same spaces. This leads to miscommunications and disruption in collation of relevant objects. Given the subjective nature of image memes themselves and the viscerally emotional states they may provoke, no separation of concern between collection and analysis activities can induce political and operational schisms within the team. These schisms do not simply re-enact broader social dynamics within the group, instead, they introduce rhetorical divergence that can induce cascading organizational failures. For example, difference in opinion can lead to bifurcation (i.e., professional relationships, friendships) even though variation in perspective on that same issue might be tolerated by a political party or company (which may have sufficient size and mechanisms for maintaining organizational coherence).

As a practical example, consider a team seeking to understand the discourse around the ongoing Russo-Ukrainian war which has collected an image meme referencing Nazi ideology in relation to Ukrainian para-military groups (see Figure 4). The intentional or unintentional strategic ambiguity embedded in the meme means both the quality of the claims and the claims themselves, are in the eye of the beholder. Thus, the team's own diversity of perspectives, without high levels of cognitive security and trust, becomes a complex threat surface [30]. One member might interpret the meme in Figure 4 as a reference to Facebook's reversal of content moderation policy regarding the Azov Battalion, which, prior to the war, had been classified alongside other white supremacist groups [35]. However, another might interpret the meme as a mockery of anti-fascist movements in the US and libel regarding the Ukrainian military. Given the difficulty of rapid synchronization of priors, unstructured discussion will almost certainly lead to disagreements in analysis which may expose meaningful, underlying sociopolitical and philosophical disagreements. Where teams lack established affordances, roles, norms, trust, and separation of concerns, there is an enhanced potential for such miscommunications which degrade trust and potentially result in disintegration of the team. Instead, a separation of concerns through tool affordances and role-based access is more likely to ensure that the analytic stage is focused on

“trending” rather than “idiosyncratic” interpretations of memetic claims.

### **Overload**

If the team manages to maintain momentum and circumvent tendency toward internal disruption and disintegration, it will next face challenges related to the volume of its collections. Depending on affordances offered by the chat platform in-use, plain-text and links might be exported from chat for analysis or may be searchable. There is rarely a simple method available to teams for the search, categorization, or export of image memes. Even if exported, brute force or manual search, as opposed to query- or attribute-based search, is likely the only method available. This being the case, the more successful the team is at unstructured collection, the less the value of any particular artifact given that the time and effort costs of brute force search increase with the size of collection. Due to this volume-value paradox, success in collections has a direct, inverse relationship with difficulty of analysis.

### **Lack of Visible Progress**

If the team continues operations to collect a relatively high number of artifacts without having resolved separation of concern, search, and collation difficulties through standards setting and compensating controls on inputs - it has likely already undergone some level of “mission creep” or deviation from original goals [36]. As such, the team will likely have no visible markers of progress, which can result in a feedback loop of decreased activity and demoralization of still-active members. If there is no clearly defined process for disintegration or removal of team members based on activity or work requirements, this feedback loop will eventually result in the team ceasing operations as opposed to formally closing.



Figure 4. "Fast Friends" Political Cartoon [37]



Figure 5. Balancing aspects of artifact collection

## Methodological Collection

After general, exploratory collection, a team may then formalize their methodology for managing collections. Potential stages and key challenges are described below and summarized in Table 2.

### Tool Selection

In order to begin collections, the team must choose a tool for artifact collection, which offers accessible input affordances (e.g., form input) for images and text (e.g., Google Sheets, Coda.io). The larger the team, the more likely it is for conflicts to arise during this process of tool selection. Unfortunately, very few tools in common use have accessible affordances for connecting data such that members could continue using their own tools while being able to collaborate on common digital assets.

### Tool Adoption and Configuration

Technical difficulties may occur with adoption and early use of tools or tools may run afoul of organizational security protocols. Moreover, individuals could simply refuse to adopt new tools due to platform or tool overload. Even where members may be using the same platforms already, if members have their own processes or organizational accounts for managing digital assets, difficulties can arise in where and how to store common assets.

During this time, the team may lose members, see a decline in activity and interest, or the team may disintegrate entirely. Poor configuration of the tool's features (e.g., incorrect sharing options leading to inaccessibility) can exacerbate the impact of these challenges.

### Maintaining Information Quality

Given a successful migration to a new tool, the team must decide how to set standards for input, such as including certain attributes, links, or other details. Here, the team has to balance the user-experience of the person performing data input with that of the person who will later perform analysis. The more detail that the team requires, the more opportunities for analysis later - but every additional required detail comes with potentially large costs.

Even a very motivated and interested team may see steep declines in activity where input controls are too strict and details required for collection are too voluminous or complicated. Making inputs optional can create an inconsistency that could potentially demotivate detail-oriented and conscientious team members. Unfortunately, any adaptation will be accompanied by tradeoffs, as the team must balance information quality controls with artifact detail requirements and rate of collection goals, each impacting one another (see Figure 5). Regardless of tool choice, whether or not the team achieves adoption, and the detail of annotation of artifacts, nearly all tools will require the user to switch back and forth between the tool and their browser during collection. This context switching is cognitively expensive, and can result in further declines in information quality and enthusiasm for collection due to poor user experience.

Regardless of scope and controls the team will also have to deal with difficulties in managing duplicates and reducing redundant collections. As the team finds image memes during their collection activities, there is no user-experience friendly method to ensure the item being viewed hasn't already been collected or if the source being accessed has already been searched for artifacts.

In terms of avoiding redundant collections, the team runs into a frustrating challenge. Many of the relevant sources of artifacts are not static web pages, but instead discussion threads that can change over time. Further, the most impactful discussion threads will change rapidly, by merit of their impact. This means the team must risk redundant collections in order to avoid missing new content.

One ameliorating approach is to rely on a "master spreadsheet" or an index to maintain a "single source of truth" for what has been collected and what sources have been searched for artifacts. This process comes with pitfalls that inevitably increase the workload and create inefficiencies. The work involved in duplicate-checking expands with the size of the extant collection of artifacts, making each additional input a contribution not only to the collection, but also to the difficulty of further collections. This constitutes a variant of the volume-value paradox discussed in relation to haphazard collection. Without a means to connect the view the analyst has of the information environment directly to existing collection data, this pain point in analysis is effectively unresolvable.





Figure 6. A collection of image memes; (A) an image meme suggesting one would have to be mentally ill to support Russia's basis for engaging in conflict, (B) an image meme suggesting Russia's handling of protestors is over-aggressive, and (C) an image meme used to represent the status of the relationship between Russia and Ukraine [38].





Figure 7. A collection of image memes; (A) an image meme critiquing profile picture changes in support of trending issues, (B) a photograph of a woman who was arrested by Russian police for holding a blank sign [39] which has been used as an image meme, (C) an image meme suggesting basis for Russian caution in provoking the United States into military action, (D) an image meme comparing Russian and EU negotiation strategy, (E) a screenshot of a subreddit's name and a recent upvoted post, used as an image meme, and (F) an image meme conveying the relationship between Putin and Obama.

#### Information Integrations and Externalization

As suggested above, maintaining information quality and the level of detail useful for analysis will inevitably come at the cost of the rate of artifact collection. Teams at this stage will often seek to address this challenge by externalizing some aspects of collection, for example, opening channels for input from other teams or the general public, or attempting to integrate already existing collections into their own. Both methods come with challenges. A team which externalizes its collection is now faced not only with processing collected items, but also with managing a crowdsourcing solution, which can be time expensive and unreliable. Attempting to integrate existing collections can be equally challenging, as the likelihood of finding an existing collection fit to the same scope as the team's is minimal, and there is no common standard for image meme citation and collection - creating the need to do additional processing work for relevance.

Further, it is unlikely that these collections will contain any provenance data, which will limit analysis a great deal. The image meme found in Figure 6-C is one such example, where the post in which it is found ties the image to the war through its title, which is not included in the image [38]. Similarly, the image meme found in Figure 7-A might be used in a variety of contexts. Without relevant metadata accompanying entry of these image memes into a shared repository, the team would have to rely on pure speculation to identify them as relevant to the memetic discourse around the Russo-Ukrainian War.

Once again, the team is faced with a fundamental trade-off. If they simply accept a slower rate of collection over attempting to externalize some aspects of collection or integrate from existing collections, the team may quickly run into problems stemming from various forms of bias - as no single team can possibly have all of the perspectives necessary to prevent it.

- Centralization bias may come in the form of collection and analysis tending to have an implicit or explicit overestimation of coordination, rational intent, and common direction or theme [40,41]. For example, the team may see the image memes found in Figure 7-C and Figure 7-D as related to the Russo-Ukrainian War or events leading up to it, even though they were both used

in reference to past events and were picked up in collections that could not be properly scoped by time-of-posting due to limitations on search engines.

- The team's own narrative will inevitably calibrate an informational niche which will create a feedback loop of bias in both collection and analysis activity. This will lead them to overestimate the importance of their own perspectives or of particular narratives relevant to their perspectives similar to "overestimation of our own importance", in the context of intelligence analysis [40], or more generally, a salience bias, leading them to prioritize that which stands out to them as relevant given their prior experiences. For example, consider the image meme found in Figure 7-E, which to some may represent a simple mark of support, as opposed to an argument-by-hypocrisy with relevant connections to other image memes such as the one featured in Figure 4.
- Sample bias will leave the team with blind spots. For example, the image memes found in Figure 3 all feature different individuals and settings. However, all of the individuals and settings featured are from the same sitcom, Parks and Recreation. The team might not be aware of this and therefore fail to mark the sitcom as a referenced entity. These kinds of details may seem inconsequential, but they can be critical for certain kinds of analyses, such as those focused on understanding the demographics involved in generating categories of artifacts, understanding the audience that artifacts were created for, or understanding and discovering coordinated activity. As another example, if the team is unaware of recent arrests of Russian protestors carrying blank signs [39], then the image meme in Figure 7-B (and especially its variations which do not include references to Putin) may not be tied as relevant to the war, but instead as merely a comical exaggeration.
- Additionally, the scope of collections itself may result in further blind spots. For example, the image in Figure 7-B is a photograph, and might not be considered an image meme by an analyst, even though it has been used as one (potentially as a protest-meme against over-moderation). As another example, 7-E is a screenshot, which, in

conjunction with salience bias, may go uncollected despite its previously mentioned potential value in analysis. As a final example, the image in 7-B is now tied to a “copy-pasta” which has spread over Twitter and Reddit:

*“A man hands out leaflets on Red Square, and the KGB arrest him. But when they get him to the station, they find that the leaflets are all blank. And he says “Well, everyone knows what the problem is, so why bother writing it down?”*

Given that copy-pasta is a text-based format, it would likely not have been considered in an image meme focused collections scope, thus limiting an analysis of 7-B’s rhetorical impact. The inability to connect collection activity to the collection activity of other groups with differing scope and collection requirements deprives later analyses of a key factor related to rate of spread and impact of content, and a key indicator of coordinated activity.

Without the ability to modularize, externalize, granularize, and connect aspects of collection to such an extent that it limits the bias of the team’s individuals on resulting information quality, biases will likely only be exacerbated by further analysis.

### **Making Analysis Useful**

The team, despite biased collections and challenges in integration and externalization, runs into its final set of difficulties, all related to ensuring the analysis they perform is actually disseminated and of use to others. The problems with rate of collection and analysis means that at this point, any analysis is likely to be a post-mortem of events related to memetic discourse, rather than a map of the current state of a relevant area of the information environment. Of course all analyses are retrospective to some extent, but in the case of memetic discourse, where the state of the environment can shift so quickly - it is likely the case that the information landscape has changed significantly by the time any form of analysis is complete. As such, a team which has been successful up to this point, might simply complete its activities by writing a report or a paper regarding their findings, as opposed to being able to offer any predictive value, actionable

recommendations, or situational awareness to stakeholders. Unfortunately, the value of these information products may be limited to archival or historical purposes unless it had been focused on fundamental research (e.g., understanding mechanism of spread of claims or mutation of image meme format over time). Even the archival value of the work is questionable. While some research work allows value to be salvaged from a project in the form of re-use of datasets, the absence of common standards and provenance data unfortunately means that the generated meme datasets may not be useful to other teams performing similar analyses.

Further, rhetorical analysis of the meaning of image memes is likely to be highly subjective among disparate groups of individuals. This significantly complicates both the process and the resulting utility of the analysis of image memes. Attempting rhetorical analysis may result in unintended, counterproductive outcomes that reinforce the biases of both the team and their stakeholders. A multi-user and multi-community process for determination of rhetorical claims could lend a greater degree of objectivity to the analysis. However, as discussed, these kinds of collaborations face a number of challenges. Focusing on the sensemaking processes of content and semiotic analysis (e.g., “What entities are referenced in this image?”, “Are any latent objects signified by the content?”), increases the likelihood of alignment, but this approach is only enabled by well-structured, voluminous collection.

## Automated Collection and Analysis

After performing initial exploration of structured collection (e.g. as described above), the team may choose to refine and elaborate its approach towards collecting artifacts via computational methods. This approach generally consists of two primary aspects: Data Engineering and Data Analysis. The challenges of automated collection and analysis can be generalized to those faced in Data Engineering and Data Analysis at large; therefore, a comprehensive discussion is beyond the scope of this white paper. However, some challenges of particular relevance to emergent teams performing automated collection and analysis of image memes are discussed below and summarized in Table 3).

## Data Engineering

Depending on the selected data source and desired outcomes, this process could consist of several phases: acquisition, cleaning, formatting, metadata collection, and de-identification.

- **Acquisition.** Data may be directly acquired through a public API (Application Programming Interface) provided by a digital platform, or by scraping memes across different websites and media formats (jpeg, png, pdf, etc). There are various types of API protocols, such as REST (Representational State Transfer) and SOAP (Simple Objects Access Protocol), each of which requires custom connections that vary in terms of difficulty of implementation.
- **Formatting.** Memes should be converted to a common file format for interpretation by a computer vision package such as OpenCV. Image memes may come in any number of sizes and shapes. Therefore, the cleaning pipeline should use uniform resizing parameters that facilitate image and text legibility, and discard memes that do not meet the minimum threshold criteria.
- **Cleaning.** This process can include correcting for errors that occur due to font type and optical character recognition (OCR), removing incomplete files, and potentially removing duplicates depending on the desired outcomes of the analysis. For example, in some circumstances, it may be helpful to understand how many times a meme has been duplicated or how many different groups and/or users engage with a given meme. However, in other circumstances, duplication may bias the results. Hence, the scope of cleaning could entail removing duplicate memes, but will certainly include removing duplication errors that sometimes occur in the scraping process.
- **Metadata Collection.** Depending on the desired outcome of the analysis, collecting additional data about the source of memes might be important. For example, the team might want to record metadata based on the source of the meme or the date that the meme was posted. When focusing on the flow of information, the user ID corresponding to the individual that posted might be of



interest. The team may also want to determine poster demographics (e.g., age, location, religion, political affiliation, level of education).

- **De-identification.** While poster demographics and psychographics can be of value to analysis, some teams may not be able to collect these kinds of data. Moreover, many institutions require IRB certification to use data from human subjects, which always mandates that the subjects are de-identified before the data is analyzed. De-identification can be done by replacing poster names with random variables. However, it is critical to also remove additional demographic information if these data could be leveraged to determine the poster identity (for example only two people over age 60 work at a specific place). These kinds of restrictions can create difficulties for collaborations among teams and limit the applicability of datasets.

### Data Analysis

Ideally, the goal of the analysis is outlined prior to data acquisition and is not established post hoc, as it may be difficult to perform specific analyses if the metadata are not properly collected. While there are infinite possible analyses with respect to forum, user, and user demographics, this section will focus on analyzing content that can be found within meme text, meme images, and the juxtaposition of images and text. Furthermore, methods that could facilitate the automated detection of rhetorical claims in image memes, such as functional annotation, categorization, and semiotic analysis will be explored here.

- **Text Analysis.** Meme text has to be extracted through optical character recognition, which converts images of text into machine-readable text. The semantic content of the text can then be analyzed through any number of natural language processing pipelines, including pipelines for sarcasm detection in memes [42].
- **Image Analysis.** Images can be analyzed for semantic content, such as objects, people, text, scenes, and activities, through any number of image analysis platforms. Gleaning semiotic content from images is more difficult. Recent efforts have attempted to

distinguish regular images from image memes [43] and explore the visual semantics of satire [44]. However, we are only just beginning to unravel the complexity of semiotic and rhetorical content embedded within image memes. While manual annotation of memes by humans can be very helpful in creating a training set of data and broad categories of image memes, manual annotation is a time-consuming task with a subjective nature that can yield variable results [45]. Machine learning can reduce the time burden of annotation, and can also be useful for semantic association and classification of images [46].

- ***Juxtaposition of Images and Text.*** The interplay between text and images in multimodal content can be quite complex, and can have a significant impact on the essence of a meme. The relationships between pictorial and textual concepts and entities are characterized by metrics that include cross-modal mutual information (conceptual overlap) and semantic correlation (meaning overlap) [47]. Moreover, content within images and text can contribute to the rhetoric in the meme in a number of ways. Meaning can be derived largely from the text, as in Figure 3, where replacing the image with any number of images is not likely to interfere with the overarching claim. In other memes the meaning is mostly image-based, as in Figure 7-E; the caption is not necessary within the context of the current Russo-Ukrainian war. Meaning can also be derived equally from image and text, as in Figure 1, where replacing either would significantly impact the underlying claim. The relative importance of images and text has been quantified in a metric called “status” [48]. Deep learning has been used to successfully categorize image-text relationships based on the metrics described above [48].
- ***Categorization of Memes.*** Analysis of the entire memome is as daunting a task as the analysis of the entire genome was at the turn of the 21st century, and we can learn a lot from the successful methods that emerged in the automation of genomic analysis. Automating the ability to understand any meme, from any source, would have to begin with coarse grained analyses (i.e. broad categories) which are then later refined to higher levels of detail (i.e. rhetorical claims).



Machine learning methods could be used for meme categorization, with appropriately labeled training data, and the largest and most successful crowdsourced set of manually annotated images and text, Wikipedia, could serve as training data to this end. Wikipedia image and text content is conveniently labeled with category terms that start off at broad levels and become increasingly refined [48]. Wikipedia categories may be a good starting point for the development of categories for a meme ontology, a hypothetical memetic analogue to the gene ontology that is broadly used for functional genomic annotation. Within the memome, there is the potential to uncover memes related to myriad stable and provisional categories. As in the genome, studying the memome will uncover motifs, i.e. recurring patterns with well-defined functions, that belong to categories (usually more than one). For example, based on the images and text, the meme in Figure 4 could belong to categories such as “comic strip”, “Nazi Germany”, “white supremacy”, “Ukraine”, “Russo-Ukrainian war”, “military”, and “hug”. While this level of categorization is far from a rhetorical analysis, it can be useful in the detection of meme types in order to understand the topics individuals or groups are interested in discussing.

- **Detection of Rhetorical Claims.** Rhetorical analysis of the claims in image memes is difficult for both humans and computers. Each person brings many years of their own unique prior cultural experience into the analysis of a meme. Computational analysis can benefit from human annotation of rhetoric in training data, although arriving at a definitive claim for every image meme is a daunting task that may have unwanted outcomes (see *Methodological Collection: Making Analysis Useful* above). In understanding rhetorical claims, there are many layers of analysis that build upon one another. At the most basic level, the concepts and entities present in an image meme are identified. Recognizing entity relationships then facilitates semantic understanding. Automated analysis of memes could be successful to this end; however, sensemaking beyond the semantic level requires a human-in-the-loop to create annotated data. Uncovering latent substance in memes requires

understanding the potential alternative significance of the meme's concepts and entities. Although semiotic comprehension can be deeply personal (e.g., not everyone thinks of Grandma when they smell lilacs) there are also signs and symbols that have been broadly adopted. An annotated dataset that links semantics and semiotics would advance our ability to uncover potential latent significance within memes, and advance efforts towards the automated detection of rhetorical claims.

| <b>Haphazard Collection</b> |                                                                                                                                                                                                                                                                                   |
|-----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Stage</b>                | <b>Key Challenges</b>                                                                                                                                                                                                                                                             |
| Initial Enthusiasm          | <ul style="list-style-type: none"> <li>• No Designated Roles</li> <li>• No Affordances for Structured Contribution</li> </ul>                                                                                                                                                     |
| Internal Disruption         | <ul style="list-style-type: none"> <li>• No Accessible Tools Designed for Collection Activity or Implementing Related Compensating Controls</li> <li>• No Separation of Concerns</li> <li>• No Role-Based Access</li> <li>• No Affordances for Structured Contribution</li> </ul> |
| Overload                    | <ul style="list-style-type: none"> <li>• No Affordances for Multi-Modal, Semantic Search</li> <li>• Limited Affordances for Structured Archiving</li> </ul>                                                                                                                       |
| Lack of Visible Progress    | <ul style="list-style-type: none"> <li>• Mission Creep</li> <li>• Lack of Ability to Measure Progress</li> </ul>                                                                                                                                                                  |

*Table 1. Key Challenges in Haphazard Collection*

| <b>Methodological Collection</b>             |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|----------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Stage</b>                                 | <b>Key Challenges</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| Tool Selection                               | <ul style="list-style-type: none"> <li>• Tool Overload</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Tool Adoption and Configuration              | <ul style="list-style-type: none"> <li>• Tool Overload</li> <li>• Lack of Tool Interoperability</li> <li>• Lack of Accessible Connectivity Affordances Among Teams and Platforms</li> </ul>                                                                                                                                                                                                                                                                                             |
| Maintaining Information Quality              | <ul style="list-style-type: none"> <li>• Poor User Experience of Collection Activity</li> <li>• Manual Connections Among Tools and Datasets</li> <li>• No Accessible Tools Designed for Collection Activity or Implementing Related Compensating Controls</li> <li>• Limited Affordances for Structured Archiving and Annotation</li> <li>• Context Switching Between Collection Tools and the Browsing Environment</li> <li>• Inability to Detect Exact and Near Duplicates</li> </ul> |
| Information Integrations and Externalization | <ul style="list-style-type: none"> <li>• Lack of Accessible Connectivity Affordances Among Teams and Platforms</li> <li>• Lack of Common Standards for Data Sharing</li> <li>• No Common Citation Method</li> <li>• Limited Affordances for Structured Archiving</li> <li>• Poor User Experience of Collection Activity</li> <li>• No Accessible Tools Designed for Collection Activity or Implementing Related Compensating Controls</li> <li>• Bias in Analysis</li> </ul>            |
| Making Analysis Useful                       | <ul style="list-style-type: none"> <li>• Collection and Analysis Activity Not Fast Enough to Provide Situational Awareness in Real-time</li> <li>• Insufficient Standardization or Provenance Data to Allow for Reusability of Datasets</li> </ul>                                                                                                                                                                                                                                      |

Table 2. Key Challenges in Methodological Collection

| <b>Automated Collection and Analysis</b> |                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Stage</b>                             | <b>Key Challenges</b>                                                                                                                                                                                                                                                                                                                                                                                                                                 |
| Data Engineering                         | <ul style="list-style-type: none"> <li>• Lack of Accessible Connectivity Affordances Between Teams and Platforms</li> </ul>                                                                                                                                                                                                                                                                                                                           |
| Data Analysis                            | <ul style="list-style-type: none"> <li>• Insufficient Human-Annotated Data to Extract the Significance from the Relationships Between Text and Images</li> <li>• Lack of Common Standards and Ontology to Leverage in Connecting Image Memes to Functional Categories or Topics</li> <li>• Limited Human-Annotated Data Connecting Semantics and Semiotics (which could help with the automated extraction of latent topics within memes).</li> </ul> |

*Table 3. Key Challenges in Automated Collection and Analysis*

# Requirements and Recommendations

Below we provide requirements for systems that might alleviate many of the key challenges for emergent teams tasked with analysis and annotation of image memes, and offer example use-cases if these requirements were made available.

## Information Where it Matters

Analysts need access to information where it matters. Being able to access details on-site about existing analyses, collected artifacts, and to simply see whether or not a relevant object (e.g. an entity reference, an image meme, a thread, or a web page) has already been the subject of collection activity would immediately and unambiguously reduce most redundancies in collections activities. This could be achieved through custom browsers or, to avoid impacts from requiring platform adoption, through web and document annotation affordances. Tools that place collected information and analysis alongside the content itself would allow analysts to see their own shared lens on the internet without requiring content providers to adopt common standards.

## Example Use-Cases

- Enriching images that have already been collected with summary information and links to extant analysis and related artifacts.
- Marking discussion threads and webpages with summary information about when they were last visited and what had been collected from them.

## Dynamic Web Annotations

The ability to annotate and enrich content with links to and presentation of existing information would provide numerous benefits to analysts, including preventing redundant analysis and duplication during collection. However, the fact that some sources of artifacts are impactful because they are expected to change often (e.g., discussion threads) creates numerous challenges. The ability to annotate, or attach, 'functions' or 'automations' (i.e., triggers to run scripts) to web pages, which dynamically update their data and contents as opposed to presenting static content (e.g., text), could alleviate many of these challenges and add new analysis capabilities.

#### Example Use-Cases

- Updating analysts when content at a given URL has changed substantially or when content with certain characteristics have been detected.
- Tracking changes to sentiment or rate of engagement, indicating the presence of recent, potentially valuable artifacts.
- Performing image-similarity searches to “track” already collected images as they spread across the internet.

#### Proximal Collection and Tagging Affordances

The constant context-switching required for use of most tools offering collection affordances decreases productivity, information quality, and user experience and engagement generally. Providing collection affordances which are proximal to the source of artifacts would greatly enhance efficiency and user experience, thus improving information quality, rate of collection and productivity, and general engagement. For example, Paperpile, a reference management platform for academics, has used HTML injection and HTML template standards in order to insert artifact collection affordances on content of both static pages and search results in order to improve the productivity and efficiency of researchers (See Figure 8). Similar approaches using community-related features and more general and customizable templates and object standards could vastly improve digital artifact collection processes outside the context of academia.

#### Example Use-Cases

- Injection of collection affordances on web pages without requiring the permission of the web page’s owners or their adoption of common standards. This would enable users to quickly add artifacts for processing.
- Providing a computational basis for instituting compensating controls on collections. If the collection affordance is proximal to the source, a great deal of metadata about the artifact can be collected automatically, greatly reducing the amount of

time taken per artifact while greatly increasing the information quality.

- Providing users and communities with the opportunity to share in common templates for computational detection and collection of different kinds of artifacts and their metadata on different websites (e.g., articles, posted comments, images).
- Offering tools that allow teams to set local scope and standards for collection and processing. As discussed, a multi-community process for annotation of semiotic content in images could lend a greater degree of objectivity to the analysis of rhetorical content. Further, the ability to set clear standards for these annotations (i.e. what and how annotation will be executed) could increase the longevity and applicability of the resulting datasets, and help communities choose the appropriate scope and level of detail for collection activity. For example, while analysts are unlikely to disagree about the semantic content within the images or text, finding consensus on rhetorical claims and the meaning of image memes may be impossible in some cases. Instead, comprehending the latent representations (i.e. semiotic content) in image memes may provide a useful common ground where human annotation can offer insight into the sensemaking that precedes the determination of rhetorical claims. Answering the question “What are the hidden representations, if any, that this meme signifies?” facilitates multiple answers. An analyst that didn’t have the appropriate background to uncover latent meaning could answer, “None.” Instead of forcing the analyst to deduce a single claim, semiotic-focused collection and analysis affords a softer approach more amenable to multi-user and multi-community analysis. In addition, an annotated dataset linking semantic and semiotic content would be re-usable by future analysts, and could be leveraged as training data to automate the process of sensemaking that underlies more complex forms of analysis (e.g. analysis of rhetorical claims).



## Structuring the Information Commons

### X – Tracking Public Sensemaking through Rhetorical Annotation of Image Memes



Figure 8. A 4-part graphic representing how Paperpile is used for digital artifact collection. (a) A detected artifact which has not yet been collected, (b) an artifact's attached collection affordance being used to search for related metadata, (c) how an artifact indicates that it has already been collected, and (d) how an artifact is represented in aggregate with other collected artifacts, with redactions of personal information.

## Semantic Multimodal Search

Semantic multimodal search may be among the most needed and difficult-to-fulfill requests for analyst capabilities, as it has constituted a challenge for as long as humans have been actively attempting to refine methods for search, sort, and summarization of information across myriad contexts, from intelligence practice to library archiving [49–51]. While AI methods are presently the most popular approach to semantic search and collation, human annotation and analysis is the oldest, most auditable, and arguably the most reliable method available, despite its drawbacks [15,49,50,52]. The most notable of these drawbacks may be speed and scalability. These drawbacks can be addressed with crowdsourcing solutions, which adds new difficulties, such as the potential for disagreement regarding both the standards for and the resulting annotations.

However, given that the intent of artifact annotation in this case is to understand underlying claims, referenced entities, and cultural references, disagreement in well-structured and standardized annotation instead becomes valuable data for analysis [15,53]. Further, artifact annotation allows teams which externalize their collection and annotation activities to transcend local narrative bias and inherently limited cultural knowledge. Human annotation, at scale, facilitated by web annotation in combination with image-similarity algorithms, AI, and traditional ontological approaches, could yield the necessary semantic multi-modal search for aggregate analysis of highly subjective content, such as image memes [15]. These approaches could also provide training data to further externalize annotation to automated systems, enable connectivity between content and concepts, and offer the basis for identifying common hidden states, themes, claims, and references running through disparate content [15].

### Example Use-Cases

- The PageRank algorithm is used to discover “which nodes [in a network] are important” whereas Reverse PageRank is “often used to determine why a particular node is important” [54]. While PageRank, and its cousin Reverse PageRank, are known for their use in internet search, the underlying mathematics are entirely general, and have been used in areas such as bioinformatics, neuroscience, literature and bibliometrics, and even sports [54]. With structured annotation, time-series metadata, and shared catalogs of image memes, there are a number of variations of search that become possible, from searching for which memes and themes or relevant URLs are or were important or trending, to collating memes which are

relevant given attribute-based search queries. Further, these forms of search may provide actionable insights for those performing or informing interventions, as they could point to image memes with highly specific attributes which can act as “stepping stones” or defenses for changing or maintaining beliefs, respectively, given audience characteristics and interests [18].

- With an annotated catalog of image memes and annotated automations for detecting image memes, it may be possible to not only search for specific memes or memes with specific attributes, but to search for where a meme has appeared and for its potential succeeding iterations. Similar to tagging and tracking methods in ecology, image memes could be tracked in real-time or through time-series projections as they move across the internet [15]. This kind of tracking, when paired with annotated automations for sentiment analysis over time, allows for search by sentiment impacts and trajectory potentials - offering the basis for dashboards, early warning systems, impact projections, and other predictive analytics and situational awareness systems [15].
- While search is generally associated with returned results from a specific query, there are numerous other forms of exploratory search, such as exploring connections in networks, interactive visualizations, and situational awareness tools [55]. Well annotated catalogs of image memes could allow for various exploration-facilitating visualizations, such as rhetorical and argument maps [56], process maps of meme lifecycles and transitions [32], and interactive network graphs [55].
- In addition, there could be forms of exploratory search as would be found in bioinformatics, such as network and graph exploration via functional annotation within the genome [57,58], allowing for analogous “memoinformatics” driven search within the memome. As functional significance has been imparted onto the human genome, our understanding of organismal biochemistry and physiology has progressed in turn. A detailed, hierarchical annotation of the memome that includes concepts, entities, semantic categories, semiotics, rhetoric, and the underlying links between these types of content, could vastly increase our understanding of

multimodal human communication, sensemaking, and cognition.

## Separate Platforms, Shared Data

The limitation of current tools and the “one platform to rule them all” [59–61], or the “one app to replace them all” paradigm [62], has created and exacerbated challenges for teams attempting to effectively and efficiently collect, analyze, and implement common standards and controls for their data. As discussed, where individuals or their organizations come to the team with tool preferences, storage protocols and nomenclature, and rules about what tools can be used, teams are not only disjointed by default, but can disintegrate before they even have a chance to begin work. In addition, the team needs to be able to externalize some aspects of collection and integrate data with other teams that allow opportunities for selective disclosure, and create separations of concern within the team based on role and information needs. Even if there was a tool available which offered all of the affordances and capabilities the team could possibly need, it may still be more efficient, inclusive, and practical to “[meet] people where they are” [63,64].

Many companies are beginning to adopt a more collaborative approach, attempting to create new value through open standards and specifications. This open approach allows for digital asset and signal exchange with third-parties, and contributes API (application programming interface) connections and affordances to the “API economy” [65,66]. The API economy is composed of integration platform as a service (iPaas) [67,68], general automation platform (GAP) [69], cloud-based integration [70], and digital ecosystem [63] approaches. Companies such as Zapier [71], IFTTT [72], Make (formerly Integromat) [73], DOMO [74], MuleSoft [75], and Workato [70] provide the ability to create automations using “trigger-action” and “if-this-then-that” frameworks [71,72] and data pipeline integration capabilities that connect and incentivize the creation and use of APIs. These capabilities reduce the time-to-impact and the development and personnel costs of linking and maintaining data across multiple organizations, applications, and services [67,68]. However, not all platforms allow for these integrations and there can be a significant amount of work involved in enabling their use.

Incentivizing API economy participation through the use of common standards and market mechanisms for exchange of data with third parties might increase venture capital attention on companies using more open approaches with their data, thereby increasing the number of interoperable platforms. It can also offer alternative revenue streams to platforms which are currently disincentivized from sharing their data due to their reliance on dwell time related revenues (i.e., advertising revenue). For example, while the meme collection site “Know Your Meme” is branded as a

meme research and collection platform [76], it is also owned by a media holding company that collects advertising revenue from its underlying brands [77]. If Know Your Meme were provided with low-cost mechanisms to bring its collections to the API economy, the potential to offset ad revenue losses and create new business value might incentivize making data easily available to third-parties for other meme search, research, and curation functions. A more robust and accessible API economy could have large impacts on interorganizational work.

### Example Use-Cases

- With accessible, low-code API integrations paired with methods for exchanging information about data standards and controls in use, the same mechanisms which help teams share calendar information between project management tools could be used to share data collections between teams in real-time.
- Role-based access controls and privileges can be difficult to manage and keep track of [55]. If teams have the ability to create ad hoc real-time connections between platforms, then, by proxy, they can implement highly complex role-based access and affordances simply by restricting membership on certain tools. This could allow for complex intelligence pipelines in which aspects of collection and expert analysis could be separated within teams, externalized to other teams with varying incentive sets (e.g., crowdsourced collection), or done in real-time collaboration with other teams and organizations.
- Some teams may have information they cannot or do not want to share, and, as discussed, some teams may have information they cannot receive without running afoul of internal ethical or other controls. API integration capabilities paired with clear opportunities for bidirectional selective disclosure could allow teams that would otherwise be unable to share information to communicate and collaborate. Further, allowing teams to set “terms of use” and related information on their offered or requested data would allow for a new level of transparency for users in how their shared information is used and governed [33].

# Conclusion

As the world's increasing complexity drives conflict into increasingly abstract spaces [78], emergent teams involved in digital discourse analysis have a vital role to play in helping organizations maintain situational awareness and synthesize disparate perspectives generally. Here we advanced previous work on rhetorical analysis of image memes by presenting several archetypes of emergent teams involved in analysis, describing their inherent challenges, and suggesting recommendations for future systems design.

In the realm of image memes, which allow for unparalleled strategic ambiguity and plausible deniability, emergent teams may be the only viable approach to sensemaking in the digital rhetorical ecosystem - as no single organizational configuration can capture all of the symbols and cultural knowledge necessary to understand or estimate the significance of the content present. As discussed, the state of the art of image meme collection by emergent teams is not commensurate with either stakeholders' or the team's needs, despite direly needed affordances being well within technological reach. In short, there is a chasm between "how it is done today" and "how it could be done" that is not proportionate to the gap between "what is available" and "what is possible" (see Figure 9).

Digital ecosystem and API economy approaches seem to be a viable route to addressing many of the challenges discussed, and for enabling and contributing to the web and document annotation approaches which address the remaining challenges. API economy approaches have gained traction in recent years, and are now being applied across various areas of the market including agriculture [67,73], engineering [70], research [72] and marketing [69]. However for API economy approaches, lack of data standardization and integration capabilities remain a problem. This problem is not specific to challenges faced by emergent teams, and addressing it could be beneficial to a variety of sectors. Polling has suggested that the average enterprise uses more than 1,200 applications [79], and that an "average knowledge worker" is using up to 28 different applications [80] and is toggling between applications up to 10 times per hour [81]. According to Deloitte's 2021 Chief Procurement Officer Survey, among the top two barriers to effective technology implementation are data quality and poor integration capabilities across applications [82].

In addition to many other domains being able to share in the benefits of developments that would resolve challenges for emergent teams conducting image meme analysis, other domains can benefit from the resulting analysis. Resolving these challenges using the approaches discussed in this white paper could result in new claims-based methods to identify counterpublics and communities that have no formal affiliations, the capacity to identify hidden states and themes running



through public discourse, and to provide early warning systems indicating where streams of memes might constitute the precursor for groups to converge on (potentially violent) action.

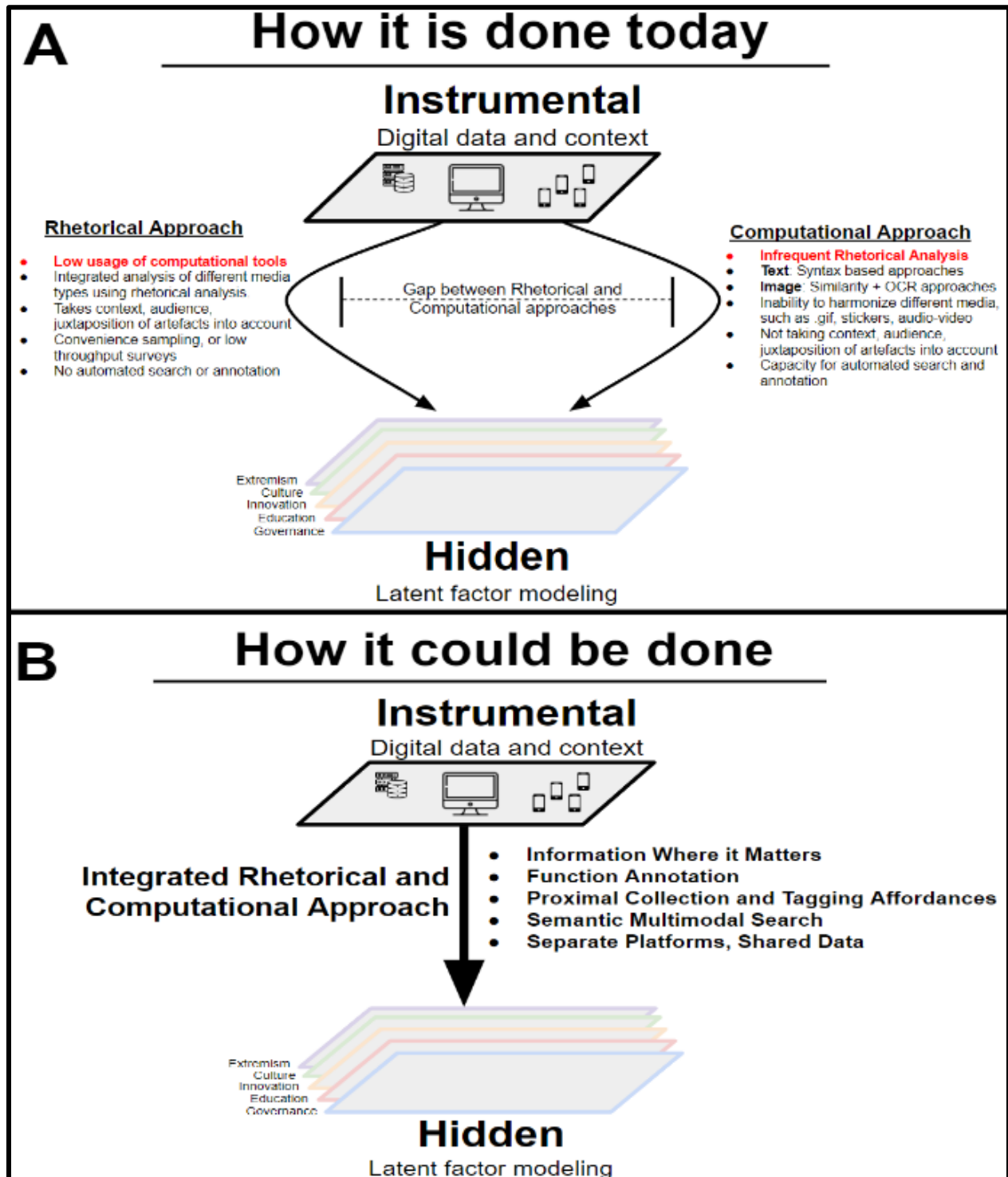


Figure 9. Graphical overview for computational and rhetorical analysis pipelines. A) How it is done today, and B) How it could be done.

## **Funding and Acknowledgements**

R.J. Cordes is funded through the NSF Convergence Accelerator Trust and Authenticity in Communication Systems Program (NSF 21-572), under award ID #49100421C0036 and is supported in research efforts through a Nonresident Fellowship with the Atlantic Council on appointment to the GeoTech Center.

Daniel A. Friedman is funded by the NSF program Postdoctoral Research Fellowships in Biology (NSF 20-077), under award ID #2010290.



# TrustFinder

## Recommendations for a Community-Based System for Finding Trusted Sources and Evaluating Claims

**Lead Designer:** R.J. Cordes

**Contributors:**

Scott David J.D., L.L.M.

Daniel Friedman, PhD

**Consultants:**

Mridula Mascarenhas, PhD

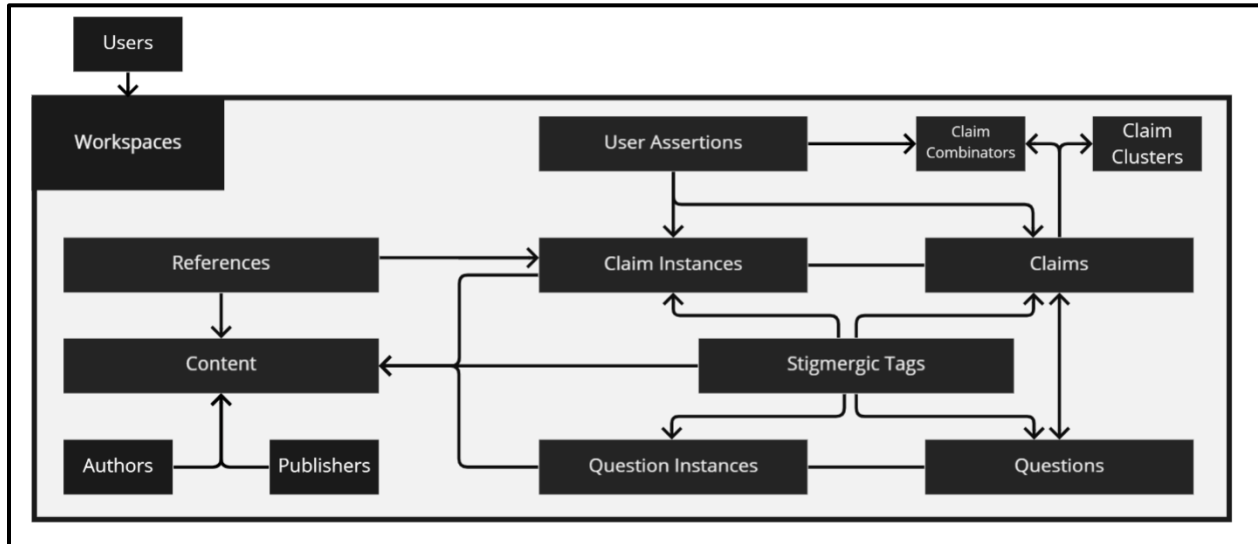
## Executive Summary

---

There is a broadly recognized need for better situational awareness within the information environment. Each year, millions of articles, books, documents, and datasets are published. Amidst this flood of information, even those with significant experience and expertise in the knowledge economy are struggling to evaluate and vet claims. This document builds on the feedback of dozens of experts across myriad fields submitted to the University of Washington Applied Physics Lab's Verified Information Exchange Environments Program, to present recommendations for a sociotechnical system, "TrustFinder", for collaborative management of the information supply chain. TrustFinder implements controls and standards, web and document annotation affordances, argument representation frameworks, and crowdsourcing design principles in order to harness the work of global research communities. The ultimate goal of TrustFinder is to structure the information environment to such an extent that it enables users to find trusted sources of information and rapidly assess concepts and claims.

---

## System Overview



*TrustFinder Environment Primary Components*

TrustFinder has one primary category of actors, “**Users**”. **Users** scope the global information environment (i.e. the internet) by creating, sharing, and adding other users to “**Workspaces**”, which represent “information commons” intended to facilitate projects related to sensemaking (e.g., a research paper, studying, exploration of a topic). **Users** within **Workspaces** use web and document annotation affordances (i.e., the ability to “mark-up”, “highlight”, take notes at the edge of a document or webpage, or to otherwise enrich content) in order to structure the information environment. **Users** and **Workspaces** can further assign “trust scores” representing expectations of the quality and intents of specific authors and publishers, as well as of the assertions and annotation contributions by other **Users** and **Workspaces**. With such enrichment tools, **Users** structure the claims and concepts they encounter in order to make the information environment more navigable and searchable, reducing future redundant work for themselves and others related to evaluation and vetting of claims and allowing for evaluation and mapping of the information supply chain (i.e., where claims originate and where they have spread).

In the TrustFinder environment, a **Workspace** can be populated with different classes of interconnected informational structures, each contributing to enrichment of the rhetorical landscape. Below are the 10 primary classes of informational elements.

**Content.** Information sources such as a book, paper, article, or video. Content provides a container for **References**, user-added metadata, **Claim Instances**, and **Question Instances**.

**References.** Relationships between **Content**, such as direct citations, are stored as references. These reference objects can be used to map the connections between **Content** and between **Claim Instances**.

**Claim.** A statement about the world can be represented as a Claim. Claims exist outside the context of any specific **Content**, can be represented using various phrasings, and can be connected to other objects. A **Workspace** can be prompted or initiated using a Claim, as a basis to help scope related work (i.e., *this work* is related to the investigation of *this Claim*). The Claim's most important feature is its ability to be connected to other Claims through **Claim Combinators** and **Claim Clusters**.

**Question.** Explicit or implicit Questions are represented by an informational structure which can be connected to both **Claims** and other Questions. Similar to **Claims**, they can exist outside the context of any specific **Content**, can be represented using various phrasings, can prompt or initiate a **Workspace**, and be connected to other objects. Questions have an important relationship with **Claims** as **Claims** can both be responses to, or prompt, Questions.

**Claim Instance.** As opposed to **Claims**, which exist outside the context of any specific **Content**, Claim Instance objects represent the instantiation, or appearance, of a particular **Claim** within a specific area of a piece of **Content** (i.e., within a particular sentence). Claim Instances can be connected to the appearances of its **Claim** within other pieces of **Content** through **References** (i.e., where there is a direct citation related to the appearance of the Claim Instance within the **Content**).

**Question Instance.** Similar to the **Claim Instance**, Question Instances are simply instantiations, or appearances, of a particular **Question** within a specific area of **Content**.

**Claim Cluster.** Claim Clusters are a simple container for **Claims** that are related in terms of their relationship to some other

object (i.e., this set of **Claims**, if all true, support this other **Claim**).

**Claim Combinator.** Claim Combinators are containers for describing the relationship between a **Claim** or a **Claim Cluster**, and a **Claim**, **Claim Cluster**, or another Claim Combinator. Claim Combinators are categorized as (i) supports, (ii) refutes, (iii) generalizes, (iv) modifies, and (v) relates to.

**User Assertion.** In addition to collecting **Claims** and marking **Claim Instances**, **Users** can also make their own assertions *about Claims* and **Claim Instances**. User Assertions are essentially a special form of **Claim Combinator**, on which they are attaching their name. User Assertions attached to **Claims** will appear within the workspace when **Users** access the **Claim**, as well as when they access instances of that claim (**Claim Instance**), allowing for contextualization of particular claims. User Assertions attached to **Claim Instances** will only appear on that particular **Claim Instance**, allowing for nuanced warnings or endorsements (e.g., if you wanted to find support for this **Claim**, this particular piece of content may not be the place to cite it from, as it is not a strong argument or works from faulty data).

**Stigmergic Tag.** Stigmergic Tags are a combination of predefined and **User-defined** tags used to further assist in querying and navigating **Workspaces**. Stigmergic Tags provide users with highly structured methods for communicating requests, directing attention, providing feedback, and marking the presence of key concepts or entities. Stigmergic Tags can be connected to nearly all other informational structures within the TrustFinder environment, including other Stigmergic Tags.

## System Purpose

The primary purpose of the sociotechnical system, “TrustFinder”, is to facilitate collaborative structuring of the information environment, enabling users to find trusted sources of information, which in turn enables them to rapidly assess concepts and claims. The secondary purposes include:

- providing infrastructure and data for the future of reference management systems,

- mapping and understanding the “supply-chain” of claims, and
- “capturing” the value of discourse and disagreement.

## Scope

This document intends to provide recommendations for the key components of the TrustFinder environment and for their structure and relationships from the perspective of knowledge management and behavioral modification in the context of crowdsourcing solutions, as well as to offer (i) relevant background information regarding the basis of these recommendations and (ii) a discussion of the potential implications of their implementation. It does not provide (i) exhaustive recommendations for user experience or presentation, or (ii) detailed recommendations or technical requirements for data structure or security assurances. Names for components within these recommendations should be adapted to optimize user experience and onboarding. **A developed TrustFinder system may differ substantially from recommendations given technical constraints or opportunities.**

## Structure of this Document

This document consists of (i) a Systems Definition section concerned with the components of the TrustFinder system, separated into 5 segments: (a) Agents and Workspaces, (b) Media, (c) Claims, (d) Questions, and (e) Reputation; (ii) an Implications section, which discusses the potential implications of explicit and implicit mechanisms within the recommended system; and (iii) a Background section, which provides a synthesis of theory and frameworks used to inform design. Within the Systems Definition section, explicit mentions of system components are bolded outside of their respective sections for reference purposes. Component attributes, related interfaces, and other objects are bolded and/or italicized for clarity where necessary.

## Definitions and Word Usage

**“Combinator”** is used within this document to describe an empty interface that allows a set of objects which do not necessarily share common methods or attributes to be used in fields which establish complex relationships between said objects. Borrowed and adapted from library organization design patterns within the Haskell programming community, wherein “combinators” are used to combine values of a given type in various ways to create more complex, and context-rich instances of that type.

**“Decorator”** is used within this document to describe an empty interface used in order to allow a set of objects which do not necessarily share common methods or

attributes to be used in a field without modifying the behavior or structure of that object.

**“Genuine Presence Testing”** describes the set of security assurances which use biometrics, computer vision, and geographic data related approaches to authenticate the presence of a particular person using a device.

**“Interface”** is used within this document to refer to general “programming interfaces” unspecific to any language, i.e., (i) an object which enables polymorphism, (ii) an object which represents a contract fulfilled by the ability to perform some function or deliver some attribute, or (iii) a vehicle for the inclusion of multiple classes of object within a field which requires type assertion (i.e., a Decorator).

**“TrustFinder Environment”** is used to describe the space of engagement with the common TrustFinder infrastructure generally, through workspaces or otherwise.

# System Definition

## Agents and Workspaces

### User

User refers to users of TrustFinder, individuals who are seeking to enrich web and document content and collect and evaluate claims. Users must engage with security assurances (e.g., genuine presence testing) in order to register and engage with certain aspects of the system (e.g., **User Assertions**).

#### *Invitation Tree*

**Users** can invite others to TrustFinder. Each **User** invited, and each invited by those invitees, up to 6 degrees of separation, are included within the inviter's Invitation Tree with their respective degree of separation (see Figure 1, degrees of separation). Invitation trees are not visible to other **Users**, and are used primarily to provide foundation for network-related impact scoring. It is recommended that in the future, there are methods devised to allow users to share the credit of invitation of new members and that invitation trees related to specific workspaces (i.e., tracking invitations to workspaces, as opposed to the platform as a whole), are implemented.

#### *Real World Credentials*

**Users** can attach real world credentials, such as higher education degrees and professional certifications to their account.

#### *Pseudonyms*

**Users** can create multiple Pseudonyms (i.e., usernames) for use within the TrustFinder environment. **Users** may selectively disclose which credentials, if any, and what aspects of those credentials to attach to Pseudonyms (e.g., "a Master's degree in computer science" as opposed to "a Master's degree from *this* university"). Pseudonyms may be used to engage with any activity within the TrustFinder environment with the exception of **User Assertions**, which must be tied directly to the **User's** account.

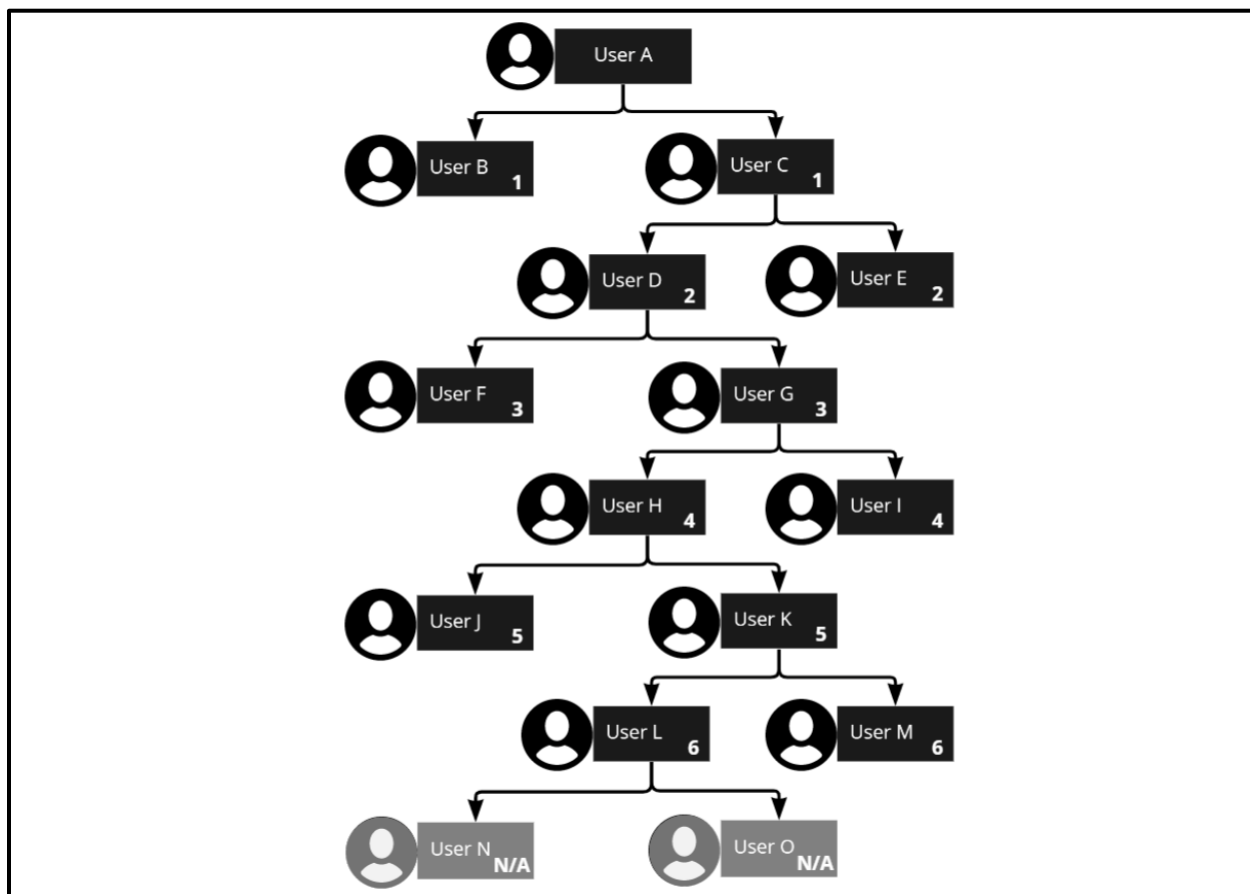


Figure 1. User Invitation Tree

## Workspace

Workspaces are the basis for engagement within the TrustFinder environment. **User's** may create and be invited to multiple Workspaces. Workspaces represent projects related to sensemaking (e.g., a research paper, studying, exploration of a topic), and are used as containers for objects relevant to that work.

- Workspaces may be instantiated using a **Claim** or **Question** (e.g., where research on a particular question is the driving motive behind intended work) and can be populated with **Workspace Objects** based on the presence of certain **Stigmergic Tags** within those objects, as well as other conditions (e.g., time period, object type).
- Workspaces make use of **Clearinghouses** in order to manage the dynamic import and export of digital goods (i.e., **Workspace Objects**).
- Workspaces may be given their own sets of **Entity Tag Types**, **Custom Tag Types**, **Contribution Trust Scores**, and **Assertion Trust Scores**.



- Workspaces have two classes of **User** within, *Administrators* and *Members*. Administrators have permission to manage high level aspects of the Workspace, including setting **Clearinghouse** import and export conditions, **Entity Tag Types**, **Custom Tag Types**, and the Workspace's **Contribution Trust Scores** and **Assertion Trust Scores**. It is recommended that, at the outset, role and permissions related governance are kept as simple as practicable, while allowing for opportunities to adapt and related features in response to need and interest.

### **Workspace Object**

Workspace Object is a decorator for the following objects: Authors, Publishers, Artifacts, URLs, Content, References, Claims, Claim Instances, Claim Combinators, Claim Clusters, Questions, Question Instances, Question Combinators, User Assertions, and Stigmergic Tags.

### **Clearinghouse**

The Clearinghouse represents the import or export channel for *Workspace Objects* between the **Workspace** and another **Workspace** or set of **Workspaces**. It contains conditional statements for managing the dynamic (i.e., active or ongoing) import and export of **Workspace Objects**, and a *Buffer*. The *Buffer* is used where **Workspace** administrators opt to approve items individually before they are added to the local **Workspace** environment or before they are available for export to external **Workspace**. Clearinghouses are directional, with export-oriented Clearinghouses making digital goods “available” based on conditional statements to the **Workspaces** specified, allowing those **Workspaces** to create respective import Clearinghouses in response; and with import-oriented Clearinghouses acting as “listening posts” waiting for exports to be made available.

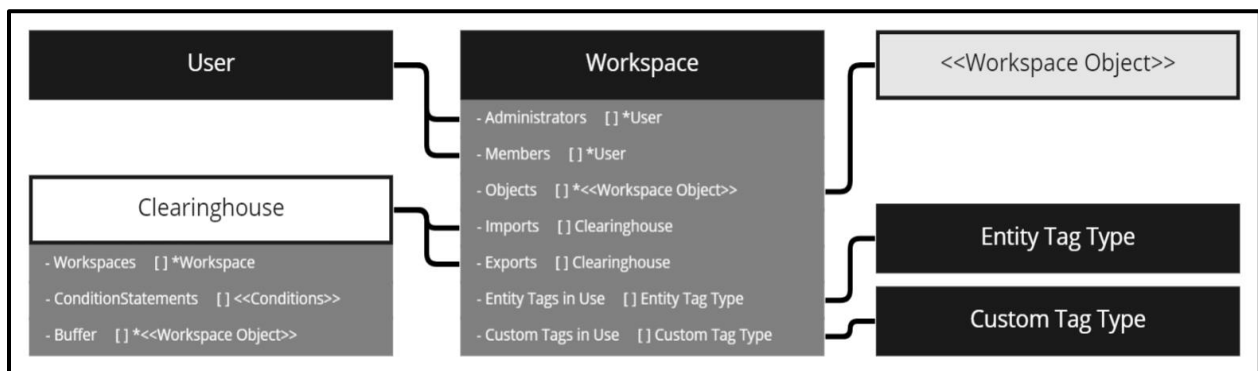


Figure 2. Workspace relationships

## Media

### Author

An Author object is used to represent authors responsible for **Content**. Authors can be assigned to **Content** (for attribution). **Users** and **Workspaces** may assign Authors an **Assertion Trust Score**. Authors may be given additional attributes over time, such as funding sources, affiliations, and academic credentials and professional certifications.

### Publisher

A Publisher object is used to represent the publisher responsible for **Content**. Publishers can be created and assigned to **Content** (for attribution). **Users** and **Workspaces** may assign Publishers an **Assertion Trust Score**. Publishers may be given additional attributes over time, such as funding sources and parent organizations.

### Artifact

Artifacts are an object used to represent a stable container for **Content**, such as a PDF or JPG. Artifacts can be linked together as “near duplicates”, where the contents and identifiers are identical, but the resulting hash of the contents are not as a result of file type, resolution, or other adaptations. It is recommended that a combination of Artifact data and data from linked **Content** objects be used as a basis for defining annotation presentation when viewing the Artifact.

### URL

The URL object is used to represent unstable, potentially dynamic, web-hosted containers for **Content**. The URL object is recommended to be paired with the use of link-rot and content change detection approaches in order to alert **Workspace** members to potential **Content** changes. It is recommended that a combination of URL data and data from linked **Content** objects be used as a basis for defining annotation presentation when viewing the URL.

### Content

The Content object is used to represent units of referenceable information. As such, it might represent an entire book, a book chapter, an area under a subheading, a segment of an image, an entry in a glossary, etc. The Content object can point to other Content objects contained within (e.g., a chapter in a book, or a subheading in a chapter, a figure in a subheading), can point to other variants (e.g., a translated version, a republishing), and be found across multiple **Artifacts**. Content is expected to be assigned an **Author**, **Publisher**, and *Date of Release*, and can contain **Claim Instances**, **Question Instances**, and **References**.

## Reference

The Reference object is a decorator for the following objects: **Direct References** and **Implied References**.

### *Direct Reference*

The Direct Reference object is used to mark labeled, explicit references within **Content** to external **Content**. A Direct Reference must be labeled with a "Type", such as "*in-text reference*", "*footnote*", "*endnote*", or "*in-text citation*", indicating the style through which it presents the reference.

### *Implied Reference*

The Implied Reference object is used to mark what the **User** believes to be an implied reference within **Content** to external **Content**. An Implied Reference must be labeled with the contributing **User's** measure of *Certainty* [0-1] about the implication (i.e., "how likely is it that the **Author** was referencing the external **Content**?").

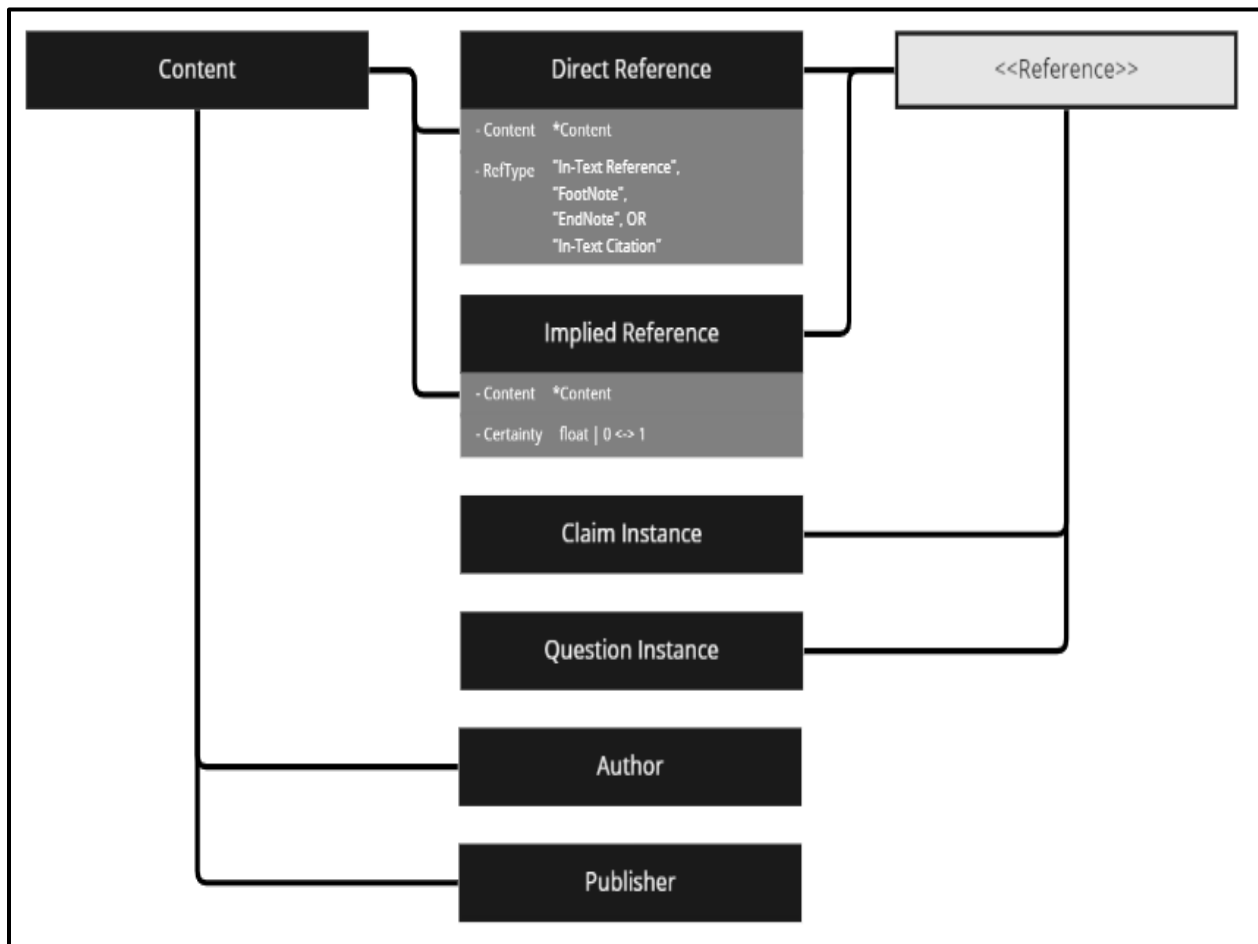


Figure 3. Reference relationships

## Structuring the Information Commons

XI – TrustFinder

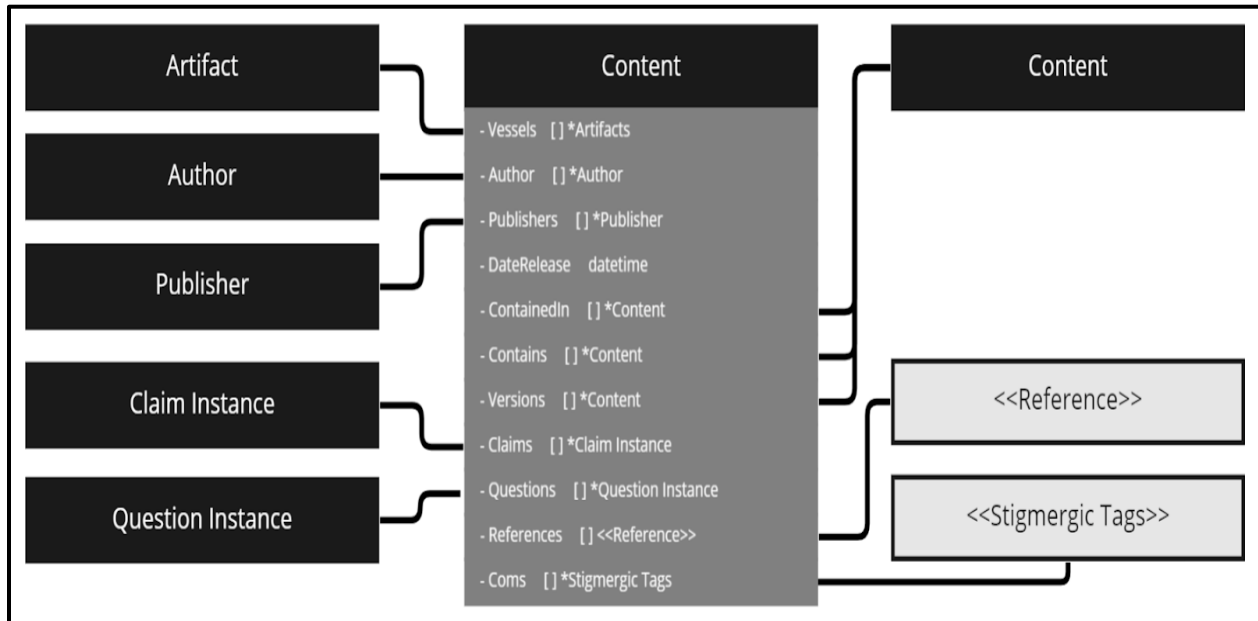


Figure 4. Content relationships

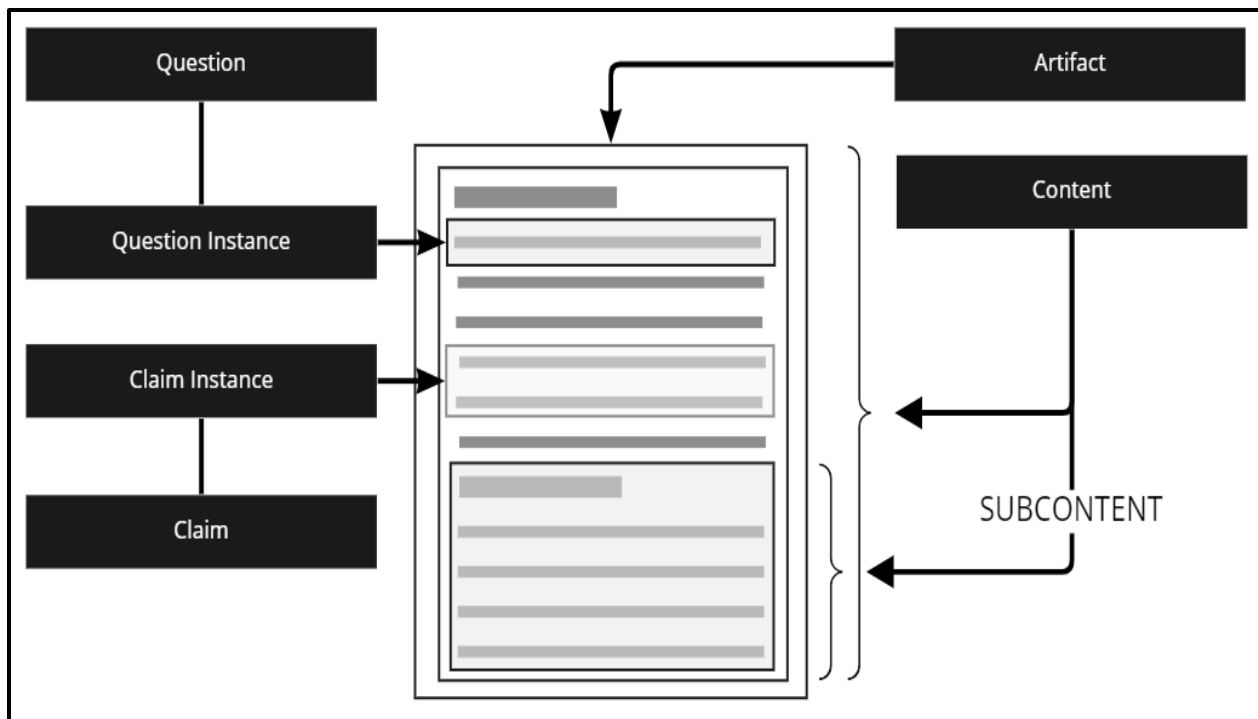


Figure 5. Graphical representation of subcontent and annotation within content

## Claims

### Claim

A Claim is an object which contains a phrase and variants on that phrase which express a “claim” or assertion. Claims also contain a field for *Counterclaims*, or Claims which assert the exact opposite of the subject Claim (e.g., “x is an integer” and “x is not an integer”).

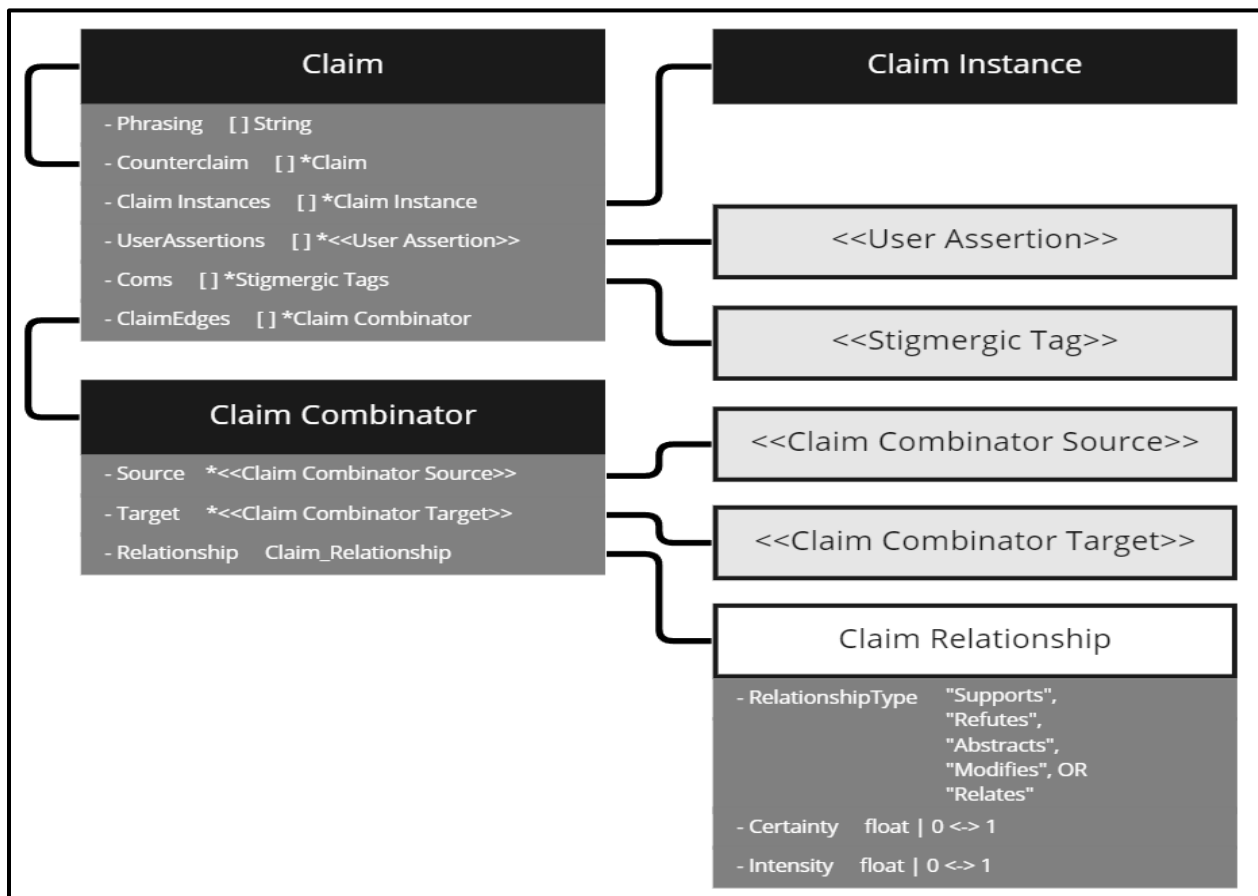


Figure 6. Claim and Claim Combinator relationships

### Claim Combinator

The Claim Combinator object is the basis for forming directional relationships, or edges, between **Claims**, **Claim Clusters**, and other Claim Combinators. Claim Combinators are composed of a **Claim Combinator Source**, **Claim Combinator Target**, and **Claim Relationship**.

#### *Claim Relationship*

A Claim Relationship adds context to a **Claim Combinator**. It is composed of a *Relationship Type*, which describes the relationship between the **Claim**

**Combinator Source** and the **Claim Combinator Target** contained within the **Claim Combinator**; and a 2 dimensional vector containing the contributing **User's** (i) *Intensity* [0-1] rating (e.g., how much does the **Claim Combinator Source** *support* the **Claim Combinator Target**) and (ii) their *Certainty* [0-1] rating (i.e., how certain the **User** is of this relationship between the **Claim Combinator Source** and the **Claim Combinator Target**). There are 5 available *Relationship Types*, and while each is directional - there is an implied bidirectionality (e.g., where Object A *supports* Object B, Object B *is supported by* Object A).

- **Supports | Is Supported By.** Where the **Claim Combinator Source** *supports* the **Claim Combinator Target** (e.g., "x is an integer less than 2" -> *supports* -> "x is equal to 1").
- **Refutes | Is Refuted By.** Where the **Claim Combinator Source** *refutes* the **Claim Combinator Target** (e.g., "x is an integer less than 2"-> *refutes* -> "x is equal to 3").
- **Generalizes | Specifies.** Where the **Claim Combinator Source** *generalizes* the **Claim Combinator Target**, in that it is a generalized version of the same claim (e.g., "x is a symbol" -> *generalizes* -> "x is a mathematical variable").
- **Modifies | Is Modified By.** Where the **Claim Combinator Source** *modifies* the **Claim Combinator Target**, in that it is a modified version of a similar claim, in that it has added conditions, refinement, or mutations (e.g., "x is a mathematical variable in the context of *this* equation" -> *modifies* -> "x is variable").
- **Relates To | Relates To.** Where the **Claim Combinator Source** *relates to* the **Claim Combinator Target**, in that it is similar, communicates something about the other, or shares a context (e.g., "x is an integer" -> *relates to* -> "y is an integer").

### ***Claim Combinator Source***

A Claim Combinator Source is a decorator for the following objects: **Claims** and **Claim Clusters**.

### ***Claim Combinator Target***

A Claim Combinator Target is a decorator for the following objects: **Claims**, **Claim Clusters**, and **Claim Combinators**.

## Structuring the Information Commons

### XI – TrustFinder

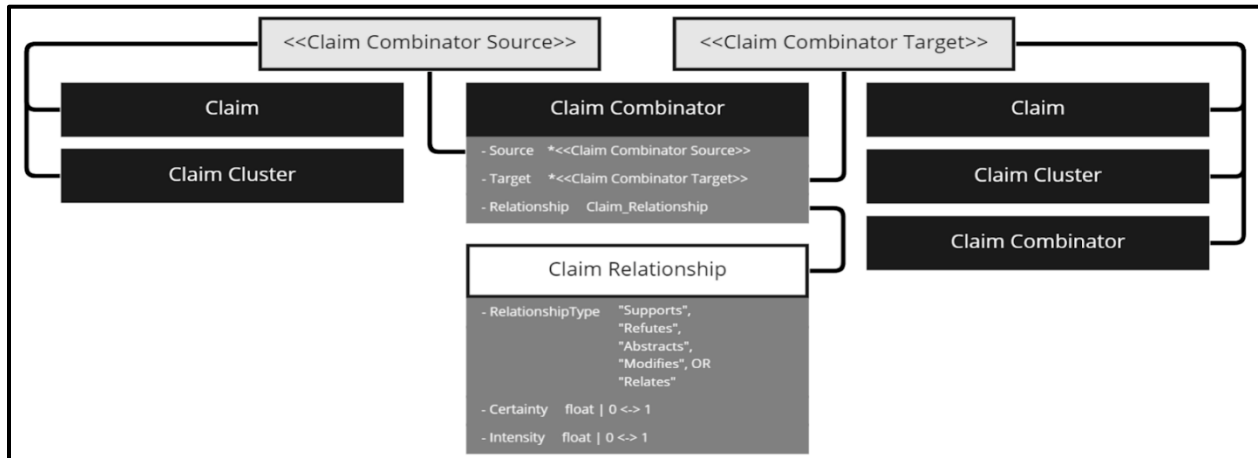


Figure 7. Claim Combinator relationships

## Claim Instance

A Claim Instance is an object representing the annotation of the presence of a **Claim** within a particular piece of **Content**. A Claim Instance must be labeled with the contributing **User's** measure of *Certainty* [0-1] about the Claim Instance (i.e., "how likely is it that this **Claim** is what the **Author** is discussing or asserting?").

- A **User** may mark a Claim Instance as *Asserted True*, *Asserted False*, or *Discussed*, in order to indicate whether the **Author** of the **Content** is asserting the relevant **Claim** is True or False, or simply discussing it, respectively.
- A **User** may mark a Claim Instance as *Explicit* or *Implicit*, in order to indicate that the **Author** of the **Content** is discussing the underlying **Claim** directly, or if the **Claim** is latent or implied in the **Content**.

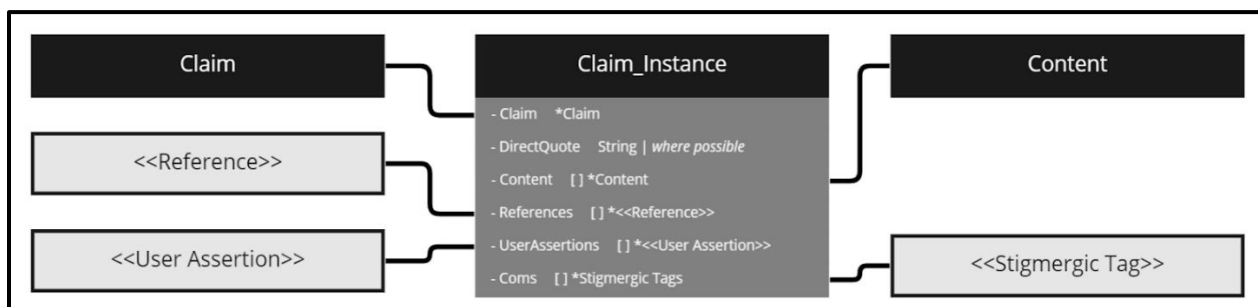


Figure 8. Claim Instance relationships

## Claim Cluster

A Claim Cluster is a container for a set of **Claims** which are grouped together for the purpose of conjecture, context, or collation (e.g., ["x is an integer", "x is a positive number", "x is a number less than 2", "x is a number greater than 0"] -> **supports** -> "x is equal to 1").

## Questions

### Question

A Question is an object which contains a phrase and variants on that phrase which express a “question”, *Prompts* (**Question Combinators** which might inspire or beg the question), and *Responses* (**Question Combinators** which might be answers or responses to the question).

### Question Combinator

A Question Combinator is a decorator for the following objects:  
**Claims, Claim Clusters, and Questions.**

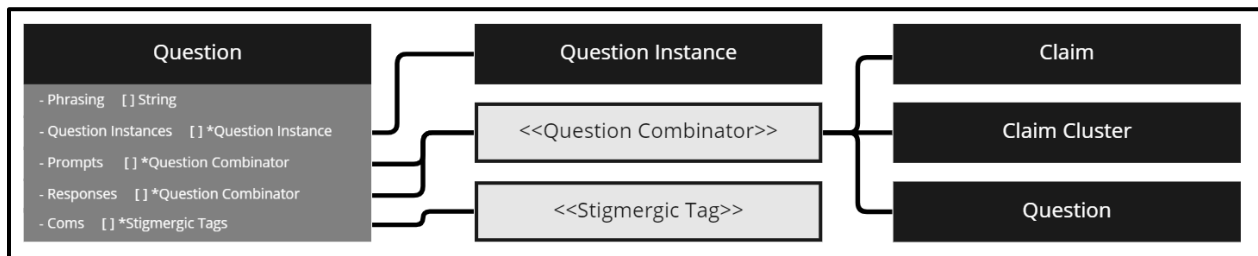


Figure 9. Question relationships

### Question Instance

A Question Instance is an object representing the annotation of the presence of a **Question** within a particular piece of **Content**. A Question Instance must be labeled with the contributing **User's** measure of *Certainty* [0-1] about the Question Instance (i.e., “how likely is it that this **Question** is what the **Author** is discussing or asking?”).

- A **User** may mark a Question Instance as *Explicit* or *Implicit*, in order to indicate whether the **Author** of the **Content** is discussing the underlying **Question** directly, or if the **Question** is latent or implied in the **Content**.



## User Communications

### User Assertion

A User Assertion is an interface representing a **User's** personal assertion about the truth or falsity of a particular **Claim** or **Claim Instance** in the form of an **Endorsement** or **Warning** object. Unlike other annotation affordances, which may be contributed to **Workspaces** by a **User's** chosen pseudonym, it must be attached to the **User's** account. A User Assertion may be attached to either a **Claim** or a **Claim Instance**, creating an option to offer either a Global or Local assertion - as a User Assertion attached to a **Claim Instance** will only be available when interacting with that particular instance of the claim in some **Content**, whereas a User Assertion attached to a **Claim** would be available both during interactions with that **Claim** object, but also during interactions with any of its instantiations (i.e., **Claim Instances**). A **User Assertion** must be accompanied by a plain text explanation, and a 2-dimensional vector containing the contributing **User's** (i) *Intensity* [0-1] rating (e.g., "How untrue or true is this claim?", and (ii) *Certainty* [0-1] rating (e.g., i.e., how certain the User is of this evaluation). It may also be accompanied by **User Assertion Support** objects, such as additional **Claims**. It is recommended that **Users** be required to engage with identity assurance tests (e.g., Genuine Presence Testing) in order to post User Assertions.

#### *User Assertion Target*

A User Assertion Target is a decorator for the following objects: **Claims** and **Claim Instances**.

#### *User Assertion Support*

A User Assertion Support is a decorator for the following objects: **Claims**, **Claim Clusters**, and **References**.

#### *Warning*

**User Assertions** which are intended to warn others of the contents of a **Claim** or **Claim Instance** (e.g., "this claim may be false", "this claim is certainly false and is likely made in bad faith")

#### *Endorsement*

**User Assertions** which are intended to endorse the contents of a **Claim** or **Claim Instance** (e.g., "this claim may be true", "this claim is certainly true, and could only be refuted in bad faith").

## Stigmergic Tag

A Stigmergic Tag is a decorator for the following objects and interfaces: **Requests**, **Rallies**, **Remarks**, **Entity Tag Instances**, and **Custom Tag Instances**, each of which is intended to structure **User** communications at scale and can be attached to nearly any other TrustFinder object (exception being **Users** and **Workspaces**) including other Stigmergic Tags.

### *Request*

A Request is an interface for **Stigmergic Tags** which ask or “ping” other **Users** within a **Workspace** to engage in a specific action. Requests can be suggested to be resolved by those who respond, and may be resolved by **Workspace** administrators or the original contributor of the Request. A Request must be accompanied by the contributing **User’s Intensity** [0-1] rating (i.e., “how urgent or important is it that this request be responded to?”).

- **Skeptical.** A request for clarification about an object or topic from a position of skepticism (i.e., uncertainty with an interest in evaluation).
- **Curious.** A request for clarification about an object or topic from a position of curiosity (i.e., uncertainty with an interest in exploration).
- **Search.** A request for more information about an object or topic which may be already known or more easily searchable by other members of the **Workspace** (e.g., “are there other papers on this specific phenomena mentioned here?”).
- **Catalog.** A request specifically intended to prompt the annotation or cataloging of information found by someone more capable (e.g., “please annotate this potential **Claim Instance**”).
- **Custom Request.** **Workspaces** can implement local, specific **Request** tags to meet their own needs.

### *Rally*

A Rally is a special **Stigmergic Tag** which adds to the *Intensity* rating of other **Stigmergic Tags** in order to help direct attention within a **Workspace** and reduce the likelihood of duplicates or simply directs attention to a particular object. A Rally must be accompanied by the contributing **User’s Intensity** [0-1] rating (i.e., “how urgent or important is it that others see this?”).

### *Remark*

A Remark is a **Stigmergic Tag** which is used to add plain text for miscellaneous comments. A Remark must be accompanied by the

contributing **User's Intensity** [0-1] rating (i.e., "how urgent or important is it that others see this?").

### Entity Tag

An Entity Tag is a tag which indicates the presence of a reference (not to be confused with **References**) in **Content** to a specific entity, such as a concept, idea, person, place, or thing.

- **Entity Tag Type.** An Entity Tag Type is a container for the schema and details of a Custom Tag (e.g., attributes and respective values of the particular Entity Tag, related entities, parent and child Entity Tags, and aliases).

### Custom Tag

A Custom Tag is an interface for **Stigmergic Tags** named and implemented by **Workspaces** for local use. It acts as a compensating control for where no other tag structure is adequate for what the **Workspace** needs to represent or mark.

- **Custom Tag Type.** Custom Tag Type is a container for the schema and details of a Custom Tag.

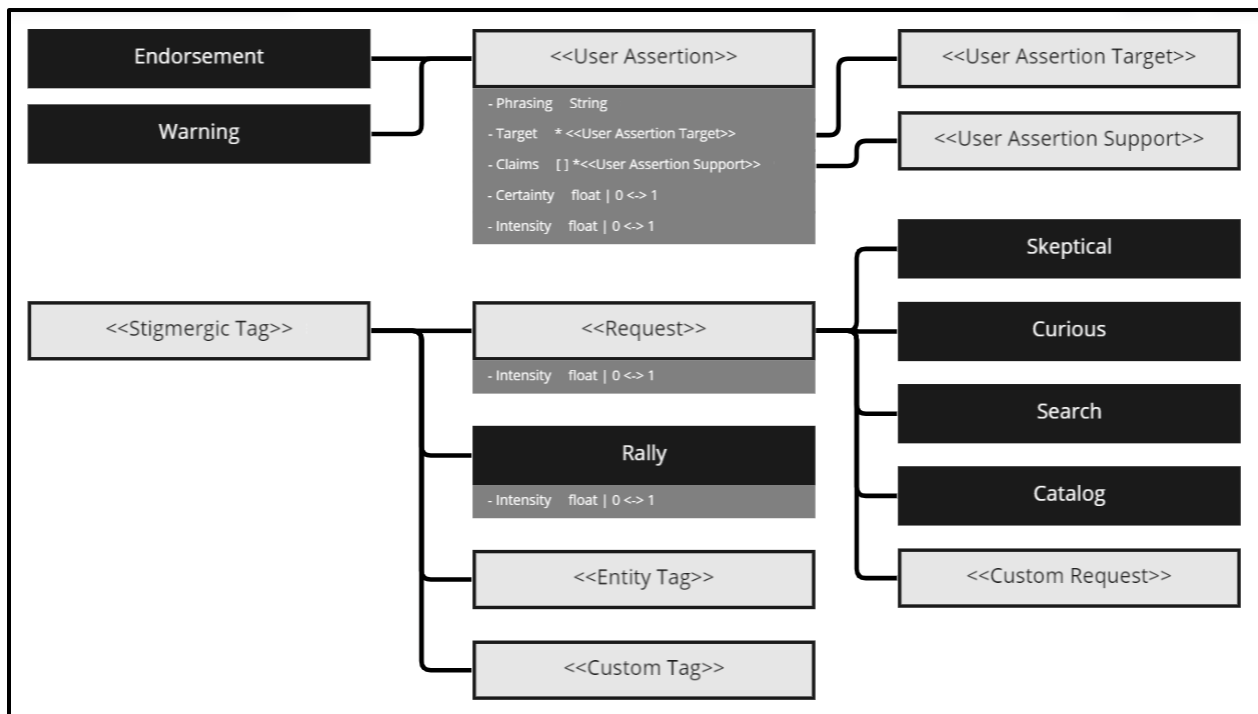


Figure 11. User Communications relationships

## Reputation

There are five separate base types of Reputation in the TrustFinder system: **CQ Annotation Score**, **Mapping Impact Score**, **Network Impact Score**, **Contribution Trust Score**, and **Assertion Trust Score**. Each is a relatively simple metric representing a signal of trust based on past interactions which can be used within the TrustFinder environment or by third parties in order to generate other, optional forms of reputation calculation metrics to **Users**. Nearly all are defined exclusively through set construction and calculation of cardinality, with the only exception being **Network Impact Score**, which uses set construction in combination with a standard decay function.

### CQ Annotation Score

The Contribution Quality (CQ) Annotation Score is a simple metric intended to represent the volume of a particular **User's** direct contributions to identifying claims and questions found in **Content** (i.e., **Claim Instances**, **Question Instances**), within the context of a **Workspace** or a collection of **Workspaces**. Relevant objects include (i) **Claim Instances** and **Question Instances** where the **User** was the initial contributor (e.g., the discoverer of a given claim), and (ii) the **Claim Instances** and **Question Instances** which were contributed within or imported to a given **Workspace** or collection of **Workspaces**. The **CQ Annotation Score** (CQAS) is defined as the cardinality of the set of **Claim Instances** and **Question Instances** formed from the intersection of the set of **Claim Instances** and **Question Instances** by a given **User** (C), with the union of the sets of **Claim Instances** and **Question Instances** associated with a given collection of **Workspaces** (W).

$$CQAS(C \{ \}, \{W_i \{ \}, \dots\}) = |C \cap (W_i \cup W_{i+1} \cup W_{i+\dots})|$$

- Every **Workspace** has a dynamically calculated **CQ Annotation Score** for each **User** which has contributed relevant objects within, either as members or as a result of imports from other **Workspaces**. This collection of scores includes scores for **Users** who are not members of that **Workspace** but have contributions present as a result of imports.
- **Users** can manually define a collection of **Workspaces** in order to calculate a respective **CQ Annotation Score**.
- **Users** can export the underlying data used to calculate the **CQ Annotation Score** (i.e., the set of all relevant **Claim Instances** and **Question Instances**) for use in third-party curation or scoring services.

- While intended for representing an individual **User's** contributions, it can also be calculated using a collection of **Users** (where score would be a sum of the individual scores of the listed **Users**), or using a given **Workspace** (where score would be the count of claims which **Users** contributed as a member of the given **Workspace**).

### Mapping Impact Score

The Mapping Impact Score is a metric intended to reflect the extent of a particular **User's** impacts on the network beyond their own contributions to identifying **Claim Instances** and **Question Instances**, such as their contributions to linking objects within the TrustFinder environment (e.g., adding **References** or **Claim Combinators**) within the context of a **Workspace** or a collection of **Workspaces**. Relevant objects include (i) **Claim Combinators**, **Question Combinators**, and **References** contributed by the **User** where they were the initial contributor, and (ii) the **Claim Combinators**, **Question Combinators**, and **References** which were contributed within or imported to a given **Workspace** or collection of **Workspaces**. The Mapping Impact Score (MIS) is defined as the cardinality of the set of **Claim Combinators**, **Question Combinators**, and **References** (referred to here as *edge objects*) formed from the intersection of the set of *edge objects* contributed by a given **User** (C), with the union of the sets of *edge objects* associated with a given collection of **Workspaces** (W).

$$\text{MIS}(C \{ \}, \{W_i \{ \}, \dots\}) = |C \cap (W_i \cup W_{i+1} \cup W_{i+\dots})|$$

- Every **Workspace** has a dynamically calculated Mapping Impact Score for each **User** which has contributed relevant objects within, either as members or as a result of imports from other **Workspaces**. This collection of scores includes scores for **Users** who are not members of that **Workspace**.
- **Users** can manually define a collection of **Workspaces** in order to calculate a respective Mapping Impact Score.
- **Users** can export the underlying data used to calculate the Mapping Impact Score (i.e., the set of all relevant **Claim Combinators**, **Question Combinators**, and **References**) for use in third-party curation or scoring services.
- While intended for representing an individual **User's** contributions, it can also be calculated using a collection of **Users**

(where score would be a sum of the individual scores of the listed **Users**), or using a given **Workspace** (where score would be the count of claims which **Users** contributed as a member of the given **Workspace**).

### Network Impact Score

The Network Impact Score is a metric intended to reflect the impact of a particular **User's** impact via the invitation of other contributors into the TrustFinder environment within the context of a **Workspace** or a collection of **Workspaces**. Relevant objects include (i) **Claim Combinators**, **Question Combinators**, **References**, **Claim Instances**, and **Question Instances** contributed by **Users**, where they were the first contributor, and where they are members of the subject **User's** (i.e., the subject of the score) *invitation tree* (e.g., where the **User** was invited by an invitee of an invitee of the subject **User**), up to a distance of 6 degrees; and (ii) **Claim Combinators**, **Question Combinators**, **References**, **Claim Instances**, and **Question Instances** which were contributed within or imported to a given **Workspace** or collection of **Workspaces**. The Network Impact Score (NIS) takes as inputs a set of **Workspaces** ( $W$ ) of length  $M$ , with each element representing a set of **Claim Combinators**, **Question Combinators**, **References**, **Claim Instances**, and **Question Instances** (referred to here as *contributions*) and a set of 2-dimensional vectors ( $u$ ) of  $n$  length with each element representing a **User** within the subject **User's invitation tree**, each vector contains (i) a set of **Claim Combinators**, **Question Combinators**, **References**, **Claim Instances**, and **Question Instances** that the element's respective **User** contributed ( $u_{i,c}$ ) and (ii) the degree of separation of the element's respective **User** in the subject **User's invitation tree** ( $u_{i,d}$ ). The Network Impact Score (NIS) is defined by (i) finding the cardinality of the intersection of the set of contributions by each given **User** in the subject **User's invitation tree** ( $u_{i,c}$ ), with the union of contributions within the given collection of **Workspaces** ( $W$ ); (ii) weighting the resulting cardinality by a parameterized decay function which takes the given **User's** degree of separation ( $u_{i,d}$ ) as an input; and (iii) summing the results for each **User**.

$$\text{NIS}(u \{[u_{i_c}, u_{i_d}], \dots\}, W \{W_a \{ \}, \dots\}) = \sum_{i=1}^n |u_{i_c} \cap (W_a \cup W_{a+1} \cup W_{a+\dots})| \frac{e^{-\frac{1}{u_{i_d}-1}}}{2}$$

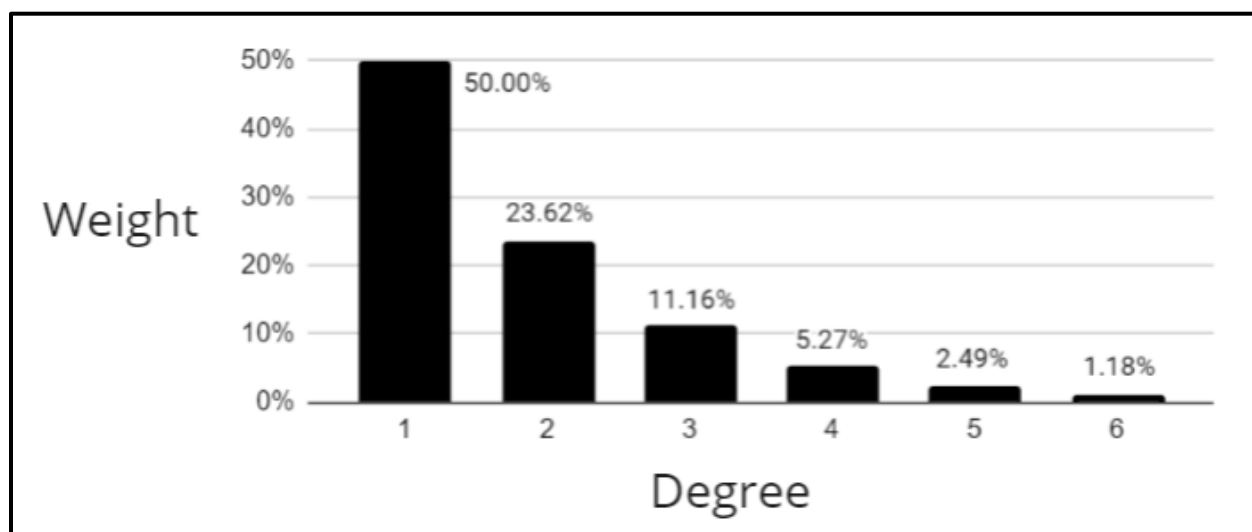


Figure 12. Degree of Separation impact on weight in Network Impact Score

### Contribution Trust Score

The Contribution Trust Score is a metric intended, generally, to represent a particular **User's** or a **Workspace's** relative level of trust in another given **User's** annotation contributions (e.g., **Claim Instances**, **References**) or another **Workspace's** imported contributions. It reflects a **User's** or **Workspace's** expectations about the quality of annotation contributions by another **User** or **Workspace** (e.g., will this annotation contain errors?, will this individual use affordances as expected?). A Contribution Trust Score is set manually by a **User** for themselves or for their **Workspace**, and can be adjusted manually at any time. It is set on a scale between -1 and 1; where a rating of -1 is intended to represent a **User's** belief that the target of the rating would, without exception, purposefully or negligently contribute flawed annotations; and where a rating of 1 is intended to represent a **User's** belief that the target of the rating would, without exception, contribute properly formatted annotations, free of errors. This rating can be used to create filters on imports of annotation contributions within a **Workspace**, and will create visible indicators on presentation of annotations. **Users** and **Workspaces** must set a default rating to apply to unrated **Users** and **Workspaces**, upon registering or instantiation, respectively.

Given the subjective nature of the contents of annotations and the nature of expertise, **Users** can set conditional Contribution Trust Scores, which use a logical statements containing "and/or" combinations of **Entity Tags** and annotation types (e.g., IF (TAG<sub>x</sub> AND TAG<sub>y</sub>) OR Type <sub>Reference</sub>) combined with a replacement rating. Where the logical statement holds true given the set of **Entity Tags** associated with a given annotation contributed by the target of the rating, the standard rating will be replaced by the defined replacement rating. This allows the marking of contextual trust, where, for example, a physicist's attempts to annotate **Claim Instances** within

the domain of physics may be trusted at a higher level than their annotations related to psychology.

- Where there is more than one rating associated with the object, such as when there is both a personal rating and workspace rating, or where multiple conditional ratings triggered, the respective indicator related to the rating should be combined with others into new visualizations.
- **Users** should be encouraged to make conditional trust the norm via user experience mechanisms (e.g., by making conditional trust easy to assign via presented annotations with suggestions related to **Entity Tags** which are already present in the annotation).
- **Users** can export their Contribution Trust Scores for use in third-party curation or scoring services.

### Assertion Trust Score

The Assertion Trust Score is intended to represent a particular **User's** or **Workspace's** relative level of trust in a **User's**, **Author's**, or **Publisher's** assertions. It reflects a **User's** expectations about the quality of **User Assertions** by a particular **User** or the quality of the contents of **Claim Instances** which are marked as *asserted* by the **Author** or **Publisher** of the **Content** in which they are found (e.g., does this person have a good grasp of the subject matter they are making assertions about? Is this person acting in good faith or are they being opportunistic?). An Assertion Trust Score is set manually by a **User** for themselves, or by a **User** for a **Workspace** and can be adjusted manually at any time. It is set on a scale between -1 and 1; where a rating of -1 is intended to represent a **User's** belief that the target of the rating would, without exception, purposefully or negligently assert false statements; and where a rating of 1 is intended to represent a **User's** belief that the target of the rating would, without exception, contribute objective and truthful statements. This rating can be used to create filters on imports of annotation contributions within a **Workspace**, and will create visible indicators on presentation of **User Assertions** and **Claim Instances**. **Users** and **Workspaces** must set a default rating to apply to unrated **Users**, **Authors**, and **Publishers**, upon registering or instantiation, respectively.

Given the subjective nature of assertions and the nature of expertise, **Users** can set conditional Assertion Trust Scores, which use logical statements containing "and/or" combinations of **Entity Tags** and annotation types (e.g., IF (TAG<sub>x</sub> AND TAG<sub>y</sub>) OR Type Reference) combined with a replacement rating. Where the logical statement holds true given the set of **Entity Tags** and annotation types associated with a given assertion,



the standard rating will be replaced by the defined replacement rating. This allows the marking of contextual trust, where, for example, a physicist's assertions within the domain of physics may be trusted at a higher level than their assertions related to psychology.

- Where there is more than one rating associated with the object, such as when there is both a personal rating and workspace rating, or where multiple conditional ratings triggered, the respective indicator related to the rating should be combined with others into new visualizations and indicators.
- **Users** should be encouraged to make conditional trust the norm via user experience mechanisms (e.g., by making conditional trust easy to assign via presented annotations with suggestions related to **Entity Tags** which are already present in the annotation).
- **Users** can export their Assertion Trust Scores for use in third-party curation or scoring services.

# Implications

The potential implications of affordances, social systems engineering mechanisms, and other aspects of the system are discussed below.

## Local Governance

The solution space for managing governance, role and process, and mediation of conflict in human interactions online are extremely diverse, and best practices are highly dependent on local conditions. Any platform-level requirements and decisions reflected in complex or complicated definitions and rules for how users mediate conflicts, offer recourse, and manage roles and processes also create platform-wide threat surfaces with the potential for goal-blocking, inefficiency, and intrusions on community and user sovereignty. The recommended TrustFinder environment embraces this paradox as being reflective of reality, and makes the causative relationships explicit, opening up access to benefits from a more distributed and scalable approach wherein inter-community conflicts are managed via the formal structure of annotations that reveal the directional, and conditional, relationships between and among workspaces while intra-community conflicts remain in the purview of community self-governance. In this way, inter-community conflicts are effectively converted into community-oriented information differentials, the collective management of which yields value for all potential users. Specific platform-level governance affordances are recommended to be added only upon request by affected communities, and not required for use by all users across all communities. Given that workspaces can be arranged in complex import and export relationships by applying simple rules, many different, locally-adapted governance affordances may be facilitated without the need for specific standardized features (e.g., role-based access).

## Empowering Communities and Users to Define and Assign Trust

Similar to the domain of governance, the solution space for managing reputation is extremely diverse, and user experience and quality control outcomes are subjective and highly dependent on local conditions. Any platform-level choice in complex or complicated definitions and rules for how user reputation is scored and impacted from behaviors of a user (or by the choices of other users) creates threat surfaces for misuse and counterproductive intrusions on individual and community level processes for deciding reputation. Further, it is not possible to create a curation or filter decision function that is free of bias, as curation and filtering is, by definition a discriminatory function. As such, any platform-level rules choice in defining curation and decision function for users will run a high likelihood of impacting users' trust in the system itself. As before, the discernment of this paradox reveals a system performance reality that is amenable to productive and value-creating management

through community-based governance affordances, but in this latter case, directed toward individual reputation variables rather than the resource-focused attention of governance.

As discussed elsewhere, the goal of TrustFinder is to structure the information environment in order to enable users to find “trusted” sources of information. “Trust” is an emergent subjective internal state of a system (including “users” as a system), that is ultimately informed by elements that are external to the system. People and organizations that are empowered to discern (and measure) the degree to which performance of elements of a given system (or system component) are reliable and predictable may more confidently rely on the future performance of said system and come to “trust” said system in a mechanistic way. Users that have the capacity to identify and cultivate system elements that are relevant to their specific circumstances and upon which they can base such mechanistic “trust” have an advantage (in terms of cost and resource efficiencies) in leveraging and de-risking future interactions that is not available to others without such capacity. To this end, the users of TrustFinder specifically are empowered to define for themselves when and how to assign easily understood measures of trust (e.g., assertion trust scores, contribution trust scores), associated with other users, workspaces, and the authors and publishers of content, and to further specify in what contexts they apply and adjust those measures. It is recommended that TrustFinder take a facilitatory role in how researchers across disciplines adjust and access the values of these signals, as opposed to an authoritative role - and that it should be anticipated that its users will exercise agency to pursue their self-interest by self-binding to rules that offer reliability and integrity across a well-structured, navigable information system.

### **Rate Limiting Mechanisms on Spread of Trust**

In complex information environments, trust may be counterproductively assigned using extrinsic signals such as affiliation and identity (or other surrogates for or abstractions of reliability and affinity) as opposed to intrinsic signals of quality and reasoning. While such assignment is understandable from the standpoint of interaction efficiency, when such a trust assignment is signaled publicly, the assignment will inevitably be affected by tribal dynamics and personal relationships and other agenda and contexts relevant to the users involved in later communications referencing such earlier trust assignments. In other words, the contextual foundations of the original abstraction of trust (e.g., to identity) is lost from the original communication, subjecting the naked communicated signal (data) to being interpreted by a later party in a different context (meaning) either through ignorance or malice, yielding so-called “mis” information and “dis” information respectively. TrustFinder makes it possible for researchers to manage communications to eliminate such “context stripping” of communications, by allowing them to manage trust signals privately.

The recommended TrustFinder environment benefits from an approach which stresses production of actor- and community-centric metrics (i.e., proximal, dynamic calculation from the perspective of a given user or workspace) which can be incorporated into more complex derivative down-stream curation and reputation analysis features by third-parties. The provision of services offering such down-stream insights have the potential to power new inter-disciplinary and trans-disciplinary insights in the academic sphere and new innovations in products, services, and markets in commercial contexts. The use of proximal calculation and presentation is applied as an alternative to a universal (i.e., platform-wide) or static reputation metric. This approach intends to limit the negative effects of context-stripped trust signals “going viral,” and to protect user and community ratings from being unduly affected by external pressures.

### **Scoping through Collaborative Work**

Scoping the information environment through the use of mission-focused workspaces intended to facilitate collaborative work may affect the environment in a number of ways:

#### **Subjectivity of Evaluation**

Human knowledge is incredibly complex. In many cases (and contrary to what is often assumed) claims may only be “true” within certain contexts. For example, “home is where people will miss you when you are gone”, in some contexts, is a “true” statement, or a statement which “rings” true, or, at the very least, a statement which may be not helpfully marked as definitively false. It may not be the technical definition of a “home” from a given personal or cultural perspective, however it may be “literally” true in some cultural contexts, or “metaphorically” true within the context of a narrative analysis. This simple statement reveals the context dependency of the concept of “truth”.

By scoping the environment around collaborative work within a defined workspace, users can collaboratively refine their community’s information environment with the necessary context for user assertions, claims, and their relationships. Within the community workspace environment as contemplated here, members of the community do not need to ask for the permission or forgiveness of any outside party to apply a given set of context. They might be said to have “context/meaning sovereignty” within that information environment. Further, they can annotate and make assertions about claims applying their context-consistent elements

without the need to fight for platform-wide consensus in order to enjoy the information environments that support them and enable them to perform work. While some may feel there is risk involved in allowing communities to define “their own truth”, a well designed system will be structured to make explicit the distinction of a contextual, community-bound “truth,” from a broader form of “truth” that is recognized across multiple contexts and multiple communities, which allows for the cultivation and management of dissenting views and innovation. In any event, fact-checking, censoring, or overriding the expressions of a given community that embraces a context bound, minority-position on a given “truth,” may be counterproductive. Generally, these kinds of interventions are only effective in terms of limiting effects of network exposure to undesired information or interpretation - but in the case of TrustFinder, said effects are already curtailed by the structure of workspaces.

### **Reduction of Information Overload**

Any given text has the potential to include an overwhelming number of entities, claims, questions, and other annotations associated with it. The use of questions, claims, clusters of claims, and relationships between claims as a basis to scope workspaces improves the likelihood that the user will find annotations relevant to the task at hand.

### **Power Dynamics**

Unbounded information collection activity results in cumulative build up of influence by committed contributors, and opportunities for “tyranny of the minority” phenomena, wherein small cliques get outsized control over what information in an environment is considered worthy of attention. With crowd-consensus mechanisms in place, the potential for tyranny of the minority is replaced by the potential for tyranny of the majority, where the interests of the majority truncate the interests of minority groups. The use of provisional and reconfigurable workspaces that can be selectively combined, abandoned, published, and republished by small teams allows for a freedom and flexibility that keeps both powerful cliques and homogenous crowds in check.

## Neutral Discovery of Claims and Questions

Separating affordances for the discovery of claims from those that convey the opinions of users reduces the likelihood of tribal and affiliation-related dynamics and creates opportunities for common ground between groups with disparate interests and perspectives on the world. For example, two communities which vehemently disagree on the truth of a claim, can find common ground in the notion that “this article has an instance of this claim”; and even in cases of extreme disagreement, can at least agree on the title and citation metadata. This separation of concerns between different levels of analysis and complexity allows communities to benefit from each other's work despite their disagreements.

## Modular and Flexible Construction of Claims Ecosystem

Traditionally, claims annotation is done on a document-by-document basis with a specific focus on the contribution of individual claims toward the argument a document is intended to advance. Allowing researchers to annotate the claims that are of value to their particular work simultaneously preserves quality of user-experience (i.e., not creating additional work for them unrelated to their current goals) and, as an incidental benefit of their self-interested annotation activities, also provides a modular, granular contribution to larger crowdsourcing solutions. As claims and references are linked to one another and are aggregated with the claims and references from other workspaces, small, individual contributions are brought together to create a rich, linked network of claims that no individual could have created alone. This is an example of familiar “network effects” of generating value, but here applied to meaning making across communities. Such emergent “meta-information” layers bear a relationship to baseline information similar to the relationship that meta-data has to baseline data, but in the case of such emergent, intercommunity context and meaning, situational awareness is extended to include formerly external components of context and meaning. Further, these relationships between claims can be represented as the key components of nearly any model of representation of argument and can be applied to any form of content (e.g., video, image, gif, text), which allows for advanced multimodal rhetorical analysis and reusability of claims information as training data in argument mining and artificial intelligence systems.

## Claims as Networked Real-Estate: Gold Rush

Being the first to mark a claim provides both a first mover advantage on setting the tone and character for description and documents participation in its discovery. The reputational gains of being first, or more importantly, being first to provide a helpfully objective interpretation of a found claim, creates the opportunity for a “gold-rush” mechanism to drive adoption and participation. Further, given that reputation metrics are impacted by both the discovery and the annotation of claims, users are incentivized to perform high-quality claims discovery and annotation

where it is most critical and valuable in both past and recent literature (e.g., finding and being associated with the discovery of claims which are at the root of a field are equally valuable to finding those which might be at the root of new fields or paradigm shifts). While such a mechanism can represent a risk to the intrinsic quality of annotation and encourage counterproductive rivalrous dynamics, there are several aspects of TrustFinder which are expected to keep these phenomena in check:

### **Consumers of Found Claims are Incentivized to Merge**

The choice to merge two duplicate claims or to choose one annotated claim over another is now within the hands of those managing that workspace, and users are highly incentivized to detect and merge duplicate claims in the interest of reference stability. The incentive for rivalrous dynamics may increase with the value of the claim, but so do the incentives for maintenance of reference integrity.

### **Competition**

Even where a user may intend to bury a rival's discovery in the interest of preserving their own status as the initial discoverer of a claim, and where they have control over a commonly referenced workspace, they do not have the affordances to maintain a control over the many other workspaces which may independently pull their rival's claims back in and merge them.

### **Game Theory of Return on Work**

Given that reputation return for contributions is tied to the breadth of use and reference of the claim, in most cases, it will likely be a more reliable strategy to simply merge claims in order to increase likelihood of spread, even if it means a slight decrease in the perceived share of the reputation impact on use of a claim. The system rewards synthesis as much as it rewards discovery.

### **Use of Security Assurances**

The affordances for annotation of personal opinions regarding claims found within content present threat surfaces for interpersonal aggression and intergroup tribal dynamics, and an opportunity for threat actors to use these vectors for purposes unrelated and contrary to the goals of the relevant community of users. As such, the TrustFinder environment requires users to engage with cyberphysical security measures in order to register an account in the system and to commit their assertions to the environment. This has several implications:

### **Cost of Engagement**

The requirement to engage with security assurances in order to annotate assertions creates task-disrupting barriers that offer “shocks to consciousness” to the user to ensure they are unambiguously aware of the gravity of their interaction. This awareness is achieved via mechanism as opposed to being provided with disclaimers - users “experience” the weight of their decisions as opposed to simply being told about them and are prompted to consider the risk of their decision given the cost of engagement.

### **Cost of Entry**

The use of security assurances creates a cost of entry to the environment that acts as hostile architecture to threat actors intending to make multiple accounts.

### **Separating Extrinsic from Intrinsic Rewards**

Extrinsic rewards are those that have visibility from the outside (e.g., titles and status), and fungibility across people (e.g., material or currency), whereas intrinsic rewards are those that are inferred or experienced by a cognitive agent, such as personal fulfillment or a sense of purpose within a community. The potential for extrinsic and intrinsic rewards has significantly different impacts on behavior. Tendency to optimize toward extrinsic rewards is natural where they are offered, but this optimization axiomatically comes at the expense of the potential intrinsic value in the solution space. This being the case, creating simplistic extrinsic rewards for writing novels might generate *more* novels, though not necessarily better ones - and attaching “eyeball” or “dwell-time” related metrics, such as *how many people saw and liked my warning/endorsement*, will create perverse incentives for users to contribute what they believe the crowd will vote for, which may be in conflict with what they believe to be true.

The TrustFinder environment supplements its ability to support relatively modular, granular, narrow solution-space tasks (e.g., claims annotation) with extrinsic reputational rewards (i.e., CQ annotation score and mapping impact score, which reflect definable network impacts and use of contributions). Given the reliance on small-team focused workspaces, user assertions and responses to requests can be left to intrinsic reputational rewards - through the impacts users *feel that they make* on their local community.

### **Structure of Claims and User Assertions**

The structure of claims paired with the attachment of user contributions to simple, self-reported levels of certainty and intensity enables the identification and



application of new metrics about information integrity, opportunities for myriad forms of cognitive modeling regarding human engagement with clusters of claims and concepts, and opportunities to create related visualizations and accessible metrics for communicating status about integrity or informational conflict at the level of claim, document, or field (e.g., through the application of system status signals based on such things as color theory and simple summary statistics). Further, the highly structured relationships between claims and the structure of user assertions means that, where conflict arises, users are incentivized to engage in such conflict in a highly structured manner - resulting in hybrid information structures (i.e., composed of competing user assertions) which can be mined for insight regarding the volatility of certain claims. When using neutral claim annotations, as opposed to user assertions, users' interest in engaging in conflict (i.e., ensuring that claims they don't agree with are undermined, and that claims they agree with are supported) is harnessed as a driving force in mapping and connecting the rhetorical landscape as they search for supporting or refuting claims.

In addition, the flexibility of entity and custom tagging affordances in conjunction with open standards for interoperability with third party tools allows for communities to layer more advanced standards onto TrustFinder structures. For example, communities interested in more advanced rhetorical analysis of discourse are empowered to layer classification information onto objects, such as categories of claims (i.e., factual, definitional, causal, value, and policy) and other related data or categories of questions (e.g., interrogative, exploratory).

# Structuring the Information Commons

## XI – TrustFinder

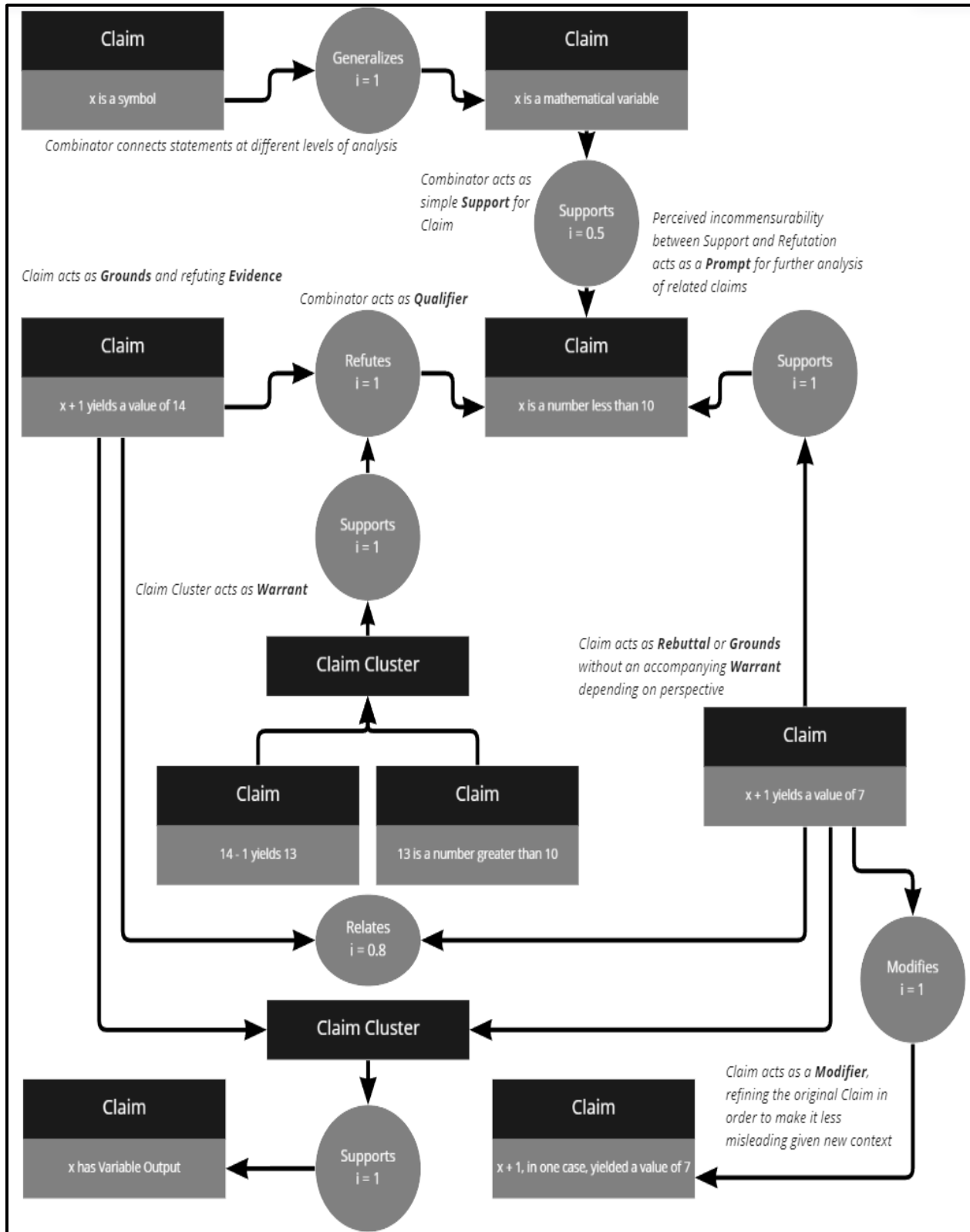


Figure 13. Graphical representation of relationships between claims as a basis for representation of complex arguments, with example intensity ratings ("i") for claim combinators.

## Compatibility with Other Systems

The structure of claims and references allows for import from (and the potential for export to) other systems which deal with claims discovery and reference management, such as Polyplexus, Swarmcheck, Paperpile, Mendeley, or Zotero.

### Polyplexus

Polyplexus is a platform for crowdsourced collection of claims from documents and for hosting of claims-based exploratory research incubators. TrustFinder's claim instance and content objects would be highly compatible with Polyplexus' schemas, offering the potential for users to:

- import Polyplexus claims and driving questions in order to instantiate a workspace,
- export TrustFinder claim instances and reference data for upload to Polyplexus,
- export a TrustFinder workspace's claim instances, reference data, or claim clusters in order to submit claims to a Polyplexus incubator, or
- import claims associated with a Polyplexus incubator in order to instantiate a workspace.

### Swarmcheck

Swarmcheck is a company which provides argument and discourse analysis and engagement tools for public and corporate use. TrustFinder's claim combinators and claim objects would be highly compatible with Swarmcheck's schemas, offering the potential for users to:

- import a Swarmcheck discourse map in order to instantiate a TrustFinder workspace, or
- export a TrustFinder workspace's claims and claim combinators in order to view and map discourse.

### Reference Managers

Paperpile, Mendeley, and Zotero are platforms which provide reference management functions for researchers. TrustFinder's references and content objects would be highly compatible with most reference management schemas, offering the potential for users to:

- import and export reference objects.

## Complex Knowledge Projects

The flexibility in creation of connections between and among workspaces allows for complex projects, constructed by multiple teams with separations of concern in workflow based on relevance of information. It also allows for individual researchers to find value even if they are isolated from all other users in the wider TrustFinder environment. Workspaces can be generated or populated with claims using queries of other workspaces to which they have access, and can have import and export integrations with other compatible systems, allowing for rapid synthesis in interdisciplinary, interorganizational work. Finally, TrustFinder workspaces can be used to help improve collection, accessibility, and dissemination of information resources for digital communities of practice at scale.

## Gradients of Common Ground

Crowdsourcing solutions for information collection and interpretation can be difficult to implement when contributors don't share ontology or common narrative. The recommended TrustFinder environment assumes a wide diversity of viewpoints and implements a separation of concerns among objects to allow for communities which might disagree at one level of analysis to nonetheless cooperate on collection and analysis activities at another level where agreement is present (see Figure, "Gradients of Common Ground"). For example, two communities may have fundamental disagreements regarding the truth of a particular statement (i.e., at the level of user assertions), but can still agree on independent notions and issues such as the ideas and concepts involved and how they support or refute the statement (claim combinators), on where the statement is made (claim instances), and the relevant entities associated with the statement (stigmatic tags). In an extreme example, where two communities cannot even agree on the relevant entities associated with a given statement, they may, at the least, be able to agree on the name of a document or author (i.e., reference information). The use of workspaces with conditional import and export allows communities that would otherwise never interact to manage information sharing agreements that circumvent unnecessary conflict.

## Mapping the Information Supply Chain

As of 2022, mapping the origin of a particular claim is a challenging, time-consuming task, even in literature with well-structured ontology and citation standards. While some reference mapping solutions exist, they are not necessarily accessible or sufficient for most use-cases, often contain errors, miss large swathes of relevant documents, and cannot keep up with the millions of new documents and datasets being generated each year. Further, even the best enterprise tools available rarely move beyond document-to-document links and references; it is only use-case specific tools, such as those found in legal study and practice, that offer affordances for semantic or conceptual provenance (e.g., precedent search). The recommended TrustFinder environment's reference and content objects, in conjunction with entity tags, claim instances, and question instances, allow for a collaborative mapping of implicit and explicit provenance of ideas across deep-time at the level of document

and claims. Further, its flexible content object structure allows for claims of provenance to extend from the higher level of books all the way down to the more granular level of paragraphs, with attribution and reference annotation affordances that enrich and clarify context of citations and references appropriate for all such levels.

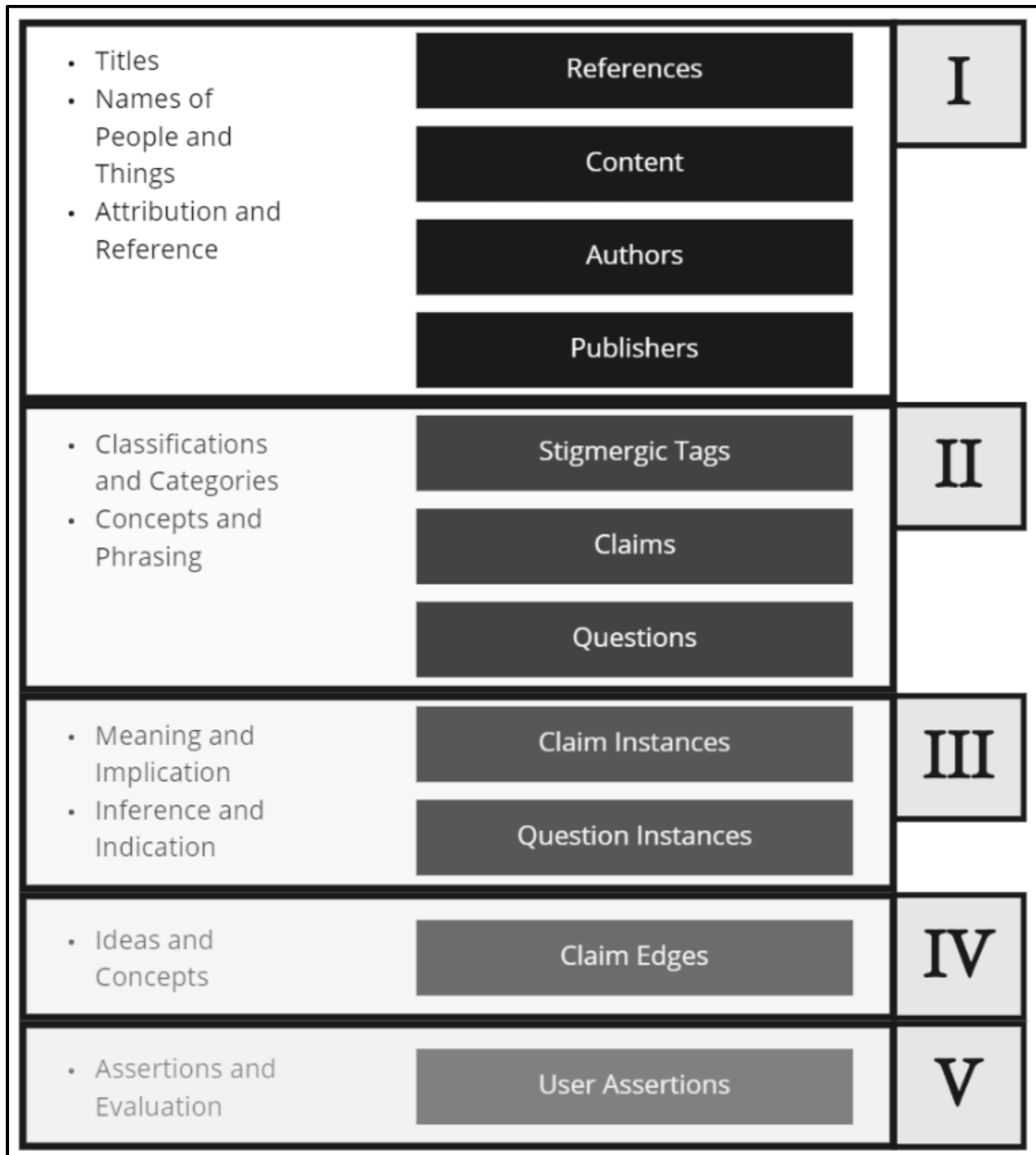


Figure 14. Gradient of Common Ground

## EOS - Entity Oriented Search

The structure of the core TrustFinder objects, such as claims, claim instances, and content, allow for numerous queries that are driven by defined entities as opposed to syntax (i.e., language based search) which can illuminate implicit and latent relationships among claims and agents. For example:

### ***By Content***

A particular piece of defined content can be used as the object of search to yield:

- Claims within and their underlying claims.
- The content's implicit and explicit references.
- Other content which has a similar set of claims or references.
- Content which references the content used in search.

### ***By Author***

A particular author can be used as the object of search to yield:

- Common claims within their work.
- Common references they use.
- The claims they've made that aren't accompanied by their common refutations (e.g., what areas within their work might be biased or assumed).
- Publishers that have published their work.

### ***By Publisher***

A particular publisher can be used as the object of search to yield:

- Common claims within the work they publish.
- Authors they've published.
- How often they publish opposing points of view.

### ***By Claim***

A particular claim can be used as the object of search to yield:

- Content which presents or contains instantiations of that claim.
- Content which has instantiations of that claim primarily accompanied by refutations of that claim (e.g., to find critique articles).
- Content which has instantiations of that claim primarily accompanied by support of that claim (e.g., to find review articles).
- Claims which have certain relationships with the claim used in the search (e.g., supporting, refuting).

### ***By Combinator Relationships***

Combinator Relationships can be used as the object of search to yield:

- Search for claims within workspace that have very few combinator relationships to find potentially underexplored areas of research.
- Search for claims within workspace that have very high consistency in combinator relationships (e.g., claims with equal support and refutation) to find areas that may have been well researched but contentious.
- Exploration of the refinement of claims, by search and review of modification trees (wherein claims are refined through modification over time).
- Exploration of the generalization of claims, by search and review of generalization trees (wherein claims are generalized and specified across fields).

### **Infrastructure for Other Systems**

The compatibility with external systems and the ability to create information “pipelines” between and among workspaces, in addition to enabling complex work, allows users to create ad hoc systems on top of TrustFinder.

### **Traditional and New Forms of Peer-Review**

Journals and other research-publishing organizations could use workspaces to manage aspects of peer review that are concerned with claims and research questions, such as finding

peer reviewers, evaluating the state of claims, and representing the rhetorical structure of the subject document. The ability to create multiple workspaces with conditional imports and exports means the potential for new forms of peer-review processes that are highly auditable and transparent, and allow for a larger number of participants.

### **OSINT SCADA**

Organizations with high information collection and analysis requirements could use layers of interconnected workspaces to generate role-based information management and intelligence pipelines that can be contributed to at-scale and monitored in real-time. Given export and web annotation affordances, a collection of interconnected workspaces could be the basis for a supervisory control and data acquisition system (SCADA) for open source intelligence (OSINT) related purposes.

### **Technical Intelligence, Narrative Wargaming, and Exploratory Exercises**

Users could build collections of interconnected, structured workspaces in order to engage in myriad narrative and technical intelligence related wargaming, collection, and exploratory exercises. For example, using separated blue (support), red (opposition), and green (communication) workspaces connected through intermediary workspaces with umpire-controlled selective disclosure. As another example, workspaces could be connected in order to allow for an adaptation of the “World Game” developed by Buckminster Fuller and others, wherein global resource availability and summary statistics are interactively and iteratively addressed by a collaborative team.



# Background

Here, key frameworks and concepts are provided from works consulted and the works within this volume which guided the recommendations for the TrustFinder environment.

## Argument Mining and Representation

### Toulmin's Framework

The rhetorical framework of Stephen Toulmin has been used to make sense of and formalize argumentation and reasoning within myriad fields, including “science, law, management, art criticism, and ethics”. The Toulmin rhetorical framework formalizes the structure of an argument through the relationships among 6 individual components:

#### **Claim**

The claim is the central *assertion* by an individual proposing an argument.

#### **Grounds**

Sometimes referred to as data, relevant facts, or evidence, the “grounds” of an argument is information that supports the claim.

#### **Warrant**

The warrant explains *why* the grounds support the claim. Warrants are claims themselves (often unstated assumptions) that must be accepted so that the original claim follows logically from the grounds. “Warrants confer different degrees of *force* on the conclusions they justify”, which is communicated through a qualifier. A single argument (claim-grounds pairing) could be supported by multiple warrants.

#### **Qualifier**

The qualifier expresses the *relative strength* of the claim. It is often expressed rhetorically, through the phrases such as “might be”, “probably”, “certainly”, or “axiomatically”.

### Backing

The “backing” component of an argument explains why the warrant has authority. The backing supports the warrant in the same way that the grounds support the claim.

### Rebuttal

The “rebuttal” or counter-claim is a claim which refutes the claim or warrant.

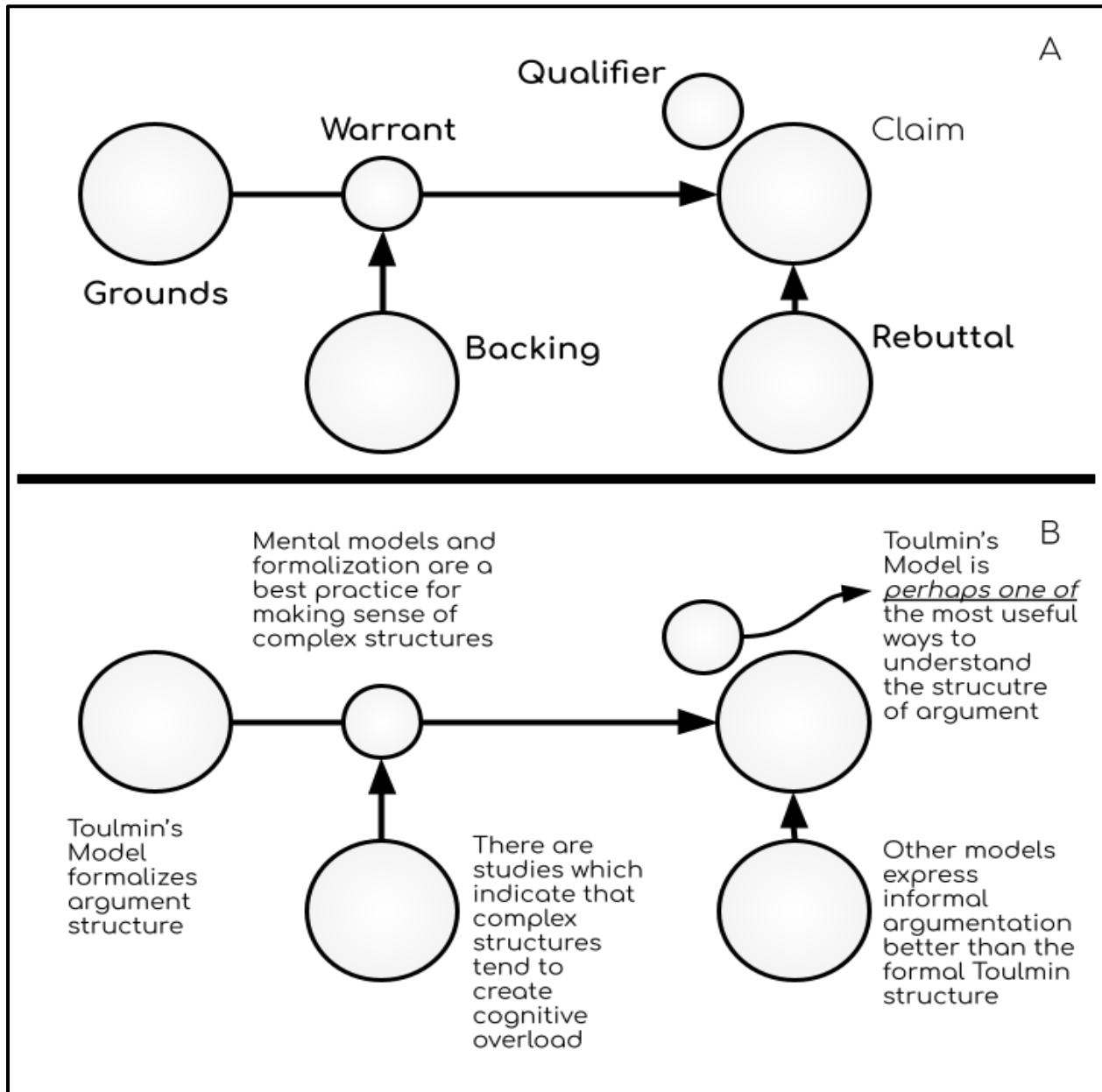


Figure 15. (A) Toulmin's Model of Argumentation and (B) an example implementation

Toulmin himself asserted that this framework was not a “final” model for argumentation. Instead, it was the product of an exploration of the layout of argument driven by the intent to see *logic* developed into a formal science built on jurisprudence (legal philosophy). As such, it carries limitations, and has served as a foundation for myriad analyses and models which seek to address or overcome these limitations. It could be argued that chief among these limitations is addressing the interconnectedness of claims and their components - as the grounds, backing, and rebuttal attached to a claim can each be claims in their own right, and as such, have their own connected structures.

### **Stab and Gurevych Model for Argument Annotation**

The Stab and Gurevych model for the annotation of argument is designed for extraction of granular and modular components of argumentation in persuasive essays. It is designed specifically for managing the relationships among claims and their support, refutations (attacks), and their own support or refutation for other claims. Of value here, is that this model uses a very simple set of rules and components in order to represent complicated arguments.

#### **Statement**

A statement is a piece of text which might contain components of argument and can be used as the basis for annotation.

#### **Major Claim**

The major claim is at the “center” of discourse, usually expressed rhetorically in the introduction of a piece of writing - indicating the author's stance on a particular topic.

#### **Claim (Support or Attack)**

This object expresses itself as *grounds* or *rebuttal* to the major claim by merit of the assigned “support” or “attack” relationship referred to as its “stance attribute”. A claim, like the major claim, is considered to be a “controversial statement” which will be supported or attacked within a text.

#### **Premise**

The premise supports (or attacks) the validity of a claim or major claim, or another premise by giving a reason.

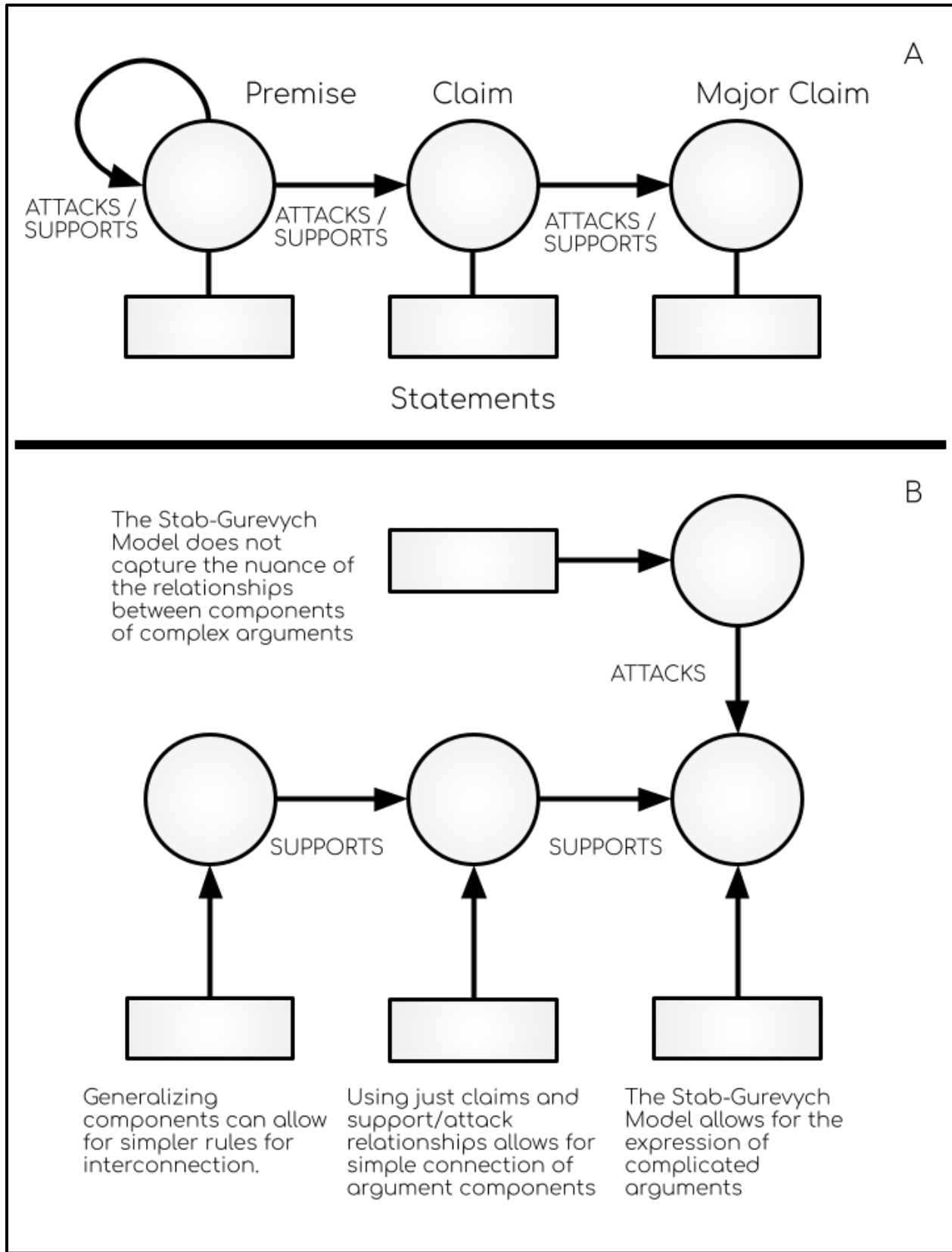


Figure 16. (A) Stab-Gurevych Model for argument annotation and (B) an example implementation.

While this more general framework allows for complex interconnections between claims and helps analyze structured discourse, it, like the Toulmin model, comes with limitations. Some of these limitations could be interpreted to be a product of an intentionally constrained scope, as the work was intended only to advance the annotation of argument structures in a particular medium. For example, it provides no equivalent component to Toulmin's qualifier, and components cannot form relationships with the relationships between components (such as the warrant in Toulmin's model, which addresses the relationship between the grounds and the claim). Further, by merit of its focus on a major claim, it is best suited for annotating documents which are built via constrained writing tasks where all other claims sit in some hierarchy beneath the central claim.

### **Digital Rhetorical Ecosystem 3-Layer Model (DRE3)**

The Digital Rhetorical Ecosystem 3-Layer Model or DRE3 model was designed to integrate rhetorical analysis with ecological theory in such a way as to make it compatible with a crowdsourced and computational analytics pipeline intended to produce a wide range of information products, such as publications and briefs, estimative and predictive metrics, and training data for automated analysis systems. It moves beyond rhetorical structure to consider object references and other content, and most importantly, is intended for analysis of argumentation communicated through multimodal content, with a specific emphasis on image memes. The DRE3 model does not structure argumentation so much as it structures the process of extraction of components and references within arguments embedded in content. The purpose of this focus is to enable analysis of argumentation at the level of public discourse, or of argumentation within the context of a rhetorical ecosystem. The process of integrating an artifact (i.e., an image-meme) is expressed in 3 stages:

#### **Entity Identification**

The first phase of DRE3 analysis is entity identification. In this phase, an analyst tags visible or implied entities, such as persons, organizations, locations, or concepts - enabling rapid collation of content with similar subjects. Further, it informs analysis in succeeding stages.

#### **Rhetorical Analysis**

The second phase of DRE3 analysis is rhetorical analysis. In this phase, an analyst decodes the relationships between the entities and their placement within the content. The objective is synthesis of these relationships into a central claim (or set of claims) made within the content.

### Hidden State Identification

The final stage of DRE3 analysis is hidden state identification. In this phase, the analyst attempts to identify underlying broad claims which are implied by the claims within the content and by similar claims across other content.

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |                                                                                                                                                                                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Entity Tagging</b> I                                                                                                                                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                 |
| <ul style="list-style-type: none"><li>• Persons<ul style="list-style-type: none"><li>◦ Steven Crowder</li></ul></li><li>• Shows/Podcasts<ul style="list-style-type: none"><li>◦ "Change my mind"</li></ul></li><li>• Other<ul style="list-style-type: none"><li>◦ Mug</li><li>◦ Microphone</li><li>◦ "Change my mind" banner</li><li>◦ College Campus</li><li>◦ Grinning Man</li></ul></li><li>• Concepts<ul style="list-style-type: none"><li>◦ Toulmin Model</li></ul></li></ul> |                                                                                                                                                                                                                                                    |
| <b>Rhetorical Analysis</b> II                                                                                                                                                                                                                                                                                                                                                                                                                                                      | <b>Hidden State Identification</b> III                                                                                                                                                                                                             |
| <ul style="list-style-type: none"><li>• Stated argument: "Toulmin model isn't useful [outside of academic contexts]"</li><li>• Implied argument, that you couldn't structure an argument using the Toulmin model that would change the mind of the poster - in support of the stated argument</li></ul>                                                                                                                                                                            | <ul style="list-style-type: none"><li>• There is a gap between academic formalization of phenomena and the phenomena itself; found in similar memes</li></ul>  |

Figure 17. Example implementation of DRE3 model

The DRE3 model, like other argumentation and argumentation analysis models, comes with its own limitations. For example, the extraction of hidden states and arguments is heavily influenced by the analyst, given the often esoteric and ambiguous nature of multimodal content. Its largest limitation may be that its value depends on the successful implementation of crowdsourcing solutions to annotate content, tag entities. and provide feedback on analyses.

# Systems Design

## Key Design Elements of Crowdsourcing Solutions

Attempts to solve problems, raise funds, collect evidence, or analyze data using large numbers of individuals is referred to as “crowdsourcing”. Crowdsourcing solutions are deployed where automated approaches may not be effective or possible, and have been successfully deployed in a myriad of use-cases even where the crowd would not necessarily be perceived as competent in addressing the relevant solution space, such as using gamers to assist in the analysis of genetic and astronomical data. In this vein, crowdsourcing solutions have to be tailored to their use case, solution space, and crowd, resulting in a number of use-case specific categories of patterns of crowdsourcing solutions, such as prediction markets, where crowds are being used to predict events; or serious games, where games or game-like mechanisms are used in order to incentivize engagement or allow for a crowd to contribute to solution spaces for which they do not have the relevant competencies. Crowdsourcing solutions have to be carefully tailored to the conditions of their implementation for functional reasons, but also because of their dependence on engagement, it is difficult to make any single approach reliable - often, attaining reliability remains difficult even within a particular domain or use-case. Analyses of crowdsourcing solutions across the spectrum of use-cases suggest there are at least a dozen interconnected elements in common which contribute to likelihood of success, below these elements are compressed into three principles relevant to our purposes:

### Task Communication

The system and users should have affordances to delineate, transmit, or broadcast task-related requests to others that are appropriate given the size of the crowd, diversity of the competencies of the crowd, complexity of the solution space, and number of requests that may be active at any given time. Difficulties in communications cost effort, time, and resources, and most importantly, impact both the likelihood of users attempting to solve tasks or their ability to broadcast tasks they cannot solve to others who can.

### Task Solution Space

The solution space of tasks should have a complexity which is appropriate given both the competence and size of the crowd. The more agents involved in a solution space, the more modular, granular, specific, and well-defined the tasks and the

measurement of their success must be in order for them to coordinate coherently. As an illustrative example, 100 people can come together to build a brick wall, but they cannot write a coherent novel. The more subjective the solution space is, and the less modular completed tasks are from one another (e.g., where each task impacts the solution space of the next), then the more individuals that are added, the more disagreement that will form within the crowd - contributing to incoherent results or lack of engagement. Where subjectivity in solution space is impossible to avoid, contributions must be well structured and as granular as is practicable.

#### **Task Motivation and Feedback**

The crowd should be given clear, relevant feedback about their interactions, and should have incentives which are appropriate given their competencies, the costs of performing tasks, and the potential impacts of incentives on outcomes. What constitutes relevant feedback or an appropriate incentive may, arguably, be more an art than a science - as some crowds may be effectively motivated and stimulated by feedback regarding their contributions to a community, whereas others may need more explicit incentives. However, incentives have to be tailored not only to the community but to the solution space itself, as extrinsic motivations such as currency or “points” can come at the expense of intrinsic motivation and therefore at the expense of the intrinsic value of the solution space. As an illustrative example, offering currency as a reward for producing 1000 words on a topic may be effective for generating words, but ineffective at generating value within them. Continuing with this example in order to illustrate the lack of standardized approaches across implementations: if individuals might have already been producing these 1000 words, and the currency was just a motivation for them to bring what they were already producing to the system, there is less risk of meaningless submissions, though moderation, reputation, and identity verification systems would still have to be put in place in order to reduce impacts on submission quality.

#### **Coonradt’s Principles of Engagement**

Coonradt, the “grandfather of gamification” asserted that activities which require extensive effort have 6 elements that must be present in order to be persistently engaging:



### **Clear Goals**

The objectives of the work are clear and well scoped, making navigation toward those goals manageable.

### **Scorekeeping**

The measurement of performance outcomes is clear, comparable, and unambiguous.

### **Feedback**

Given the clarity of objectives and performance outcomes, individuals participating in a game or gamified system have reasonable basis to consider the impact of certain behaviors on results.

### **Choice**

Games and game mechanisms provide players with choices, some clearer than others - the clearer the choices, the more valuable feedback becomes, and the more opportunities are provided for players to invest in understanding the impacts of their choices on outcomes and in innovating or adapting those choices.

### **Field of Play**

The time and space in which the game is played are well scoped, so players have clear expectations entering this scope: they know what to expect, what is expected of them, and that the game will eventually end, and therefore that they will have time to rest if they exert themselves.

### **Skin in the Game**

This concept from game theory was communicated to a much wider audience in the book of the same name by Nassim Taleb—that players need to acknowledge some value on the table, some potential cost or gain at stake that is tied to their performance in order to play effectively and fairly.

## **Key Principles for Social Systems Engineering**

Social Systems Engineering (SSE) is concerned with the design of systems which involve or are driven by interactions between social agents. In traditional engineering, final system states can often be defined completely and provide highly reliable behavior through the use of (i) separations of concern among components, (ii) clear causal relationships and formal interfaces resulting in mathematically or algorithmically predictable phenomena, (iii) high reliability controls on interfaces,

and (iv) predictably adaptive components with highly accurate feedback mechanisms. Humans have hidden states, hidden interests, and highly adaptive policy. As such, any system which includes human inputs will have a reliability which holds a nonlinear relationship with the degrees of freedom of said inputs and their impact on the system. Any system which has outputs that depend on the interactions between flexible human inputs is thus, by default, a complex system. The company AIE Nexus offers the following principles to help SSE clients define requirements and set expectations:

**Simple Rules Create Complex Structures**

Rules for interfaces and mechanisms should be as simple as possible, be moderated only by local conditions, result in modular and granular products, and rarely, if ever, contain exceptions. The relationships between the resulting granular products should be equally simple, and allow for flexible modularity in order to seed opportunities for the emergence of complex subsystems and structures.

**You Cannot Design the Social System's Mature State**

For the majority of cases, you cannot predict from the starting state or from mechanisms or infrastructure what the resulting mature system will look like or if it will ever reach a mature state, even if a prior system had identical mechanisms and infrastructure and arguably equal starting state. While it is tempting to attempt rigorous definition and design of the mature state, the focus should instead be placed on requirements, controls, and standards which reduce likelihood of system failure and withdrawal of users, provide the users with value, control the structure of the systems outputs, and allow for iterative adaptation over time.

**Retreatism and Withdrawal are the Default**

Social systems implemented from scratch should have their mechanisms and rules designed with the assumption that new users are looking for a reason to leave until they have enough stake in the system to look for reasons to stay. Thus, the mechanisms and rules for interaction should be designed in such a way that individuals, by merit of use, are always accumulating stake in the system.

**Harness Rebellion, Error, and Conflict**

Assume that circumvention of the rules and use of the system's human interfaces will be misused, abused, and rebelled against

and that users will come into conflict. Do not assume that any component of the system is foolproof against any error or misuse. Instead, consider what adaptations or supplementary mechanisms can allow users or moderators to address or quarantine misuse and enable engineers to understand misuse in order to iteratively adapt the system over time.

#### **Humans are Components in the System, Not Just Consumers**

Social systems should be designed with the assumption that humans are “components” within that system, in addition to their roles as “users.” With this expanded perspective, considering the “engineering” of human behavior (both as individuals and in their capacity as organizational actors) to increase reliability of outcomes becomes a default.

#### **Meet the User Where They Are**

Engineering user behavior or creating incentives from scratch is a perilous and generally unreliable process. Humans are not blank slates, and controlled environments with captured audiences can create misunderstandings about how game-theoretically-sound incentives may work in the wild. Wherever possible, mechanisms should be designed to harness, facilitate, and accommodate existing incentives, motivations, interests, processes, norms and expectations, and activities.

#### **Trade-Offs are Inevitable, Prioritize Wisely**

Every social system will be accompanied by trade-offs. For example, efficiency comes at the expense of reliability and quality and vice versa, and quality controls will negatively impact user experience in the short term in exchange for positive impacts in the long term. Trade-offs must be made explicit for participant evaluation, considered and prioritized carefully, and recognized as both unavoidable and amenable to co-management for enhanced system sustainability and resilience.

#### **If Value to the User Depends on other Users, the System must be Seeded**

If the value to the user depends on other users, then organic adoption in early stages is unlikely, as a lone user will likely not stay long enough to await the arrival of other users. A system must provide notable value to users in isolation or be seeded

with inorganic users (e.g., paid users, stakeholders) in advance of achieving scale and maturity that is prerequisite to organic growth.

**Clear, Meaningful Feedback is Good,  
Embodiment is Better**

Clear, objective, and consistent feedback is standard practice for behavioral modification. However, wherever possible, behavior should be modified via affordances and structure to enhance reliability of system performance. For example, where users should exercise caution, it is more effective to implement affordances which require them to act out a process or ritual that requires caution or careful thought than it is to inform them to be cautious or to provide feedback where they failed to exercise caution.

**MMOS Recipe for Serious Games**

While there are numerous serious games designed for both research and education purposes, those implemented by the company Massively Multiplayer Online Science (MMOS) have been among the most impactful in the history of the field. To some extent, this success is due to their focus on finding ways to harness effort that is already being expended through existing activities, as opposed to building new activities entirely from scratch. The founders of MMOS have discussed a “recipe” for converting those individuals already engaged with digital activities into “virtually limitless computation engines for citizen science” <sup>3</sup>. An outline of this recipe, originally developed for use in the game EVE Online, is adapted for general use here:

**Task Discovery**

Find large-volume, modular tasks which require human annotation, analysis, or evaluation and cannot be effectively or reliably automated.

**User Discovery**

Find activities with which users with relevant competencies and capabilities are already highly engaged.

**Task Mapping**

Map the modular tasks to adaptations within the existing activities that harness or add to existing incentives while facilitating the performance of said tasks.

### **Theme Mapping**

Make adaptations to the activity “aesthetically fitting and thematically adoptable” by the users.

### **Feedback**

Make it clear to the users that by participating, they are making impacts beyond their own community.

### **Integration with Automation**

Use the resulting data as training data for automated systems.

## **Active Inference Principles of Trust**

The paper “Active Inference in Modeling Conflict: A Framework for Modeling Conflict in Business, Operations, Legal, Technical, and Social Contexts” presents 5 insights regarding trust and its impact on operations, informed by the Active Inference cognitive modeling framework. In conjunction with the ability to use ontology and formalization as a basis for behavioral engineering, these 5 insights can be argued to be principles for the design of collaborative systems:

### **Trust is Synonymous with Reliability**

Trust can be characterized as a high level of certainty regarding the expectations of the policies and actions of another object, actor, or system. For example, we can trust a machine to function or not function, just as we could trust another person to act or not act.

### **Trust can be Externalized to Interfaces**

Actors do not need to build trust with other actors if a higher level of trust can be assigned to an intermediary or interface through which they can instead engage. For example, we can externalize our trust to receive payment from a stranger to a payment system, as opposed to requiring trust in the stranger.

### **Trust can be Externalized to Symbols and Signals**

Actors do not need to build trust with other actors if a higher level of trust can be assigned to symbols which reliably predict expectations about the environment. For example, “traffic signals allow drivers to externalize their trust to signals which inform the projection of other drivers’ behavior, as opposed to being left to develop trust with other drivers in order to share the road”.

### **Trust is a Prerequisite for Efficient Information Sharing**

There are high costs associated with vetting information or sources of information, making communication without symbols, signals, interfaces, protocols, or pre-established personal trust cognitively expensive. Communication without externalization of trust or personal trust is axiomatically inefficient, either by merit of the costs of vetting, or the probabilistic risk of accepting low quality information or disinformation in lieu of vetting.

### **Trust is a Prerequisite for Collaborative Enterprise**

In order to engage in collaborative enterprise, actors must have trust in relevant actors or externalize trust to a degree that is commensurate with associated risks.

## **Principles Related to Sustainability of a Commons**

The study of “commons management” is rooted in the analysis and design of shared-resource systems, such as fisheries and grazing lands. While originally focused on natural resource management, commons management principles and research has found use in approaching other systems, with both real and abstract, or tangible and intangible resources, that encounter similar problems of common-resource use, such as conflicts over use, overuse, pollution, congestion, free-riding, unequal distribution, and availability of recourse. Hess and Ostrom, in their book, *Understanding Knowledge as a Commons*, provided eight principles for “robust, long-enduring, common-pool resource institutions”:

### **Clear Boundaries**

Where boundaries over what constitutes the common-pool being managed are blurred; responsibilities, needs, requirements, protocols, rules, and jurisdictional authority are blurred as well.

### **Rules are Well Matched to Local Needs**

Empirical studies on common-pool resource governance have consistently indicated that “no single set of specific rules... had a clear association with success”. Instead, rules needed to be adapted and adjusted to local requirements in order to sustain a resource commons.

### **Those Affected by Rules can Participate in Modifying Them**

A commons “is often most efficient and durable when individuals affected by a resource regime” can participate in

modifying its rules. This is in part because those who are affected are in the best position to understand how rules need to be adapted to map well to local needs, and more importantly are in the best position to understand what rules will be maladaptive or dysfunctional. Adaptive, sustainable governance systems tend to have the following characteristics:

- Information availability
- Recourse capabilities
- Rule compliance capability
- Rule-related infrastructure
- Preparation for and expectation of change

All of these characteristics require that rules be functional and well-mapped to the local environment and that those who are within the system participate in modifying them over time.

#### **Right to Establish Local Rules**

In order to enable rules which are mapped to local needs and avoid rules which generate dysfunction or encourage subversion, those affected by rules must be able to participate in modifying them. Those affected by rules cannot participate in modifying them if external authorities do not recognize their right to engage in establishing and modifying local rules.

#### **Community is Empowered to Self-Monitor**

Sustainability requires ongoing monitoring and evaluation. Those that are engaging in the interactions within the commons are in the best position to spot wrong-doing, negligence, or failure to meet standards.

#### **Graduated System of Sanctions of Bad Behavior**

Effective governance requires that there are “reasonable standards for small variations that [will] always occur due to errors, forgetfulness, and urgent problems”, and a graduated system of sanctions which become more severe to those “who do not learn’ from initial, more lenient encounters. The system itself also needs to graduate over time, increasing its severity and specificity. A governance system will often need to begin somewhat informally, as too many requirements for compliance too early can create disincentives for participation,

and then develop over time into having more strict and clear sanctions for undesired behavior.

### **Simple and Low-Cost Mechanisms for Conflict Resolution**

Conflict can provide opportunities for information discovery and refinement if facilitated and tempered in a controlled environment, in much the same way an engine produces work from heat. The goal of the governance system is not necessarily to end all potential for conflict, but to harness it to help the system as a whole reduce externalities and the potential for conflict to be destructive. Conflict resolution affordances need to be available, accessible, and affordable in order to avoid uncontrolled conflict.

### **Nested Enterprise**

Sustainable commons tend to be those which have “nested enterprises” or those which have conflict resolution, monitoring, sanctioning, and other governance activities nested within a larger structure with “multiple layers” of activity and organizational components.

## **Infinite Games for Infinite Teams**

The white paper “Infinite Games for Infinite Teams” introduced a role-based “case management [system] for knowledge mapping”. This system is expressed as a game which acts as a crowdsourcing solution for mapping narrative, arguments, and concepts together. The game begins with a “workspace” which is initialized with a “seed-meme”, such as “the central argument of a paper” or a hypothesis being investigated. The game has two modes, explore and exploit. In explore mode, “all team members can see all information”. In exploit mode, players then take on a role as either a Red, Blue, or Green contributor, each attaching concepts, documents, and arguments to the seed-meme.

### **Blue Contributor**

Blue contributors take a defensive stance in making connections to the seed-meme, considering questions such as:

- What have previous thinkers/movements/stories done to counter this meme?
- How might the meme or narrative be instantly and transparently debunked?



### **Red Contributor**

Red contributors take a more aggressive approach to contribution, considering questions such as:

- What would be an effective approach to changing people's mind, not just informing them or "raising awareness"?
- What is the most direct and devastating attack on the ignorance surrounding this topic? \

### **Green Contributor**

Blue and Red contributors focus on evidence and logic, whereas Green contributors focus on "evocation of emotion, anecdotes, and narrative." Green introduces "kairos in the system, that is an understanding, sense, and sequence to the memes in a space", considering questions such as:

- How can ideas be communicated to multiple audiences?
- How might the same messaging be effective across audiences & media formats?

The contributions, when taken together, map an emergent, stigmergic memetic landscape. Disparate concepts from multimodal digital media are linked, providing a unique form of situational awareness around a topic.

## **Narrative Information Management**

The paper, Narrative Information Management asserts that fields and specializations which intend to design and implement systems, protocols, and procedures to manage, synthesize, curate, and search digital information generally need to account for the provision of the following features:

### **Managing Information Gaps**

The ability to recognize gaps in the knowledge base in order to direct attention, or to recognize gaps in personal knowledge and address them using an existing knowledge base.

### **Facilitating Situational Awareness**

The ability to stay apprised or be notified of changes and updates in the relevant environment despite pressures of information volume, complexity, and rate of change.

### **Providing Descriptive and Explanatory Information**

The ability to “dig” into particular components and objects for summaries and background information.

### **Compression**

The ability to compress complex information structures using visualization, structure, collation, curation, ontological, and interactive mechanisms.

### **Case Management and Providing Prescriptive Information.**

The ability to follow particular chained events or objects and be provided with actionable procedure-related information, such as best practices or next steps.

### **Synthesizing Intelligence**

The ability to synthesize information in the knowledge base in order to generate new information products.

### **Facilitating Communication**

The ability for users of the knowledge base to coordinate in a structured and coherent manner even where roles or expertise are heterogeneous.

### **Handling of Errors and Inconsistencies.**

The ability for users to be directed toward and remediate errors and inconsistencies.

### **Management of Trust Signals**

The ability for users to send, receive, assign, parse, and isolate signals of trust related to evaluation of information and of the intents and competencies of actors.

### **Social Systems Engineering**

The ability for the system to adjust and modify behavior of the users in a way which promotes the health and sustainability of the system.

## **Framework for Synthetic Intelligence Guilds**

The paper “The Synthetic Intelligence Guid: A Social Technology for a Digital Bazaar”, in proposing the foundations for a sensemaking-oriented community of practices, offers the basis for a number of generalizable prerequisites for decentralized knowledge management systems:

**Prevent Race-to-the-Bottom and Rivalrous Mechanics**

Mechanism design should prevent, address, or offset the impacts of Hobbesian, multipolar, and Thucydidean traps, coordination failures, negative-sum game theoretic dynamics, free rider, principal-agent problems, and other related dynamics.

**Prevent Centralized Capture**

Mechanism and underlying structure design should prevent, circumvent, or deincentivize the centralized capture or clique-control of any particular aspect of the system.

**Shared Situational Awareness, Decision-Making, and Dissemination**

Mechanism and underlying structure design should allow for and facilitate shared situational awareness of the information environment, support decision making activity, and allow for directed dissemination.

**Clearinghouses**

The system should provide simple clearinghouses for setting of information-related contracts and exchange of information products and services in order to break up silos and allow the flow of critical information between specialized groups.

**Direction of Attention toward Opportunities and Gaps in the Knowledge Base**

Mechanism and underlying structure design should incentivize search for and direct attention to opportunities and gaps in the knowledge base (e.g., “low hanging fruit”).

**Domain-Specific Agents and Teams as opposed to Large Central Bureaucracy**

As opposed to central bureaucratic structure, autonomous agents and teams should be incentivized and empowered to address challenges within the information environment.

**Standards for Crowdsourcing and Crowdsourced Standards**

The system should have structure and standards allowing for contributions at scale, and allow for the implementation, development, and spread of locally developed standards.

**Group Transferable, Network Maintained Reputation Systems**

Communities should be empowered to develop and manage local reputation systems with opportunities for information sharing between groups.

**Right to Bundle, Buy, and Broker**

Communities should have affordances to bundle, buy, and broker information products.

---

## **Works Cited**

---

## Chapter I Works Cited

---

1. Heuer RJ. Psychology of Intelligence Analysis. Center for the Study of Intelligence; 1999.
2. Avey AE. Function and Forms of Thought. Henry Holt and Company; 1927.
3. Fox JR, Park B, Lang A. When Available Resources Become Negative Resources: The Effects of Cognitive Overload on Memory Sensitivity and Criterion Bias. *Communic Res.* 2007;34: 277–296.
4. Bawden D, Robinson L. The dark side of information: overload, anxiety and other paradoxes and pathologies. *J Inf Sci Eng.* 2009;35: 180–191.
5. Sutcliffe K, Weick KE. Information overload revisited. *The Oxford handbook of organizational decision making.* Oxford University Press; 2009.
6. Weick KE. The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. *Adm Sci Q.* 1993;38: 628–652.
7. O'Reilly CA. Individuals and Information Overload in Organizations: Is More Necessarily Better? *AMJ.* 1980;23: 684–696.
8. Bennett M, Waltz E. Counterdeception principles and applications for national security. Artech House; 2007.
9. Edmunds A, Morris A. The problem of information overload in business organisations: a review of the literature. *Int J Inf Manage.* 2000;20: 17–28.
10. Kight AC, Narayanan S. Cognitive engineering in algorithm development for multisensor data fusion in military applications. *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2004.* International Society for Optics and Photonics; 2004. pp. 148–155.

11. Kirsh D. A Few Thoughts on Cognitive Overload. *Intellectica Revue de l'Association pour la Recherche Cognitive*. 2000;1: 19–51.
12. Fayol M, Largy P, Lemaire P. Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. A study in French written language. *The Quarterly Journal of Experimental Psychology Section A*. 1994;47: 437–464.
13. Lyons TF. Role clarity, need for clarity, satisfaction, tension, and withdrawal. *Organ Behav Hum Perform*. 1971;6: 99–110.
14. Singh J, Goolsby JR, Rhoads GK. Behavioral and Psychological Consequences of Boundary Spanning Burnout for Customer Service Representatives. *J Mark Res*. 1994;31: 558–569.
15. Dieter M. Dark Patterns: Interface Design, Augmentation and Crisis. *Postdigital Aesthetics*. 2015. pp. 163–178. doi:10.1057/9781137437204\_13
16. Waltzman R. The Weaponization of Information: The Need for Cognitive Security. RAND Corporation; 2017 Apr. doi:10.7249/CT473
17. Cordes RJ, Friedman DA. Emergent Teams for Complex Threats. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
18. Shoup JR, Hinrichs TW. *Literature and Leadership: The Role of the Narrative in Organizational Sensemaking*. Routledge; 2019.
19. Andersen D, Ravn S, Thomson R. Narrative sense-making and prospective social action: methodological challenges and new directions. *Int J Soc Res Methodol*. 2020;23: 367–375.
20. Zeff SA. *The Routledge Companion to Accounting History*. Accounting, Business & Financial History. 2010. pp. 107–112. doi:10.1080/09585200903504298
21. Szubin HZ, Porten B. An Aramaic Joint Venture Agreement: A New Interpretation of the Bauer-Meissner Papyrus. *Bull Am Schools Orient Res*. 1992;288: 67–84.
22. Verboven K, editor. *Complexity Economics: Building a New Approach to Ancient Economic History*. Springer Nature; 2020.
23. Newman JH. *Essay on the Development of Christian Doctrine*. 1845.
24. McLuhan M. *The Gutenberg Galaxy*. 1962. New York: Signet. 1969. Available: [http://www.robertdelossa.net/docs/ca2/Marshall%20McLuhan%20handout\\_1\\_CA2.doc](http://www.robertdelossa.net/docs/ca2/Marshall%20McLuhan%20handout_1_CA2.doc)

25. Kuhn TS. *The Structure of Scientific Revolutions: 50th Anniversary Edition*. University of Chicago Press; 2012.
26. Olsson G. *Abysmal: A Critique of Cartographic Reason*. University of Chicago Press; 2010.
27. Cordes RJ. *The Pipeline*. The Pipeline: EdTech, Knowledge Management, and NATSEC. COGSEC; 2021.
28. Dickey E. *Ancient Greek scholarship : a guide to finding, reading, and understanding scholia, commentaries, lexica, and grammatical treatises, from their beginnings to the Byzantine period*. New York; Oxford: Oxford University Press; 2006.
29. Dedijer S. Ragusa Intelligence and Security (1301-1806): A Model for the Twenty-First Century? *Int J Intell CounterIntelligence*. 2002;15: 101–114.
30. Šundrica Z. *The Intelligence Activities of the Dubrovnik Republic in the Austro-Turkish War 1737-1739*. *Dubrovnik Annals*; 2001. Report No.: 8299. Available: [https://hrcak.srce.hr/index.php?id\\_clanak=8299&show=clanak](https://hrcak.srce.hr/index.php?id_clanak=8299&show=clanak)
31. Austin NJE, Rankov NB. *Exploratio: Military and Political Intelligence in the Roman World from the Second Punic War to the Battle of Adrianople*. Psychology Press; 1998.
32. Cordes RJ, Friedman DA. *The Facilitator's Catechism*. In: Friedman DA, Cordes RJ, editors. *The Great Preset: Remote Teams and Operational Art*. COGSEC; 2020.
33. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. Big Data: Astronomical or Genomical? *PLoS Biol*. 2015;13: e1002195.
34. International Society for Biocuration. *Biocuration: Distilling data into knowledge*. *PLoS Biol*. 2018;16: e2002846.
35. Rydning J, Reinsel D, Gantz JF. *Worldwide Global DataSphere Forecast, 2020–2024: The COVID-19 Data Bump and the Future of Data Growth*. IDC; 2020 Apr. Report No.: US44797920. Available: <https://www.idc.com/getdoc.jsp?containerId=US44797920>
36. Rydning J, Reinsel D. *Worldwide Global StorageSphere Forecast, 2021–2025: To Save or Not to Save Data, That Is the Question*. IDC; 2021 Mar. Report No.: US47509621. Available: <https://www.idc.com/getdoc.jsp?containerId=US44797920>
37. National Science Foundation. *Publications output: US Trends and international comparisons*. {National Science Foundation}; 2019. Report No.: NSB-2020-6. Available: <https://nces.nsf.gov/pubs/nsb20206/>



38. Brainard J. Scientists are drowning in COVID-19 papers. Can new tools keep them afloat? In: ScienceMag [Internet]. 2020 [cited 5 Aug 2020]. doi:10.1126/science.abc7839
39. Cohen PR. The Big Mechanism Program and the Future of Scholarship. DARPA Biological Technologies Office Symposium; 2015 Feb 11. Available: [https://www.youtube.com/watch?v=GCaXSyY\\_C9Y](https://www.youtube.com/watch?v=GCaXSyY_C9Y)
40. Cohen PR. Acquiring Big Mechanisms by Reading Primary Literature Seminar. John Hopkins Whiting School of Engineering; 2015. Available: <https://www.clsp.jhu.edu/events/acquiring-big-mechanisms-by-reading-primary-literature-paul-cohen-information-innovation-office-darpa-and-the-university-of-arizona/#.XL0nLehKiUk>
41. Fournie DA. Harsh Lessons: Roman Intelligence in the Hannibalic War. *Int J Intell CounterIntelligence*. 2004;17: 502–538.
42. Friedman DA, Cordes RJ. Knowledge Management Archipelago. Zenodo. COGSEC; 2021. doi:10.5281/zenodo.5034809
43. Zegart A. Intelligence Isn't Just for Government Anymore. *Foreign Affairs*. 2 Nov 2020. Available: <https://www.foreignaffairs.com/articles/united-states/2020-11-02/intelligence-isnt-just-government-anymore>. Accessed 4 Nov 2020.
44. Schwartz D. *Encyclopedia of Knowledge Management*. Idea Group Inc (IGI); 2005.
45. Dalkir K. *Knowledge management in theory and practice*. London, England: Routledge; 2013. doi:10.4324/9780080547367
46. Girard J, Girard J. Defining knowledge management: Toward an applied compendium. *Online Journal of Applied Knowledge Management*. 2015. Available: [https://www.iiakm.org/ojakm/articles/2015/volume3\\_1/OJAKM\\_Volume3\\_1pp1-20.pdf](https://www.iiakm.org/ojakm/articles/2015/volume3_1/OJAKM_Volume3_1pp1-20.pdf)
47. Bonovitz C. Making Sense of COVID-19 with Child Patients: From Screen to Screen. *Psychoanal Dialogues*. 2020;30: 628–629.
48. Waltz E. *Knowledge Management in the Intelligence Enterprise*. Artech House; 2003.
49. Black MB. *Coalition Command, Control, Communications, Computer, and Intelligence Systems Interoperability: A Necessity or Wishful Thinking? Illustrated edition*. BiblioScholar; 2012.

50. Lebeuf C, Storey M-A, Zagalsky A. How Software Developers Mitigate Collaboration Friction with Chatbots. arXiv [cs.HC]. 2017. Available: <http://arxiv.org/abs/1702.07011>
51. Cordes RJ, Friedman DA. Reimagining Maps. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
52. Kujansivu P. Is there something wrong with intellectual capital management models? Knowledge Management Research & Practice. 2009;7: 300–307.
53. Fakhar Manesh M, Pellegrini MM, Marzi G, Dabic M. Knowledge Management in the Fourth Industrial Revolution: Mapping the Literature and Scoping Future Avenues. IEEE Trans Eng Manage. 2021;68: 289–300.
54. Neng-fu K. Chinese Library Science in the Twelfth Century. Libr Cult. 1991;26: 357–371.
55. Bray DA. Literature review - knowledge management research at the organizational level. Emory University; 2007. Available: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=991169](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=991169)
56. Harvey JF. Toward a definition of international and comparative library science. Int Libr Rev. 1973;5: 289–319.
57. El-Akruti K, Dwight R. A framework for the engineering asset management system. Journal of Quality in Maintenance Engineering. 2013;39: 88.
58. Daneels A, Salter W. What is SCADA? 1999. Available: <https://accelconf.web.cern.ch/ica99/papers/mc1i01.pdf>
59. Singh P, Klobas J, Anderson K. Information seeking behaviour of electronic records management systems (ERMS) users: implications for records management practices. Human IT. 2007;9: 135–181.
60. Dong W, Ji QD, He LN. An Enterprise-Wide Digital Archive Management System Infrastructure. Appl Mech Mater. 2014;513-517: 2420–2423.
61. Bailey S. Managing the Crowd: Rethinking Records Management for the Web 2.0 World. Facet Publishing; 2008.
62. Bailey S, Vidyarthi J. Human-computer interaction: the missing piece of the records management puzzle? Records Management Journal. 2010;39: 88.
63. Bawden D, Robinson L. Introduction to Information Science. Facet Publishing; 2015.
64. Borko H. Information science: What is it? Am doc. 1968;19: 3–5.

65. Zarri GP. Representation and Management of Narrative Information: Theoretical Principles and Implementation. Springer Science & Business Media; 2009.
66. Ryan M-L. Toward a definition of narrative. In: Herman D, editor. The Cambridge Companion to Narrative. Cambridge University Press; 2007.
67. O'Donnell RA. The Case for Narrative Led Operations in the Canadian Armed Forces. cfc.forces.gc.ca. 2017. Available: <https://www.cfc.forces.gc.ca/259/290/402/305/odonnell.pdf>
68. Lawlor C. Consumption and Literature: The Making of the Romantic Disease. Springer; 2006.
69. Adebayo AL, Davidson Mhonde R, DeNicola N, Maibach E. The Effectiveness of Narrative Versus Didactic Information Formats on Pregnant Women's Knowledge, Risk Perception, Self-Efficacy, and Information Seeking Related to Climate Change Health Risks. *Int J Environ Res Public Health*. 2020;17: 6969.
70. Hinyard LJ, Kreuter MW. Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview. *Health Educ Behav*. 2007;34: 777–792.
71. Brownson RC, Dodson EA, Stamatakis KA, Casey CM, Elliott MB, Luke DA, et al. Communicating evidence-based information on cancer prevention to state-level policy makers. *J Natl Cancer Inst*. 2011;103: 306–316.
72. Herman D. Introduction to the Cambridge Companion to Narrative. In: Herman D, editor. The Cambridge Companion to Narrative. Cambridge University Press; 2007.
73. Rudrum D. On the Very Idea of a Definition of Narrative: A Reply to Marie-Laure Ryan. *Narrative*. 2006;14: 197–204.
74. Carroll S. The Big Picture: On the Origins of Life, Meaning, and the Universe Itself. Illustrated edition. Dutton; 2016.
75. Tufte ER. Visual Explanations: Images and Quantities, Evidence and Narrative. *Computers in Physics*. 1998;12. doi:10.1063/1.168637
76. Cordes RJ, David S, Maan A, Ruiz A, Sapp E, Scannell P, et al. The Narrative Campaign Field Guide - First Edition. 1st ed. Cordes RJ, editor. Narrative Strategies Ink; 2021.
77. Cordes RJ, Friedman DA, Phelan SE. The Innovator's Catechism. In: Friedman DA, Cordes RJ, editors. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.

78. Vyatkin A, Metelkin I, Mikhailova A, Cordes RJ, Friedman DA. Active Inference and Behavior Engineering for Teams. In: Friedman DA, Cordes RJ, editors. *The Great Preset: Remote Teams and Operational Art*. COGSEC; 2020.
79. Vasil J, Badcock PB, Constant A, Friston K, Ramstead MJD. A World Unto Itself: Human Communication as Active Inference. *Front Psychol*. 2020;11: 417.
80. Bouizegarene N, Ramstead M, Constant A, Friston K, Kirmayer L. Narrative as active inference. 2020. doi:10.31234/osf.io/47ub6
81. Freeman M. Narrative as a mode of understanding. *The handbook of narrative analysis*. 2015; 21–37.
82. Ezzyat Y, Davachi L. What constitutes an episode in episodic memory? *Psychol Sci*. 2011;22: 243–252.
83. Benjamin AS. Memory is More than just Remembering: Strategic Control of Encoding, Accessing Memory, and Making Decisions. Benjamin AS, Ross BH, editors. *Psychol Learn Motiv*. 2007;48: 175–223.
84. Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, Norman KA. Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron*. 2017;95: 709–721.e5.
85. Black JB, Bower GH. Episodes as chunks in narrative memory. *Journal of Verbal Learning and Verbal Behavior*. 1979;18: 309–318.
86. León C. An architecture of narrative memory. *Biologically Inspired Cognitive Architectures*. 2016;16: 19–33.
87. Larsen SF. Personal Context in Autobiographical and Narrative Memories. In: Conway MA, Rubin DC, Spinnler H, Wagenaar WA, editors. *Theoretical Perspectives on Autobiographical Memory*. Dordrecht: Springer Netherlands; 1992. pp. 53–71.
88. Li B, Lee-Urban S, Appling D, Riedl M. Crowdsourcing Narrative Intelligence. *Advances in Cognitive Systems*. 2012;1: 1–18.
89. Mateas M, Sengers P. Narrative Intelligence. *aaai.org*; 2003. Report No.: FS-99-01. Available: <https://www.aaai.org/Papers/Symposia/Fall/1999/FS-99-01/FS99-01-001.pdf>
90. De Bra P, Pechenizkiy M, Van Der Sluijs K, Smits D. GRAPPLE: Integrating adaptive learning into learning management systems. *EdMedia+ Innovate Learning. Association for the Advancement of Computing in Education (AACE)*; 2008. pp. 5183–5188.

91. Akpotu C. Exploring knowledge gaps: The knowledge to knowledge process. *ijires.org*; 2019 [cited 1 Jul 2021]. Available: [http://www.ijires.org/administrator/components/com\\_jresearch/files/publications/IJIRES\\_1568\\_FINAL.pdf](http://www.ijires.org/administrator/components/com_jresearch/files/publications/IJIRES_1568_FINAL.pdf)
92. Hipel KW, Ben-Haim Y. Decision making in an uncertain world: information-gap modeling in water resources management. *IEEE Trans Syst Man Cybern C Appl Rev*. 1999;29: 506–517.
93. Ben-Haim Y. *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*. Elsevier; 2006.
94. Shimada T, Futakata A. Automatic link generation and repair mechanism for document management. *Proceedings of the Thirty-First Hawaii International Conference on System Sciences*. [ieeexplore.ieee.org](http://ieeexplore.ieee.org); 1998. pp. 226–235 vol.2.
95. Wilensky R. Personal libraries: Collection management as a tool for lightweight personal and group document management. *Joint Conference on Digital Libraries*. Citeseer; 2001. Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.4238&rep=rep1&type=pdf>
96. Waterson P, Glenn Y, Eason K. Preparing the ground for the “paperless hospital”: a case study of medical records management in a UK outpatient services department. *Int J Med Inform*. 2012;81: 114–129.
97. Wittenberg D. Time. In: Garrett M, editor. *The Cambridge Companion to Narrative Theory*. Cambridge University Press; 2018.
98. Mingo AC, Cerrillo-I-Martinez A. Improving records management to promote transparency and prevent corruption. *Int J Inf Manage*. 2018;38: 256–261.
99. Schenkel A. Investigating the Influence that Media Richness has on Learning in a Community of Practice: A Case Study at Øresund Bridge. In: Hildreth PM, Kimble C, editors. *Knowledge Networks: Innovation Through Communities of Practice*. 2004.
100. Armstrong MB, Landers RN. An Evaluation of Gamified Training: Using Narrative to Improve Reactions and Learning. *Simul Gaming*. 2017;48: 513–538.
101. Adamson K, Sengsavang S, Charise A, Wall S, Kinross L, Balkaran M. Narrative Training as a Method to Promote Nursing Empathy Within a Pediatric Rehabilitation Setting. *J Pediatr Nurs*. 2018;42: e2–e9.
102. Leyzorek M. A missing feature in some records management systems. *Prairie Village*. 1998;32: 46–49.

103. Eslami M, Vaccaro K, Karahalios K, Hamilton K. “be careful; Things can be worse than they appear”: Understanding biased algorithms and users’ behavior around them in rating platforms. ICWSM. 2017;11. Available: <https://ojs.aaai.org/index.php/ICWSM/article/view/14898>
104. Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. Hum Factors. 1995; 9–42.
105. Salmon P, Stanton N, Walker G, Green D. Situation awareness measurement: a review of applicability for C4i environments. Appl Ergon. 2006;37: 225–238.
106. Endsley MR. Measurement of Situation Awareness in Dynamic Systems. Hum Factors. 1995;37: 65–84.
107. Nofi AA. Defining and measuring shared situational awareness. CENTER FOR NAVAL ANALYSES ALEXANDRIA VA; 2000 Nov. Available: <https://apps.dtic.mil/sti/citations/ADA390136>
108. Sarter NB, Woods DD. Situation Awareness: A Critical But Ill-Defined Phenomenon. Int J Aviat Psychol. 1991;1: 45–57.
109. Rosen MA, Fiore SM, Salas E, Letsky M, Warner N. Tightly coupling cognition: understanding how communication and awareness drive coordination in teams. C2 Journal. 2008;2. Available: <https://apps.dtic.mil/sti/citations/ADA513557>
110. Endsley MR, Garland DJ. Situation Awareness Analysis and Measurement. CRC Press; 2000.
111. Dyer M, Lehnert W. Memory Organization and Search Processes for Narratives. Yale University; 1980 Apr. Report No.: ADA084141.
112. Brewer WF. Plan Understanding, Narrative Comprehension, and Story Schemas. AAAI. aaii.org; 1982. pp. 262–264.
113. Schank R. SAM - A Story Understander. Yale; 1975. Report No.: 43.
114. Schank RC. Interestingness: Controlling inferences. Artif Intell. 1979;12: 273–297.
115. Lehnert WG. Narrative Text Summarization. AAAI. 1980. pp. 337–339.
116. Carbonell JG. Counterplanning: A strategy-based model of adversary planning in real-world situations. Artif Intell. 1981;16: 295–329.
117. Paris AH, Paris SG. Assessing narrative comprehension in young children. Read Res Q. 2003;38: 36–76.
118. Siu C-Y. Assessing narrative comprehension in young pre-school. Development. 2011;233: 222–242.

119. Devine RT, Hughes C. Measuring theory of mind across middle childhood: Reliability and validity of the Silent Films and Strange Stories tasks. *J Exp Child Psychol.* 2016;149: 23–40.
120. Cao Y, Kim Y-SG. Is Retell a Valid Measure of Reading Comprehension? *Educ Res.* 2021;32. doi:10.1016/j.edurev.2020.100375
121. Welland RJ, Lubinski R, Higginbotham DJ. Discourse comprehension test performance of elders with dementia of the Alzheimer type. *J Speech Lang Hear Res.* 2002;45: 1175–1187.
122. Alonzo J, Basaraba D, Tindal G, Carriveau RS. They Read, but How Well Do They Understand?: An Empirical Look at the Nuances of Measuring Reading Comprehension. *Assess Eff Interv.* 2009;35: 34–44.
123. allkindsofminds.org. Reading Comprehension - The Attention Components. In: allkindsofminds.org [Internet]. [cited 22 Jul 2021]. Available: <https://www.allkindsofminds.org/mastering-the-challenges-of-reading-reading-comprehension-the-attention-components>
124. Maedche A, Morana S, Schacht S, Werth D, Krumeich J. Advanced User Assistance Systems. *Business & Information Systems Engineering.* 2016;58: 367–370.
125. Vandlandingham G. Through a Glass, Darkly: The Promise and Reality of State Transparency Websites. *Citeseer.* 2015. doi:10.1.1.1074.5098
126. Giuse NB, Williams AM, Giuse DA. Integrating best evidence into patient care: a process facilitated by a seamless integration with informatics tools. *J Med Libr Assoc.* 2010;98: 220–222.
127. Schmidt A. Bridging the Gap Between Knowledge Management and E-Learning with Context-Aware Corporate Learning. *Professional Knowledge Management.* Springer Berlin Heidelberg; 2005. pp. 203–213.
128. Ning K, Sullivan DO. Context modeling and measuring for context aware knowledge management. *Int J Mach Learn Comput.* 2012;2: 287–291.
129. De Silva H, Moessener K, Carrez F. Group knowledge management for context-aware group applications and services. 2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications. [ieeexplore.ieee.org](http://ieeexplore.ieee.org); 2009. pp. 2851–2855.
130. Roy S, Searle M. Scope Creep and Purposeful Pivots in Developmental Evaluation. *CJPE.* 2020;35. doi:10.3138/cjpe.56898



131. Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD. Humans use directed and random exploration to solve the explore–exploit dilemma. *J Exp Psychol Gen.* 2014;143: 2074–2081.
132. Hills TT, Todd PM, Lazer D, Redish AD, Couzin ID, Cognitive Search Research Group. Exploration versus exploitation in space, mind, and society. *Trends Cogn Sci.* 2015;19: 46–54.
133. Tschantz A, Seth AK, Buckley CL. Learning action-oriented models through active inference. *PLoS Comput Biol.* 2020;16: e1007805.
134. White RW, Marchionini G, Muresan G. Evaluating exploratory search systems. *Inf Process Manag.* 2008;44: 433.
135. Adams B, McKenzie G, Gahegan M. Frankenplace: Interactive Thematic Mapping for Ad Hoc Exploratory Search. Proceedings of the 24th International Conference on World Wide Web. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee; 2015. pp. 12–22.
136. Damiano R, Lieto A, Lombardo V. Testing the Effectiveness of Narrative Archetypes for Archive Exploration. Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference. *aaai.org*; 2017. Available: <https://www.aaai.org/ocs/index.php/AIIDE/AIIDE17/paper/viewPaper/15840>
137. Singh J, Nejdl W, Anand A. Expedition: A Time-Aware Exploratory Search System Designed for Scholars. Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. New York, NY, USA: Association for Computing Machinery; 2016. pp. 1105–1108.
138. Abbott DH, editor. *The Handbook of Fifth-Generation Warfare (5GW)*. Nimble Books; 2010.
139. Mayer RE, Moreno R. Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educ Psychol.* 2003;38: 43–52.
140. Wang J, Mendori T, Hoel T. Strategies for Multimedia Learning Object Recommendation in a Language Learning Support System: Verbal Learners Vs. Visual Learners. *International Journal of Human–Computer Interaction.* 2019;35: 345–355.
141. Isabelle LA. *The fruit, the tree, and the serpent: why we see so well*. Harvard University Press; 2009.
142. Latour B. Visualization and Cognition: Drawing things Together. *Studies in the Sociology of Culture Past and Present.* 1990;6: 1–40.



143. Siu Y-T, Presley I. Access Technology for Blind and Low Vision Accessibility. APH Press, American Printing House for the Blind; 2020.
144. Cohn N. Visual narrative structure. *Cogn Sci.* 2013;37: 413–452.
145. Peterson JB. Maps of Meaning: The Architecture of Belief. Psychology Press; 1999.
146. Neumann E. The Origins and History of Consciousness. Routledge & Kegan Paul; 1954.
147. Jaynes J. The origin of consciousness in the breakdown of the bicameral mind. Houghton Mifflin Harcourt; 2000.
148. Kurbatskaya TB, Dobrotvorskaya SG, Ustin PN. Pareidolia illusions in advertising with “hidden” symbolic images. *Life Sci J.* 2014;11: 337–340.
149. Brantley C. The Heilmeier Catechism. IEEE-USA Today’s Engineer. 2012.
150. Goertzel B. The Hidden Pattern: A Patternist Philosophy of Mind. Universal-Publishers; 2006.
151. Gobet F, Simon HA. Expert chess memory: revisiting the chunking hypothesis. *Memory.* 1998;6: 225–255.
152. Gobet F, Simon HA. Templates in chess memory: a mechanism for recalling several boards. *Cogn Psychol.* 1996;31: 1–40.
153. Kent S. Strategic Intelligence for American World Policy. Princeton University Press; 1949.
154. Buja A, Cook D, Swayne DF. Interactive High-Dimensional Data Visualization. *J Comput Graph Stat.* 1996;5: 78–99.
155. Janvrin DJ, Raschke RL, Dilla WN. Making sense of complex data using interactive data visualization. *Journal of Accounting Education.* 2014;32: 31–48.
156. Kanter J. Clinical case management: definition, principles, components. *Hosp Community Psychiatry.* 1989;40: 361–368.
157. Franklin JL, Solovitz B, Mason M, Clemons JR, Miller GE. An evaluation of case management. *Am J Public Health.* 1987;77: 674–678.
158. Khurana H, Basney J, Bakht M, Freemon M, Welch V, Butler R. Palantir: a framework for collaborative incident response and investigation. Proceedings of the 8th Symposium on Identity and Trust on the Internet. New York, NY, USA: Association for Computing Machinery; 2009. pp. 38–51.

159. Campbell AG. Bernardo investigation review. Solicitor General and Correctional Services; 1996.
160. Gebert H, Geib M, Kolbe L, Brenner W. Knowledge-enabled customer relationship management: integrating customer relationship management and knowledge management concepts. *Journal of knowledge management*. 2003. Available: <https://www.emerald.com/insight/content/doi/10.1108/13673270310505421/full/html>
161. Chalmers R. Methodology for customer relationship management. *J Syst Softw*. 2006;79: 1015–1024.
162. Hughes PD. Analysis, Selection, and Implementation of a Case Management System for Local City Government Attorney's Office. Regis University. 2007. Available: <https://publications.regis.edu/theses/440/>
163. Yang M. Man arrested in mistaken identity case locked in Hawaii mental health hospital for two years. *The Guardian*. 4 Aug 2021. Available: <http://www.theguardian.com/us-news/2021/aug/04/hawaii-man-mistaken-identity-case-mental-hospital-two-years>. Accessed 5 Aug 2021.
164. Wang WM, Cheung CF. A Computational Knowledge Elicitation and Sharing System for mental health case management of the social service industry. *Comput Ind*. 2013;64: 226–234.
165. DARPA. Context Management for Effective Transitions (COMET). 2021. Report No.: HR001121S0007-19.
166. Mattis J. Commander's Guidance for Effects-Based Operations. {US Joint Forces Command}; 2018.
167. De Werd P. Critical intelligence studies? A contribution. *Journal of European and American Intelligence Studies*. 2018;1: 109–148.
168. Moore DT. Sensemaking: A Structure for an Intelligence Revolution. Center for Strategic Intelligence Research, National Defense Intelligence College; 2011.
169. Carney DR, Cuddy AJC, Yap AJ. Power posing: brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychol Sci*. 2010;21: 1363–1368.
170. Linnér RK, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet*. 2019;51: 245–257.

171. Gluth S, Rieskamp J, Büchel C. Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. *J Neurosci*. 2012;32: 10686–10698.
172. Brown PB. Reconsidering the Advice in 3 Popular Personal Finance Books. *The New York Times*. 11 Oct 2019. Available: <https://www.nytimes.com/2019/10/11/business/suze-orman-robert-kiyosaki-dave-ramsey-books.html>. Accessed 11 Aug 2021.
173. Remund DL. Financial literacy explicated: The case for a clearer definition in an increasingly complex economy. *J Consum Aff*. 2010;44: 276–295.
174. Lin JT, Bumcrot C, Mottola G, Walsh G. Investors in the United States: A Report of the National Financial Capability Study. FINRA; 2019 Dec.
175. Investment Adviser Association. SEC Fiduciary Standards. In: [investmentadviser.org](https://www.investmentadviser.org) [Internet]. 2019 [cited 13 Sep 2021]. Available: <https://www.investmentadviser.org/home/side-content/sec-standard>
176. Online Trading Platforms & Tools. [cited 11 Aug 2021]. Available: <https://www.tdameritrade.com/tools-and-platforms.html>
177. Borri N. A Bloomberg Terminal Primer. *I Quaderni di Minerva Bancaria*. 2018 [cited 28 Jul 2021]. Available: <https://www.bloomberg.com/professional/solution/bloomberg-terminal/>
178. Brazier J. Dark pool and OTC crypto trading growing as market begins to resemble traditional asset classes. 29 May 2019 [cited 28 Jul 2021]. Available: <https://www.thetradenews.com/dark-pool-otc-crypto-trading-growing-market-begins-resemble-traditional-asset-classes/>
179. Sen. Warren warns of cryptocurrency risks, presses SEC on oversight authority. *Reuters*. 8 Jul 2021. Available: <https://www.reuters.com/technology/sen-warren-warns-cryptocurrency-risks-presses-sec-oversight-authority-2021-07-08/>. Accessed 11 Aug 2021.
180. Al Guindy M, Naughton JP, Riordan R. The Chilling Effect of Regulation FD: Evidence from the SEC Endorsement of Social Media. 2020. doi:10.2139/ssrn.3385180
181. Ethical Journalism. *The New York Times*. 5 Jan 2018. Available: <https://www.nytimes.com/editorial-standards/ethical-journalism.html>. Accessed 11 Aug 2021.

182. Cheng Y. Tesla vs. its Stock Price: “Herd Theory” at work? [cited 11 Aug 2021]. Available: <https://www.subr.edu/assets/subr/COBJournal/Tesla--Cheng-and-Griffin--E-Journal.pdf>
183. Michaels D, Elliott R. Tesla failed to oversee Elon Musk’s tweets, SEC argued in letters. Wall Street journal (Eastern ed). 1 Jun 2021. Available: <https://www.wsj.com/articles/tesla-failed-to-oversee-elon-musk-s-tweets-sec-argued-in-letters-11622582765>. Accessed 11 Aug 2021.
184. Thomson G. Primal arrogance and greed: Elon Musk slammed for crypto manipulation. In: Cointelegraph [Internet]. 9 Jun 2021 [cited 11 Aug 2021]. Available: <https://cointelegraph.com/news/primal-arrogance-and-greed-elon-musk-slammed-for-crypto-manipulation>
185. futurewise research. Genealogy Products and Services Market By Category, By Products & Services and By Region: Global Industry Analysis, Market Share, Revenue Opportunity, Competition and Forecast 2020-2027. 2020. Report No.: HC-1137. Available: <https://www.futurewiseresearch.com/healthcare-market-research/Genealogy-Products-and/1273>
186. prnewswire.com. Genealogy Products and Services Gaining Traction in Forensics. In: prnewswire.com [Internet]. 16 Apr 2019 [cited 10 May 2021]. Available: <https://www.prnewswire.com/news-releases/genealogy-products-and-services-gaining-traction-in-forensics-thereby-impacting-market-growth-at-11-4-cagr-during-2018-2024--details-factmr-study-300832714.html>
187. Regalado A. More than 26 million people have taken an at-home ancestry test. MIT Technology Review. 11 Feb 2019. Available: <https://www.technologyreview.com/2019/02/11/103446/more-than-26-million-people-have-taken-an-at-home-ancestry-test/>. Accessed 10 May 2021.
188. Hughes K. In Covid Lockdown, the Family Historian Has a Captive Audience. In: WSJ Online [Internet]. 14 Jul 2020 [cited 10 May 2021]. Available: <https://www.wsj.com/articles/covid-coronavirus-lockdown-genealogy-family-tree-history-ancestry-11594659103>
189. Rubinstein W d. History and “amateur” history. In: Lambert P, Schofield P, editors. Making History: An Introduction to the History and Practices of a Discipline. Routledge; 2004.
190. Lambert P, Schofield P. Making History: An Introduction to the History and Practices of a Discipline. Routledge; 2004.
191. ancestry.com. Ancestry.com Company Facts. In: Ancestry.com [Internet]. [cited 10 May 2021]. Available: <https://www.ancestry.com/corporate/about-ancestry/company-facts>

192. myheritage.com. Free Family Tree, Genealogy, Family History, and DNA Testing. In: myheritage.com [Internet]. [cited 13 May 2021]. Available: <https://www.myheritage.com/>
193. Worthington R. Dealing with Information Overload. Youtube; 2015 May 2; Washington DC Family History Center. Available: <https://www.youtube.com/watch?v=XYLgdmNO0PI>
194. Bidlack RE. Genealogy today. 1983. Available: [http://www.ideals.illinois.edu/bitstream/handle/2142/7294/librarytrendsv32i1c\\_opt.pdf](http://www.ideals.illinois.edu/bitstream/handle/2142/7294/librarytrendsv32i1c_opt.pdf)
195. Darby P, Clough P. Investigating the information-seeking behaviour of genealogists and family historians. *J Inf Sci Eng.* 2013;39: 73–84.
196. Agirre E, de Lacalle OL, Clough P, Stevenson M. PATHS: Personalised Access To cultural Heritage Spaces. *researchgate.net*. 2012. Available: [https://www.researchgate.net/profile/Eneko\\_Agirre/publication/228359402\\_PATHS\\_Personalised\\_Access\\_To\\_cultural\\_Heritage\\_Spaces/links/02bfe50e3412d372f200000/PATHS-Personalised-Access-To-cultural-Heritage-Spaces.pdf](https://www.researchgate.net/profile/Eneko_Agirre/publication/228359402_PATHS_Personalised_Access_To_cultural_Heritage_Spaces/links/02bfe50e3412d372f200000/PATHS-Personalised-Access-To-cultural-Heritage-Spaces.pdf)
197. Clough P, Sanderson M, Tang J. Examining the limits of crowdsourcing for relevance assessment. *IEEE Internet Comput.* 2012. Available: <https://ieeexplore.ieee.org/abstract/document/6231614/>
198. Minelli S, Marlow J, Clough P. Gathering requirements for multilingual search of audiovisual material in cultural heritage. *Citeseer*. 2007. Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.364.137&rep=rep1&type=pdf>
199. Wolinsky H. Genetic genealogy goes global. Although useful in investigating ancestry, the application of genetics to traditional genealogy could be abused. *EMBO Rep.* 2006;7: 1072–1074.
200. ancestry.com. Ancestry Communities Tool. In: ancestry.com [Internet]. 2021 [cited 19 May 2021]. Available: <https://support.ancestry.com/s/article/DNA-Genetic-Communities>
201. Reid K. How Ancestry.com's Find A Grave Encourages Bad Actors and Bad Data. 5 Aug 2019. Available: <https://onezero.medium.com/the-digital-under-takers-inside-the-obsessive-community-racing-to-put-the-dead-online-f3199147e968>. Accessed 10 May 2021.
202. Warner M. The Past and Future of the Intelligence Cycle. In: Phythian M, editor. *Understanding the Intelligence Cycle*. Routledge; 2013.

203. U.S. Joint Chiefs of Staff. Joint Publication 2-01: Joint and National Intelligence Support to Military Operations. Washington, DC: Government Printing Office; 2017.
204. Ancestry.com. Getting Started on Ancestry.com. In: Ancestry.com [Internet]. 2021 [cited 26 May 2021]. Available: <https://support.ancestry.com/s/article/Getting-Started-Lesson-1-Starting-Your-Tree>
205. Wei L. Tracking User Behavior with Big Data: A Model for Detecting Pain Points in the User Experience. Focus on Asia. 2017 [cited 13 Jan 2021]. Available: <https://uxpamagazine.org/tracking-user-behavior/>
206. Sternberg RJ. Toward a unified theory of human reasoning. *Intelligence*. 1986. pp. 281–314. doi:10.1016/0160-2896(86)90001-2
207. Biermann J. A Knowledge-Based Approach to Information Fusion for the Support of Military Intelligence. MILITARY ACADEMY IN BRNO (CZECH REPUBLIC) COMMAND AND STAFF FACULTY; 2004 Mar. Available: <https://apps.dtic.mil/sti/citations/ADA428705>
208. Waltz E. Quantitative Intelligence Analysis: Applied Analytic Models, Simulations, and Games. Illustrated edition. Rowman & Littlefield Publishers; 2014.
209. Cowan C. How To Use Those Shaky Leaf Hints. {ancestry.com}; 2016 Aug 23. Available: <https://www.youtube.com/watch?v=bsQp6RWjLg>
210. digitalclassicist.org. Digital tools for historians - Recommendations. [cited 19 May 2021]. Available: [https://wiki.digitalclassicist.org/Digital\\_tools\\_for\\_historians](https://wiki.digitalclassicist.org/Digital_tools_for_historians)
211. doinghistoryinpublic.org. Top 3 Digital Tools for Doing History. 10 Nov 2020 [cited 19 May 2021]. Available: <https://doinghistoryinpublic.org/2020/11/10/top-3-digital-tools-for-doing-history/>
212. ProgrammingHistorian.org. The Programming Historian. [cited 19 May 2021]. Available: <https://programminghistorian.org/en/>
213. history.ac.uk. History Online - Tools for Historians. [cited 19 May 2021]. Available: <https://www.history.ac.uk/history-online>
214. Cunningham S. Archive Skills and Tools for Historians. In: archives.history.ac.uk [Internet]. [cited 26 May 2021]. Available: [https://archives.history.ac.uk/makinghistory/resources/articles/archive\\_skills\\_and\\_tools\\_for\\_historians.html](https://archives.history.ac.uk/makinghistory/resources/articles/archive_skills_and_tools_for_historians.html)
215. Horowitz D. Ask The Expert Source Citations on the MyHeritage Website. myheritage.com; 2020 Jul 9. Available: <https://www.youtube.com/watch?v=gxRQMN-0Nbo>

216. ancestry.com. Managing Sources in Trees. In: ancestry.com [Internet]. [cited 19 May 2021]. Available: <https://support.ancestry.com/s/article/Managing-Sources-in-Trees>
217. Oulanov A, Pajarillo EJY. Academic librarians' perception of Lexis-Nexis. *The Electronic Library*. 2003;21: 123–129.
218. city-data.com. Ancestry tree - How to record multiple aliases. In: city-data.com [Internet]. Nov 2017 [cited 26 May 2021]. Available: <http://www.city-data.com/forum/genealogy/2849507-ancestry-tree-how-record-multiple-aliases.html>
219. Pease A. *Ontology: A Practical Guide*. Articulate Software Press; 2011.
220. Ball CA, Barber MJ, Byrnes JK, Carbonetto P, Chahine KG, Curtis DB, et al. DNA Circles. {ancestry.ca}; 2014. Available: <https://www.ancestry.ca/dna/resource/whitePaper/AncestryDNA-DNA-Circles-White-Paper.pdf>
221. Putman AL, Cole KL. All hail DNA: the constitutive rhetoric of AncestryDNATM advertising. *Critical Studies in Media Communication*. 2020;37: 207–220.
222. Charland M. Constitutive rhetoric: The case of the peuple québécois. *Q J Speech*. 1987;73: 133–150.
223. Tate H. The Ideological Effects of a Failed Constitutive Rhetoric: The Co-option of the Rhetoric of White Lesbian Feminism. *Women's Studies in Communication*. 2005;28: 1–31.
224. City-Data User Handle: MaryleeIII. Now I understand why some ancestry.com trees are private. In: city-data.com [Internet]. 2013 [cited 10 May 2021]. Available: <http://www.city-data.com/forum/genealogy/875430-now-i-understand-why-some-ancestry-7.html>
225. RootsChat User Handle: mrs.tenacious. Private Tree on Ancestry Photos Copied. In: RootsChat.Com [Internet]. 11 Aug 2013 [cited 10 May 2021]. Available: <https://www.rootschat.com/forum/index.php?topic=657263.0>
226. Family History Daily. The Huge Genealogy Mistake We All Need to Stop Making Now. In: familyhistorydaily.com [Internet]. 17 Oct 2015 [cited 10 May 2021]. Available: <https://familyhistorydaily.com/genealogy-help-and-how-to/the-huge-genealogy-mistake-we-all-need-to-stop-now/>
227. Roberts GHL, Park DS, Coignet MV, McCurdy SR, Knight SC, Partha R, et al. AncestryDNA COVID-19 Host Genetic Study Identifies Three Novel Loci. *medRxiv*.



2020. Available:  
<https://www.medrxiv.org/content/10.1101/2020.10.06.20205864v1.abstract>
228. Villa R, Clough PD, Hall MM, Rutter SA. Search or Browse? Casual Information Access to a Cultural Heritage Collection. Citeseer. 2013. Available:  
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.401.8178&rep=rep1&type=pdf>
229. Fulton C. The pleasure principle: the power of positive affect in information seeking. *Aslib Proceedings*. 2009;39: 88.
230. ancestry.com. What Can I Do With All My Family History Research? In: ancestry.com [Internet]. 2020 [cited 29 May 2021]. Available:  
<https://www.ancestry.com/corporate/blog/what-can-i-do-with-all-my-family-history-research>
231. Dyer N. How to Prepare DNA Research Reports for Sharing. In: familylocket.com [Internet]. 7 Nov 2020 [cited 29 May 2021]. Available:  
<https://familylocket.com/how-to-prepare-dna-research-reports-for-sharing/>
232. Hulnick AS. Intelligence Theory: Seeking better models. In: Phythian M, editor. *Understanding the Intelligence Cycle*. Routledge; 2013.
233. Gill P, Phythian M. From Intelligence Cycle to Web of Intelligence: Complexity and the conceptualisation of intelligence. In: Phythian M, editor. *Understanding the Intelligence Cycle*. Routledge; 2013.
234. Hribar G, Podbregar I, Ivanuša T. OSINT: A “Grey Zone”? *Int J Intell CounterIntelligence*. 2014;27: 529–549.
235. Johnson LK. Making the intelligence “Cycle” work. *Int J Intell CounterIntelligence*. 1986;1: 1–23.
236. Frini A, Boury-Brisset A-C. An intelligence process model based on a collaborative approach. DEFENCE RESEARCH AND DEVELOPMENT CANADA VALCARTIER (QUEBEC); 2011 Jun. Available: <https://apps.dtic.mil/sti/citations/ADA547105>
237. Tam L, Glassman M, Vandenwauver M. The psychology of password management: a tradeoff between security and convenience. *Behav Inf Technol*. 2010;29: 233–244.
238. Mendel T, Toch E. My Mom was Getting this Popup: Understanding Motivations and Processes in Helping Older Relatives with Mobile Security and Privacy. *Proc ACM Interact Mob Wearable Ubiquitous Technol*. 2019;3: 1–20.



239. Gaw S, Felten EW. Password management strategies for online accounts. Proceedings of the second symposium on Usable privacy and security. New York, NY, USA: Association for Computing Machinery; 2006. pp. 44–55.
240. Geng H. Data Center Handbook. John Wiley & Sons; 2014.
241. Benson T, Anand A, Akella A, Zhang M. Understanding data center traffic characteristics. Proceedings of the 1st ACM workshop on Research on enterprise networking. New York, NY, USA: Association for Computing Machinery; 2009. pp. 65–72.
242. Gong Z, Zhu J-W, Li C-P, Jiang S, Ma L-N, Tang B-X, et al. An online coronavirus analysis platform from the National Genomics Data Center. Zool Res. 2020;41: 705–708.
243. Miller HE, Engemann KJ. Business Continuity Management in Data Center Environments. IJITSA. 2019;12: 52–72.
244. Mell P, Grance T, Others. The NIST definition of cloud computing. 2011. Available: <http://faculty.winthrop.edu/domanm/csci411/Handouts/NIST.pdf>
245. Breuer PC. Reducing Information Overload via an Analog Model for Cyber Risk. Boger DC, editor. Doctor of Philosophy in Information Sciences, Naval Postgraduate School. 2020. Available: <https://calhoun.nps.edu/handle/10945/65458>
246. Veeramachaneni K, Arnaldo I, Korrapati V, Bassias C, Li K. AI<sup>2</sup>: training a big data machine to defend. 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS). IEEE; 2016. pp. 49–54.
247. De Domenico A, Liu Y-F, Yu W. Optimal Virtual Network Function Deployment for 5G Network Slicing in a Hybrid Cloud Infrastructure. IEEE Trans Wireless Commun. 2020;19: 7942–7956.
248. Javadi B, Abawajy J, Buyya R. Failure-aware resource provisioning for hybrid Cloud infrastructure. J Parallel Distrib Comput. 2012;72: 1318–1331.
249. Dhirani LL, Newe T. Hybrid Cloud SLAs for Industry 4.0: Bridging the Gap. Annals of Emerging Technologies in. 2020. Available: <https://papers.ssrn.com/abstract=3785455>
250. Meadows D. Leverage Points: Places to Intervene in a System - The Donella Meadows Project. 19 Oct 1999 [cited 4 Aug 2021]. Available:

- <https://donellameadows.org/archives/leverage-points-places-to-intervene-in-a-system/>
251. Pauwels E. The New Geopolitics of Converging Risks. UN Centre for Policy Research; 2019. Available: <http://collections.unu.edu/eserv/UNU:7308/PauwelsAlGeopolitics.pdf>
252. Halkort M. The Geopolitics of the Cloud. ror-n.org; 2013. Available: <http://www.ror-n.org/-blog/the-geopolitics-of-the-cloud>
253. Roguski P. The Geopolitics of Cloud Computing. ruj.uj.edu.pl; 2020 [cited 4 Aug 2021]. Available: [https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/262562/roguski\\_the\\_geopolitics\\_of\\_cloud\\_computing\\_2020.pdf?sequence=1&isAllowed=y](https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/262562/roguski_the_geopolitics_of_cloud_computing_2020.pdf?sequence=1&isAllowed=y)
254. Aliapoulios M, Bevensee E, Blackburn J, Bradlyn B, De Cristofaro E, Stringhini G, et al. An Early Look at the Parler Online Social Network. arXiv [cs.SI]. 2021. Available: <http://arxiv.org/abs/2101.03820>
255. Cordes RJ, Friedman DA. Emergent Teams for Complex Threats. 2020. Available: [https://www.remotorconsulting.com/uploads/4/8/4/2/48428829/emergent\\_teams\\_for\\_complex\\_threats\\_\\_1\\_.pdf](https://www.remotorconsulting.com/uploads/4/8/4/2/48428829/emergent_teams_for_complex_threats__1_.pdf)
256. Atlantic Council GeoTech Center. Report of the Commission on the Geopolitical Impacts of New Technologies and Data. 2021 May. Available: <https://www.atlanticcouncil.org/content-series/geotech-commission/exec-summary/>
257. Colomo-Palacios R, Fernandes E, Sabbagh M, de Amescua Seco A. Human and intellectual capital management in the cloud: software vendor perspective. J Univers Comput Sci. 2012;18: 1544–1557.
258. Baumann FW, Roller D. Additive Manufacturing, Cloud-Based 3D Printing and Associated Services—Overview. J Mater Process Manuf Sci. 2017;1: 15.
259. Fisher O, Watson N, Porcu L, Bacon D, Rigley M, Gomes RL. Cloud manufacturing as a sustainable process manufacturing route. Journal of Manufacturing Systems. 2018;47: 53–68.
260. Turton W, Mehrotra K. Hackers Breached Colonial Pipeline Using Compromised Password. Bloomberg News. 4 Jun 2021. Available: <https://www.bloomberg.com>. Accessed 18 Aug 2021.
261. Greenberg A. The Full Story of the Stunning RSA Hack Can Finally Be Told. Wired. 20 May 2021. Available: <https://www.wired.com/story/the-full-story-of-the-stunning-rsa-hack-can-finally-be-told/>. Accessed 23 Aug 2021.

262. Kanipe CR, Yeo DW. A Wargame for Introducing Cybersecurity Considerations to First-year Engineering Students. AIAA Scitech 2019 Forum. American Institute of Aeronautics and Astronautics; 2019.
263. Russo L, Binaschi F, De Angelis A. Cybersecurity exercises: Wargaming and red teaming. *Next Generation CERTs*. 2019;54: 44.
264. Moore CJ. Preparing the Virtual Battlefield for war: A cyber threat “survival kit” for commanders. Naval War College; 2004 May. Available: <https://apps.dtic.mil/sti/citations/ADA426069>
265. Rieb A, Lechner U. Operation Digital Chameleon: Towards an Open Cybersecurity Method. *Proceedings of the 12th International Symposium on Open Collaboration*. New York, NY, USA: Association for Computing Machinery; 2016. pp. 1–10.
266. Abu-Alhaija M. Cyber Security: Between Challenges and Prospects. *ICIC express letters Part B, Applications: an international journal of research and surveys*. 2020;11: 1019–1028.
267. Sandhu R. Speculations on the science of web user security. *Computer Networks*. 2012;56: 3891–3895.
268. Fenrich K. Securing your control system: the “CIA triad” is a widely used benchmark for evaluating information system security effectiveness. *Power engineering*. Feb 2008;112: 44+.
269. Elifoglu IH, Abel I, Taşseven Ö. Minimizing insider threat risk with behavioral monitoring. *Review of business*. 2018;38: 61–73.
270. Wang S, Liu Y, Wilkins-Diehr N, Martin S. SimpleGrid Toolkit: Enabling Efficient Learning and Development of TeraGrid Science Gateway. *International Workshop on Grid Computing Environments*. academia.edu; 2007. Available: [https://www.academia.edu/download/42971209/SimpleGrid\\_toolkit\\_Enabling\\_efficient\\_le20160223-13548-umtggz.pdf](https://www.academia.edu/download/42971209/SimpleGrid_toolkit_Enabling_efficient_le20160223-13548-umtggz.pdf)
271. Gesing S, Others. Hands-On Tutorial on Grid Portal development focused on Biomedical Image and Signal Processing. *HealthGrid Applications and Technologies Meet Science Gateways for Life Sciences*. 2012;175: 104.
272. Wang S, Liu Y, Wilkins-Diehr N, Martin S. SimpleGrid toolkit: Enabling geosciences gateways to cyberinfrastructure. *Comput Geosci*. 2009;35: 2283–2294.
273. Nazir S, Patel S, Patel D. Assessing and augmenting SCADA cyber security: A survey of techniques. *Comput Secur*. 2017;70: 436–454.

274. Sivaraman V, Gharakheili HH, Fernandes C, Clark N, Karliychuk T. Smart IoT Devices in the Home: Security and Privacy Implications. *IEEE Technol Soc Mag*. 2018;37: 71–79.
275. Kephart JO, Chess DM. The vision of autonomic computing. *Computer* . 2003;36: 41–50.
276. alz.org. Alzheimer's Facts and Figures. In: alz.org [Internet]. 2021 [cited 24 Aug 2021]. Available: <https://www.alz.org/alzheimers-dementia/facts-figures>
277. Liu P-P, Xie Y, Meng X-Y, Kang J-S. History and progress of hypotheses and clinical trials for Alzheimer's disease. *Signal Transduct Target Ther*. 2019;4: 29.
278. Anderson RM, Hadjichrysanthou C, Evans S, Wong MM. Why do so many clinical trials of therapies for Alzheimer's disease fail? *Lancet*. 2017;390: 2327–2329.
279. Yiannopoulou KG, Anastasiou AI, Zachariou V, Pelidou S-H. Reasons for Failed Trials of Disease-Modifying Treatments for Alzheimer Disease and Their Contribution in Recent Research. *Biomedicines*. 2019;7. doi:10.3390/biomedicines7040097
280. Penzes P, Cahill ME, Jones KA, VanLeeuwen J-E, Woolfrey KM. Dendritic spine pathology in neuropsychiatric disorders. *Nat Neurosci*. 2011;14: 285–293.
281. Sasaguri H, Nilsson P, Hashimoto S, Nagata K, Saito T, De Strooper B, et al. APP mouse models for Alzheimer's disease preclinical studies. *EMBO J*. 2017;36: 2473–2487.
282. Bales KR. The value and limitations of transgenic mouse models used in drug discovery for Alzheimer's disease: an update. *Expert Opin Drug Discov*. 2012;7: 281–297.
283. Reardon S. Frustrated Alzheimer's researchers seek better lab mice. *Nature*. 2018;563: 611–612.
284. Oxford AE, Stewart ES, Rohn TT. Clinical Trials in Alzheimer's Disease: A Hurdle in the Path of Remedy. Hindawi Limited; 2020 [cited 16 Aug 2021]. doi:10.1155/2020/5380346
285. Das R. Are Direct-To-Consumer Ads For Drugs Doing More Harm than Good? *Forbes Magazine*. 2019 [cited 16 Aug 2021]. Available: <https://www.forbes.com/sites/reenitadas/2019/05/14/direct-to-consumer-drug-ads-are-they-doing-more-harm-than-good/?sh=6bc831ce4dfc>
286. Triunfol M, Gouveia FC. What's not in the news headlines or titles of Alzheimer disease articles? #InMice. *PLoS Biol*. 2021;19: e3001260.

287. Hamilton J. Future Alzheimer's Treatments Aim To Do More Than Clear Plaques From The Brain. NPR. 9 Aug 2021. Available: <https://www.npr.org/sections/health-shots/2021/08/09/1026207499/new-treatments-look-for-ways-to-protect-the-brain-from-alzheimers>. Accessed 16 Aug 2021.
288. Shen Y, Li R. What do we know from clinical trials on exercise and Alzheimer's disease? *J Sport Health Sci*. 2016;5: 397–399.
289. Watson C. Health researchers report funder pressure to suppress results. *Nature*. 2021. doi:10.1038/d41586-021-02242-x
290. Rothenberg KH, Bush LW. *The Drama of DNA: Narrative Genomics*. 1st ed. The Drama of DNA: Narrative Genomics. 1st ed. Oxford University Press;
291. Hurst M. Narrative Genetics: Following the Trail of Spit. In: *Narrative Genetics: Following the Trail of Spita* [Internet]. [cited 26 May 2021]. Available: <https://medhum.med.nyu.edu/magazine/?p=147>
292. Bush LW. Narrative Genomics: Creating a Stage for Inquiry and Bioethics Education. *J Health Care L & Pol'y*. 2019;22: 177.
293. Tasaki KM. Circular causality in integrative multi-scale systems biology and its interaction with traditional medicine. *Prog Biophys Mol Biol*. 2013;111: 144–146.
294. Noble R, Tasaki K, Noble PJ, Noble D. Biological Relativity Requires Circular Causality but Not Symmetry of Causation: So, Where, What and When Are the Boundaries? *Front Physiol*. 2019;10: 827.
295. GenBank and WGS Statistics. [cited 11 Aug 2021]. Available: <https://www.ncbi.nlm.nih.gov/genbank/statistics/>
296. Milestones in Genomic Sequencing. [cited 11 Aug 2021]. Available: <https://www.nature.com/articles/d42859-020-00099-0>
297. Genetic Information Discrimination: The Evolving Meaning of the Genetic Information Nondiscrimination Act. Bureau of National Affairs; 2008.
298. Roberts JL. Preempting discrimination: Lessons from the genetic information nondiscrimination act. *Vanderbilt Law Rev*. 2010;63: 439.
299. Longino HE. *Studying Human Behavior: How Scientists Investigate Aggression and Sexuality*. University of Chicago Press; 2013.
300. Panofsky A. Rethinking scientific authority: Behavior genetics and race controversies. *American Journal of Cultural Sociology*. 2018;6: 322–358.

301. Johnson KA. Research, Race and Profit. In: Johnson KA, editor. *Medical Stigmata: Race, Medicine, and the Pursuit of Theological Liberation*. Singapore: Springer Singapore; 2019. pp. 73–124.
302. Jablonka E, Lamb MJ. Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life. *Life and mind*. 2005;462. Available: <https://psycnet.apa.org/fulltext/2005-04046-000.pdf>
303. Keller EF. *The Mirage of a Space between Nature and Nurture*. Duke University Press; 2010.
304. Cyranoski D. What CRISPR-baby prison sentences mean for research. *Nature*. 2020;577: 154–155.
305. Chamary JV. How Genetic Genealogy Helped Catch The Golden State Killer. *Forbes Magazine*. 30 Jun 2020. Available: <https://www.forbes.com/sites/jvchamary/2020/06/30/genetic-genealogy-golden-state-killer/>. Accessed 11 Aug 2021.
306. Hurst M, Lieber C, Lewis LJ, Grob R. Family stories: narrative genetics and conceptions of heritability in pregnant women. *J Midwifery Womens Health*. 2011;56: 26–32.
307. Spiers J, Smith JA, Ferrer-Duch M, Moldovan R, Roche J, MacLeod R. Evaluating a genetic counseling narrative group session for people who have tested positive for the Huntington's disease expansion: An interpretative phenomenological analysis. *J Genet Couns*. 2020;29: 1015–1025.
308. Snell K, Helén I. “Well, I knew this already” - explaining personal genetic risk information through narrative meaning-making. *Sociol Health Illn*. 2020;42: 496–509.
309. Hirst R. Scientific Jargon, Good and Bad. *Journal of Technical Writing and Communication*. 2003;33: 201–229

## Chapter II Works Cited

---

1. Benkler Y, Faris R, Roberts H. Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics. Oxford University Press; 2018.
2. Rowell L, Call-Cummings M. Knowledge Democracy, Action Research, the Internet and the Epistemic Crisis. *Journal of Futures Studies*. 2020;24: 73–81.
3. Miller M, Kirwan J. The Modern Origins of Our Epistemic Crisis: A Media & Democracy Workshop. 7 May 2019 [cited 24 Jun 2021]. Available: <https://items.ssrc.org/from-our-programs/the-modern-origins-of-our-epistemic-crisis-a-media-democracy-workshop/>
4. Firth J, Torous J, Stubbs B, Firth JA, Steiner GZ, Smith L, et al. The “online brain”: how the Internet may be changing our cognition. *World Psychiatry*. 2019;18: 119–129.
5. Menczer F, Hills T. Information Overload Helps Fake News Spread, and Social Media Knows It. *Scientific American*. doi:10.1038/scientificamerican1220-54
6. Morris JD, Woo C, Singh AJ. Elaboration likelihood model: A missing intrinsic emotional implication. *Journal of Targeting, Measurement and Analysis for Marketing*. 2005;14: 79–98.
7. SanJosé-Cabezudo R, Gutiérrez-Arranz AM, Gutiérrez-Cillán J. The combined influence of central and peripheral routes in the online persuasion process. *Cyberpsychol Behav*. 2009;12: 299–308.
8. Dunbar NE, Connelly S, Jensen ML, Adame BJ, Rozzell B, Griffith JA. Fear Appeals, Message Processing Cues, and Credibility in the Websites of Violent, Ideological, and Nonideological Groups. *J Comput Mediat Commun*. 2014;19: 871–889.

9. Tindale CW. Replicating Reasons: Arguments, Memes, and the Cognitive Environment. *Philosophy & Rhetoric*. 2017;50: 566–588.
10. Denisova A. Meme. [cited 6 Aug 2021]. Available: <https://sk.sagepub.com/reference/the-sage-encyclopedia-of-mass-media-and-society/i10854.xml>
11. Shifman L. Memes in a Digital World: Reconciling with a Conceptual Troublemaker. *J Comput Mediat Commun*. 2013;18: 362–377.
12. Shifman L. The Cultural Logic of Photo-Based Meme Genres. *Journal of Visual Culture*. 2014;13: 340–358.
13. Catchphrases. 7 Mar 2012 [cited 24 Jun 2021]. Available: <https://knowyourmeme.com/memes/cultures/catchphrases>
14. Taecharungroj V, Nueangjamnong P. Humour 2.0: Styles and Types of Humour and Virality of Memes on Facebook. *Journal of Creative Communications*. 2015;10: 288–302.
15. Guenther L, Ruhrmann G, Bischoff J, Penzel T, Weber A. Strategic Framing and Social Media Engagement: Analyzing Memes Posted by the German Identitarian Movement on Facebook. *Social Media + Society*. 2020;6: 2056305119898777.
16. Facebook: From Election to Insurrection. [cited 11 Aug 2021]. Available: [https://secure.avaaz.org/campaign/en/facebook\\_election\\_insurrection/?slideshow](https://secure.avaaz.org/campaign/en/facebook_election_insurrection/?slideshow)
17. Mascarenhas M. Memes as Quasi-Argument. *Local Theories of Argument*. 2021. pp. 385–390. doi:10.4324/9781003149026-65
18. Rapp C. Aristotle's Rhetoric. Zalta EN, editor. *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University; 2010. Available: <https://plato.stanford.edu/archives/spr2010/entries/aristotle-rhetoric/>
19. Harrell J, Linkugel WA. On Rhetorical Genre: An Organizing Perspective. *Philosophy & Rhetoric*. 1978;11: 262–281.
20. Heiskanen B. Meme-ing electoral participation. *Eur J Am Stud*. 2017;12. doi:10.4000/ejas.12158
21. Hakoköngäs E, Halmesvaara O, Sakki I. Persuasion Through Bitter Humor: Multimodal Discourse Analysis of Rhetoric in Internet Memes of Two Far-Right Groups in Finland. *Social Media + Society*. 2020;6: 2056305120921575.
22. Huntington HE. Affect and effect of Internet memes: assessing perceptions and influence of online user-generated political discourse as media, The. *Colorado*



- State University. 2017. Available:  
[https://mountainscholar.org/bitstream/handle/10217/183936/Huntington\\_colostate\\_0053A\\_14303.pdf](https://mountainscholar.org/bitstream/handle/10217/183936/Huntington_colostate_0053A_14303.pdf)
23. Ling C, AbuHilal I, Blackburn J, De Cristofaro E, Zannettou S, Stringhini G. Dissecting the Meme Magic: Understanding Indicators of Virality in Image Memes. *Proc ACM Hum-Comput Interact.* 2021;5: 1–24.
  24. Charland M. Constitutive rhetoric: The case of the peuple québécois. *Q J Speech.* 1987;73: 133–150.
  25. Kofman A. Bruno Latour, the Post-Truth Philosopher, Mounts a Defense of Science. *The New York Times.* 25 Oct 2018. Available:  
<https://www.nytimes.com/2018/10/25/magazine/bruno-latour-post-truth-philosopher-science.html>. Accessed 11 Aug 2021.
  26. Ceccarelli L. Polysemy: Multiple meanings in rhetorical criticism. *Q J Speech.* 1998;84: 395–415.
  27. Wikipedia contributors. *Condescending Wonka*. In: *Wikipedia, The Free Encyclopedia* [Internet]. 1 Jun 2021. Available:  
[https://en.wikipedia.org/w/index.php?title=Condescending\\_Wonka&oldid=1026354342](https://en.wikipedia.org/w/index.php?title=Condescending_Wonka&oldid=1026354342)
  28. Beveridge A, Chemers M. *The Game of Game of Thrones : Networked Concordances and Fractal Dramaturgy*. Reading Contemporary Serial Television Universes. Routledge; 2018. pp. 201–225.
  29. Hybrid ecosystem of narratives. 12 Apr 2009 [cited 25 Mar 2021]. Available:  
<https://tihane.wordpress.com/2009/04/12/hybrid-ecosystem-of-narratives/>
  30. Druschke CG. A Trophic Future for Rhetorical Ecologies. *Enculturation: a journal of rhetoric, writing, and culture.* 2019 [cited 24 Jun 2021]. Available:  
<http://enculturation.net/a-trophic-future>
  31. Fiaidhi J, Mohammed S. Virtual care for cyber–physical systems (VH\_CPS): NODE-RED, community of practice and thick data analytics ecosystem. *Comput Commun.* 2021;170: 84–94.
  32. Wang H, Liu Z, Guo Y, Chen X, Zhang M, Xu G, et al. An Explorative Study of the Mobile App Ecosystem from App Developers' Perspective. *Proceedings of the 26th International Conference on World Wide Web*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee; 2017. pp. 163–172.

33. Fiorina C. HP Carly Fiorina Speech: The Digital Ecosystem. [cited 27 Apr 2021]. Available:  
[http://www.hp.com/hpinfo/execteam/speeches/fiorina/ceo\\_worldres\\_00.html](http://www.hp.com/hpinfo/execteam/speeches/fiorina/ceo_worldres_00.html)
34. Games SJ. Transhuman Space Toxic Memes. First Ed edition. Steve Jackson Games, Incorporated; 2004.
35. Ford T, Krohn R, Weninger T. Competition Dynamics in the Meme Ecosystem. arXiv [cs.SI]. 2021. Available: <http://arxiv.org/abs/2102.03952>
36. Zannettou S, Caulfield T, Blackburn J, De Cristofaro E, Sirivianos M, Stringhini G, et al. On the Origins of Memes by Means of Fringe Web Communities. arXiv [cs.SI]. 2018. Available: <http://arxiv.org/abs/1805.12512>
37. Druschke CG, McGreavy B. Why rhetoric matters for ecology. *Front Ecol Environ*. 2016;14: 46–52.
38. Friedman D, Tschantz A, Ramstead MJD, Friston K, Constant A. Active inferants: The basis for an active inference framework for ant colony behavior. *Front Behav Neurosci*. 2021;15: 126.
39. Edbauer J. Unframing Models of Public Distribution: From Rhetorical Situation to Rhetorical Ecologies. *Rhetor Soc Q*. 2005;35: 5–24.
40. Equihua M, Espinosa Aldama M, Gershenson C, López-Corona O, Munguía M, Pérez-Maqueo O, et al. Ecosystem antifragility: beyond integrity and resilience. *PeerJ*. 2020;8: e8533.
41. Marcellino W, Helmus TC, Kerrigan J. Detecting Conspiracy Theories on Social Media: Improving Machine Learning to Detect and Understand Online Conspiracy Theories. RAND Corporation; 2021.
42. Cordes RJ, Friedman DA. Reimagining Maps. *The Great Preset: Remote Teams and Operational Art*. COGSEC; 2020.
43. Ramstead MJD, Constant A, Badcock PB, Friston KJ. Variational ecology and the physics of sentient systems. *Phys Life Rev*. 2019. doi:10.1016/j.plrev.2018.12.002
44. Budaev S, Jørgensen C, Mangel M, Eliassen S, Giske J. Decision-Making From the Animal Perspective: Bridging Ecology and Subjective Cognition. *Frontiers in Ecology and Evolution*. 2019;7: 164.
45. Ratcliffe JM, Phelps SM. Twenty-five years of cognitive ecology. *Anim Behav*. 2019;147: 127–128.

46. Power DA, Watson RA, Szathmáry E, Mills R, Powers ST, Doncaster CP, et al. What can ecosystems learn? Expanding evolutionary ecology with learning theory. *Biol Direct*. 2015;10: 69.
47. Sharov AA. Evolutionary biosemiotics and multilevel construction networks. *Biosemiotics*. 2016;9: 399–416.
48. Gilbert SF. Ecological Developmental Biology: Interpreting Developmental Signs. *Biosemiotics*. 2016;9: 51–60.
49. Ruhl JB. Governing Cascade Failures in Complex Social-Ecological-Technological Systems: Framing Context, Strategies, and Challenges. 2019. Available: <https://papers.ssrn.com/abstract=3471945>
50. Cordes RJ, Friedman DA. Emergent Teams for Complex Threats. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
51. Zhao B. Why Processing is Not Swarm Intelligence. *Proceedings of the 2019 DigitalFUTURES*. Springer Singapore; 2020. pp. 257–264.
52. Xing X, Rongpeng L, Zhifeng Z, Honggang Z. Stigmergic Independent Reinforcement Learning for Multi-Agent Collaboration. *arXiv preprint arXiv*. 2019. Available: <https://arxiv.org/abs/1911.12504>
53. Finn KR, Silk MJ, Porter MA, Pinter-Wollman N. The use of multilayer network analysis in animal behaviour. *Anim Behav*. 2019;149: 7–22.
54. cadCAD – A Python package for designing, testing and validating complex systems through simulation. [cited 26 Mar 2020]. Available: <https://cadcad.org/>
55. Bradbury JW, Vehrencamp SL. Complexity and behavioral ecology. *Behav Ecol*. 2014. Available: <https://academic.oup.com/beheco/article-abstract/25/3/435/510920>
56. Leonhardt SD, Menzel F, Nehring V, Schmitt T. Ecology and Evolution of Communication in Social Insects. *Cell*. 2016;164: 1277–1287.
57. Gordon DM. Measuring collective behavior: an ecological approach. *Theory Biosci*. 2019. doi:10.1007/s12064-019-00302-5
58. Chambers JC, Allen CR, Cushman SA. Operationalizing Ecological Resilience Concepts for Managing Species and Ecosystems at Risk. *Frontiers in Ecology and Evolution*. 2019;7: 241.
59. Clua E, Beliaeff B, Chauvet C, David G, Ferraris J, Kronen M, et al. Towards multidisciplinary indicator dashboards for coral reef fisheries management. *Aquat Living Resour*. 2005;18: 199–213.

60. O'Brien A, Townsend K, Hale R, Sharley D, Pettigrove V. How is ecosystem health defined and measured? A critical review of freshwater and estuarine studies. *Ecol Indic.* 2016;69: 722–729.
61. Jørgensen S, Xu L, Costanza R. *Handbook of Ecological Indicators for Assessment of Ecosystem Health.* CRC Press; 2016.
62. De Garine-Wichatitsky M, Binot A, Ward J, Caron A, Perrotton A, Ross H, et al. "Health in" and "Health of" Social-Ecological Systems: A Practical Framework for the Management of Healthy and Resilient Agricultural and Natural Ecosystems. *Front Public Health.* 2020;8: 616328.
63. Reporting Bird Sightings. 25 Sep 2015 [cited 13 May 2021]. Available: <https://ny.audubon.org/about-us/reporting-bird-sightings>
64. National Geographic Society. BioBlitz Program. [cited 7 Oct 2021]. Available: <https://www.nationalgeographic.org/projects/bioblitz/>
65. Hood N, Littlejohn A. Hacking history: Redressing gender inequities on Wikipedia through an editathon. *Int Rev Res Open Distrib Learn.* 2018;19. doi:10.19173/irrodl.v19i5.3549
66. March L, Dasgupta S. Wikipedia Edit-a-thons as Sites of Public Pedagogy. *Proc ACM Hum-Comput Interact.* 2020;4: 1–26.
67. Zamora ME, Espinosa M, Gershenson C, López-Corona O, Munguia M, Pérez-Maqueo O, et al. Ecosystem antifragility: Beyond integrity and resilience. *PeerJ PrePrints.* 2019.
68. Falk DA, Watts AC, Thode AE. Scaling Ecological Resilience. *Frontiers in Ecology and Evolution.* 2019;7: 275.
69. Asllani M, Lambiotte R, Carletti T. Structure and dynamical behavior of non-normal networks. *Sci Adv.* 2018;4: eaau9403.
70. Bruder A, Frainer A, Rota T, Primicerio R. The Importance of Ecological Networks in Multiple-Stressor Research and Management. *Front Environ Sci Eng China.* 2019;7: 59.
71. Marshall JAR. Generalizations of Hamilton's rule applied to non-additive public goods games with random group size. *Frontiers in Ecology and Evolution.* 2014;2: 40.
72. Pangallo M, Heinrich T, Doyne Farmer J. Best reply structure and equilibrium convergence in generic games. *Sci Adv.* 2019;5: eaat1328.

73. Fisher B, Turner RK, Morling P. Defining and classifying ecosystem services for decision making. *Ecol Econ*. 2009;68: 643–653.
74. Gravel D, Guichard F, Loreau M, Mouquet N. Source and sink dynamics in meta-ecosystems. *Ecology*. 2010. pp. 2172–2184. doi:10.1890/09-0843.1
75. Sievanen L. How do small-scale fishers adapt to environmental variability? Lessons from Baja California, Sur, Mexico. *Marit Stud*. 2014;13: 9.
76. Brown ED, Williams BK. Resilience and Resource Management. *Environ Manage*. 2015;56: 1416–1427.
77. Tosa MI, Dziedzic EH, Appel CL, Urbina J, Massey A, Ruprecht J, et al. The Rapid Rise of Next-Generation Natural History. *Frontiers in Ecology and Evolution*. 2021;9: 480.
78. Newell E, Jurgens D, Saleem HM, Vala H, Sassine J, Armstrong C, et al. User Migration in Online Social Networks: A Case Study on Reddit During a Period of Community Unrest. Tenth International AAAI Conference on Web and Social Media. 2016. Available: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/viewPaper/13137>
79. Leurs K, Prabhakar M. Doing digital migration studies: Methodological considerations for an emerging research focus. *Qualitative research in European migration studies*. Springer, Cham; 2018. pp. 247–266.
80. Buhs JB. *The Fire Ant Wars: Nature, Science, and Public Policy in Twentieth-Century America*. University of Chicago Press; 2004.
81. Cordes RJ, David S, Maan A, Ruiz A, Sapp E, Scannell P, et al. *The Narrative Campaign Field Guide - First Edition*. 1st ed. Cordes RJ, editor. Narrative Strategies Ink; 2021.
82. Shine R. The ecological impact of invasive cane toads (*Bufo marinus*) in Australia. *Q Rev Biol*. 2010;85: 253–291.
83. Roman A, Dasgupta D, Pleimling M. A theoretical approach to understand spatial organization in complex ecologies. *J Theor Biol*. 2016;403: 10–16.
84. Jones ME, Apfelbach R, Banks PB, Cameron EZ, Dickman CR, Frank A, et al. A Nose for Death: Integrating Trophic and Informational Networks for Conservation and Management. *Frontiers in Ecology and Evolution*. 2016;4: 124.
85. LaFrance A. The Internet Is Mostly Bots. *The Atlantic*. 31 Jan 2017. Available: <https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043/>. Accessed 11 Aug 2021.

86. Zeifman I, McKeever G, Hewitt N, Hasson E. Bot Traffic Report 2016. Imperva; 31 Mar 2021 [cited 11 May 2021]. Available: <https://www.imperva.com/blog/bot-traffic-report-2016/?redirect=Incapsula>
87. Vyatkin A, Metelkin I, Mikhailova A, Cordes RJ, Friedman DA. Active Inference and Behavior Engineering for Teams. In: Friedman DA, Cordes RJ, editors. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
88. Mitterhofer S, Kruegel C, Kirda E, Platzer C. Server-Side Bot Detection in Massively Multiplayer Online Games. IEEE Security Privacy. 2009;7: 29–36.
89. Rheault L, Musulan A. Efficient detection of online communities and social bot activity during electoral campaigns. Journal of Information Technology & Politics. 2021;18: 324–337.
90. Rosenberg KV, Dokter AM, Blancher PJ, Sauer JR, Smith AC, Smith PA, et al. Decline of the North American avifauna. Science. 2019; eaaw1313.
91. Lafuite A-S, de Mazancourt C, Loreau M. Delayed behavioural shifts undermine the sustainability of social-ecological systems. Proc Biol Sci. 2017;284: 20171192.
92. Grasso F. Towards Computational Rhetoric. IL. 2002;22. doi:10.22329/il.v22i3.2589
93. Wojcik MW. Thesis: Inventing computational rhetoric. search.proquest.com. 2013. Available: <https://search.proquest.com/openview/594bb5e36be4a69803db5ef14dc7f830/1?pq-origsite=gscholar&cbl=18750>
94. Walsh L, Boyle C. Topologies as Techniques for a Post-Critical Rhetoric. Walsh L, Boyle C, editors. Palgrave Macmillan, Cham; 2017.
95. Cordes RJ. Games with serious impacts: The next generation of serious games - Atlantic Council. In: atlanticcouncil.org [Internet]. 21 May 2021 [cited 3 Jun 2021]. Available: <https://www.atlanticcouncil.org/blogs/geotech-cues/games-with-serious-impacts-the-next-generation/>
96. preferred reading. [cited 17 Sep 2021]. doi:10.1093/oi/authority.20110803100342875
97. Dietze MC, Fox A, Beck-Johnson LM, Betancourt JL, Hooten MB, Jarnevich CS, et al. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proc Natl Acad Sci U S A. 2018;115: 1424–1432.
98. Cordes RJ, David S, Maan A, Ruiz A, Sapp E, Scannell P, et al. The Narrative Campaign Field Guide. 1st ed. Cordes RJ, editor. Narrative Strategies Ink; 2021.

99. Bradshaw S, Howard PN. The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation. Project on Computational Propaganda; 2019.
100. Union of Concerned Scientists. The Disinformation Playbook. In: UCSUSA.org [Internet]. oct 10 2017 [cited 8 Jan 2021]. Available: <https://www.ucsusa.org/resources/disinformation-playbook>
101. DisinfoCloud. Disinfo Cloud. In: disinfocloud.com [Internet]. 22 Apr 2021 [cited 23 Apr 2021]. Available: <https://disinfocloud.com/>
102. Military Intelligence Professional Bulletin - Collection Management. {US Army Intelligence Center of Excellence}; 2020.
103. Varga S, Brynielsson J, Horndahl A, Rosell M. Automated Text Analysis for Intelligence Purposes: A Psychological Operations Case Study. In: Tayebi MA, Glässer U, Skillicorn DB, editors. Open Source Intelligence and Cyber Crime: Social Media Analytics. Cham: Springer International Publishing; 2020. pp. 221–251.
104. Christianson C, Duncan J, Onyshkevych B. Overview of the DARPA LORELEI Program. Machine Translation. 2018;32: 3–9.
105. Naderpour M, Lu J, Zhang G. A fuzzy dynamic bayesian network-based situation assessment approach. 2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). [ieeexplore.ieee.org](http://ieeexplore.ieee.org); 2013. pp. 1–8.
106. Winklerova Z. Ontological approach to the representation of military knowledge. MILITARY ACADEMY IN BRNO (CZECH REPUBLIC) COMMAND AND STAFF FACULTY; 2004 Mar. Available: <https://apps.dtic.mil/sti/citations/ADA428705>
107. Davis Z, Gac F, Rager C, Reiner P, Snow J, editors. Strategic Latency Unleashed: The Role of Technology in a Revisionist Global Order and the Implications for Special Operations Forces. Center for Global Security Research; 2021.
108. Rosen MA, Fiore SM, Salas E, Letsky M, Warner N. Tightly coupling cognition: understanding how communication and awareness drive coordination in teams. C2 Journal. 2008;2. Available: <https://apps.dtic.mil/sti/citations/ADA513557>
109. Letsky M, Warner N, Fiore SM, Rosen M, Salas E, OFFICE OF NAVAL RESEARCH ARLINGTON VA. Macrocognition in Complex Team Problem Solving. OFFICE OF NAVAL RESEARCH ARLINGTON VA; 2007 Jun. Available: <https://apps.dtic.mil/sti/citations/ADA481422>



110. Alberts JK, Tracy SJ, Trethewey A. An Integrative Theory of the Division of Domestic Labor: Threshold Level, Social Organizing and Sensemaking. *J Fam Commun.* 2011;11: 21–38.
111. Goertzel B. *The Hidden Pattern: A Patternist Philosophy of Mind.* Universal-Publishers; 2006.
112. Bach J. *Principles of Synthetic Intelligence PSI: An Architecture of Motivated Cognition.* Oxford University Press, USA; 2009.
113. Bose R. Competitive intelligence process and tools for intelligence analysis. *Industrial Management & Data Systems.* 2008;108: 510–528.
114. Kent S. *Strategic Intelligence for American World Policy.* Princeton University Press; 1949.
115. Treverton GF, Jones SG, Boraz S, Lipsky P, RAND CORP. Toward a theory of intelligence. Workshop report. RAND CORP ARLINGTON VA NATIONAL SECURITY RESEARCH DIV; 2006 Jan. Available: <https://apps.dtic.mil/sti/citations/ADA449313>
116. Kramer RM. A Failure to Communicate: 9/11 and the Tragedy of the Informational Commons. *International Public Management Journal.* 2005;8: 397–416.
117. Linn BM. *The Echo of Battle.* Harvard University Press; 2009.
118. Austin NJE, Rankov NB. *Exploratio: Military and Political Intelligence in the Roman World from the Second Punic War to the Battle of Adrianople.* Psychology Press; 1998.
119. Dedijer S. Ragusa Intelligence and Security (1301-1806): A Model for the Twenty-First Century? *Int J Intell CounterIntelligence.* 2002;15: 101–114.
120. Šundrica Z. *The Intelligence Activities of the Dubrovnik Republic in the Austro-Turkish War 1737-1739.* Dubrovnik Annals; 2001. Report No.: 8299. Available: [https://hrcak.srce.hr/index.php?id\\_clanak=8299&show=clanak](https://hrcak.srce.hr/index.php?id_clanak=8299&show=clanak)
121. Sweeney WC. *Military Intelligence, a New Weapon in War.* Frederick A. Stokes Company; 1924.
122. Moore DT. *Sensemaking: A Structure for an Intelligence Revolution.* Center for Strategic Intelligence Research, National Defense Intelligence College; 2011.
123. Walsh PF. *Intelligence and intelligence analysis: Exploring the state of intelligence since 9/11.* Crehan AC, editor. Charles Sturt University. 2013. Available: <https://researchoutput.csu.edu.au/en/publications/intelligence-and-intelligence-analysis-exploring-the-state-of-int>



124. Waltz E. Knowledge Management in the Intelligence Enterprise. Artech House; 2003.
125. Hribar G, Podbregar I, Ivanuša T. OSINT: A “Grey Zone”? *Int J Intell CounterIntelligence*. 2014;27: 529–549.
126. Bennett M, Waltz E. Counterdeception principles and applications for national security. Artech House; 2007.
127. Kambere G, Goh PH, Kumar P, Msafir F. The Financing of Lashkar-e-Taiba. *Combating Terrorism Exchange*. 2011;1: 18.
128. Glanz J, Rotella S, Sanger DE. In 2008 Mumbai attacks, piles of spy data, but an uncompleted puzzle. In: *New York Times* [Internet]. 2014. Available: <http://www.nytimes.com/2014/12/22/world/asia/in-2008-mumbai-attacks-piles-of-spy-data-but-an-uncompleted-puzzle.html>
129. Jackson BA. Groups, Networks, or Movements: A Command-and-Control-Driven Approach to Classifying Terrorist Organizations and Its Application to Al Qaeda. *Stud Conflict Terrorism*. 2006;29: 241–262.
130. Phythian M. Beyond the Intelligence Cycle? In: Phythian M, editor. *Understanding the Intelligence Cycle*. Routledge; 2013.
131. Gill P, Phythian M. From Intelligence Cycle to Web of Intelligence: Complexity and the conceptualisation of intelligence. In: Phythian M, editor. *Understanding the Intelligence Cycle*. Routledge; 2013.
132. Herman M. *Intelligence Power in Peace and War*. Cambridge University Press; 1996.
133. Johnson LK. Strategic intelligence: An American perspective. *Int J Intell CounterIntelligence*. 1989;3: 299–332.
134. Warner M. The Past and Future of the Intelligence Cycle. In: Phythian M, editor. *Understanding the Intelligence Cycle*. Routledge; 2013.
135. Hulnick AS. Intelligence Theory: Seeking better models. In: Phythian M, editor. *Understanding the Intelligence Cycle*. Routledge; 2013.
136. Lakoff G, Johnson M. *Metaphors we live by*. University Of Chicago Press; 1980.
137. McCreddie R, Macdonald C, Ounis I, Osborne M, Petrovic S. Scalable distributed event detection for Twitter. 2013 IEEE International Conference on Big Data. [ieeexplore.ieee.org](http://ieeexplore.ieee.org); 2013. pp. 543–549.

138. Morstatter F, Pfeffer J, Liu H, Carley KM. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. Seventh international AAAI conference on weblogs and social media. [aaai.org](http://aaai.org); 2013. Available: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewPaper/6071>
139. Office of the Director of National Intelligence. U.S. National Intelligence: An Overview. 2011.
140. Margolis JE. Estimating state instability. *Studies in Intelligence*. 2012;56: 13–24.
141. Peppler B. Dealing with the longer-term in intelligence practice: The application of a foresight approach. *Journal of the Australian Institute of Professional Intelligence Officers*. 2014;22: 35–51.
142. Stack KP. Competitive intelligence. *Intell Natl Sec*. 1998;13: 194–202.
143. Friedman JA, Zeckhauser R. Assessing Uncertainty in Intelligence. *Intell Natl Sec*. 2012;27: 824–847.
144. Wirtz JJ. Indications and Warning in an Age of Uncertainty. *Int J Intell CounterIntelligence*. 2013;26: 550–562.
145. Gentry JA, Gordon JS. *Strategic Warning Intelligence: History, Challenges, and Prospects*. Georgetown University Press; 2019.
146. Goldman J. “Actionable” Intelligence Is Not “Warning” Intelligence. Proceedings of the XIXth International Conference. [iksconference.ro](http://iksconference.ro); 2014. Available: [https://iksconference.ro/wp-content/uploads/2017/04/IKS\\_2014.pdf#page=21](https://iksconference.ro/wp-content/uploads/2017/04/IKS_2014.pdf#page=21)
147. Hulnick AS. What's wrong with the Intelligence Cycle. *Intell Natl Sec*. 2006;21: 959–979.
148. Johnson LK. Making the intelligence “Cycle” work. *Int J Intell CounterIntelligence*. 1986;1: 1–23.
149. Fournie DA. Harsh Lessons: Roman Intelligence in the Hannibalic War. *Int J Intell CounterIntelligence*. 2004;17: 502–538.
150. Johnson LK. Preface to a Theory of Strategic Intelligence. *Int J Intell CounterIntelligence*. 2003;16: 638–663.
151. Lukyanenko R, Parsons J, Wiersma Y, Wachinger G, Huber B, Meldt R. Representing Crowd Knowledge: Guidelines for Conceptual Modeling of User-generated Content. *Journal of the Association for Information Systems*. 2017;18: 2.

152. Eidson M, Kramer L, Stone W, Hagiwara Y, Schmit K, New York State West Nile Virus Avian Surveillance Team. Dead bird surveillance as an early warning system for West Nile virus. *Emerg Infect Dis.* 2001;7: 631–635.
153. Lukyanenko R, Parsons J, Wiersma Y. Citizen Science 2.0: Data Management Principles to Harness the Power of the Crowd. *Service-Oriented Perspectives in Design Science Research.* Springer Berlin Heidelberg; 2011. pp. 465–473.
154. Lukyanenko R, Parsons J, Wiersma YF. The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content. *Information Systems Research.* 2014;25: 669–689.
155. He J, van Ossenbruggen J, de Vries AP. Fish4label: accomplishing an expert task without expert knowledge. *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval.* groups.inf.ed.ac.uk; 2013. pp. 211–212.
156. Chasmer LE, Ryerson RA, Coburn CA. Educating the Next Generation of Remote Sensing Specialists: Skills and Industry Needs in a Changing World. *Can J Remote Sens.* 2021; 1–16.
157. Ekins E. Poll: 62% of Americans Say They Have Political Views They're Afraid to Share. In: *cato.org* [Internet]. 22 Jul 2020 [cited 27 Jul 2021]. Available: <https://www.cato.org/survey-reports/poll-62-americans-say-they-have-political-views-theyre-afraid-share?queryID=98d4c30397de4e786f8625a2b87a2380>
158. Musto J. School principal sues over dismissal for sharing conservative memes on Facebook. In: *foxnews.com* [Internet]. 9 Dec 2020 [cited 27 Jul 2021]. Available: <https://www.foxnews.com/us/school-principal-sues-after-being-fired-for-sharing-conservative-memes>
159. Reuters. Disney's Lucasfilm ditches "The Mandalorian" star Gina Carano over social media posts. *Reuters.* 11 Feb 2021. Available: <https://www.reuters.com/article/us-walt-disney-lucasfilm-carano-idUSKBN2AB0PL>. Accessed 27 Jul 2021.
160. Zuckerman E. Mistrust, efficacy and the new civics: understanding the deep roots of the crisis of faith in journalism. 2017 [cited 27 Jul 2021]. Available: <https://dspace.mit.edu/handle/1721.1/110987?show=full>
161. Alexander S. NYT Is Threatening My Safety By Revealing My Real Name, So I Am Deleting The Blog. In: *slatestarcodex.com* [Internet]. 23 Jun 2020 [cited 27 Jul 2021]. Available: <https://slatestarcodex.com/2020/06/22/nyt-is-threatening-my-safety-by-revealing-my-real-name-so-i-am-deleting-the-blog/>

162. Middlebrook C. The Grey Area: Instagram, Shadowbanning, and the Erasure of Marginalized Communities. Shadowbanning, and the Erasure of Marginalized. 2020. doi:10.2139/ssrn.3539721
163. Le Merrer E, Morgan B, Trédan G. Setting the Record Straighter on Shadow Banning. arXiv [cs.SI]. 2020. Available: <http://arxiv.org/abs/2012.05101>
164. Jaidka K, Mukerjee S, Lelkes Y. An audit of Twitter’s shadowban sanctions in the United States. [yahootechpulse.easychair.org](https://yahootechpulse.easychair.org/publications/preprint_download/z4jt); 2021 [cited 27 Jul 2021]. Available: [https://yahootechpulse.easychair.org/publications/preprint\\_download/z4jt](https://yahootechpulse.easychair.org/publications/preprint_download/z4jt)
165. Chambers S. The amazing banned memes from China. In: Index on Censorship [Internet]. 1 Mar 2018 [cited 27 Jul 2021]. Available: <https://www.indexoncensorship.org/2018/03/amazing-banned-memes-china/>
166. Zadrozny B. Twitter troll arrested, accused of election interference related to disinformation campaign. In: NBC News [Internet]. 27 Jan 2021 [cited 27 Jul 2021]. Available: <https://www.nbcnews.com/tech/internet/twitter-troll-arrested-election-interference-related-disinformation-campaign-n1255864>
167. Allen-Ebrahimian B. University of Minnesota student jailed in China for tweets critical of government. 23 Jan 2020 [cited 27 Jul 2021]. Available: <https://www.axios.com/china-arrests-university-minnesota-twitter-e495cf47-d895-4014-9ac8-8dc76aa6004d.html>
168. U.S. Army. Open-Source Intelligence ATP 2-22.9. 2012.
169. Tufte ER. Visual Explanations: Images and Quantities, Evidence and Narrative. Computers in Physics. 1998;12. doi:10.1063/1.168637
170. Cordes RJ, Friedman DA. The Facilitator’s Catechism. In: Friedman DA, Cordes RJ, editors. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
171. Lippi M, Torroni P. Context-Independent Claim Detection for Argument Mining. Twenty-Fourth International Joint Conference on Artificial Intelligence. [aaai.org](http://aaai.org); 2015. Available: <https://www.aaai.org/ocs/index.php/IJCAI/IJCAI15/paper/viewPaper/10942>
172. Lawrence J, Reed C. Argument mining: A survey. Comput Linguist. 2020. Available: <https://direct.mit.edu/coli/article-abstract/45/4/765/93362>
173. Lippi M, Torroni P. Argument mining: A machine learning perspective. International Workshop on Theory and Applications of. 2015. Available: [https://link.springer.com/chapter/10.1007/978-3-319-28460-6\\_10](https://link.springer.com/chapter/10.1007/978-3-319-28460-6_10)
174. Peterson C, Rauschnot M, Mc Clain T. Shark alert: Early warning system. [digitalcommons.calpoly.edu](http://digitalcommons.calpoly.edu); 2015 [cited 27 Aug 2021]. Available:

- <http://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=1341&context=eesp>
175. Brownstein JS, Freifeld CC. HealthMap: the development of automated real-time internet surveillance for epidemic intelligence. *Euro Surveill*. 2007;12: E071129.5.
  176. Daneels A, Salter W. What is SCADA? 1999. Available: <https://accelconf.web.cern.ch/ica99/papers/mc1i01.pdf>
  177. Yu X, Wu P, Han W, Zhang Z, Others. A remote SCADA system for keeping fruits and vegetables fresh with ozone. *Int J Food Agric Environ*. 2013;11: 187–192.
  178. Jiang J, Wang P, Nan J, Guo L, Ran M, Others. Songhua River monitoring system on point sources and water quality with SCADA and GIS. *Environ Sci Technol*. 2013;36: 132–136.
  179. Cannon P. Taking an “Ecological Turn” in the Evaluation of Rhetorical Interventions. University of South Florida. 2019. Available: <https://scholarcommons.usf.edu/etd/8010>
  180. Godwin M. Meme, Counter-meme. *Wired*. 1 Oct 1994. Available: <https://www.wired.com/1994/10/godwin-if-2/>. Accessed 12 Aug 2021.
  181. Projet Debater - IBM Research AI. [cited 17 Sep 2021]. Available: <https://www.research.ibm.com/interactive/project-debater/>
  182. Swarmcheck.ai. Swarmcheck. In: [swarmcheck.ai](https://www.swarmcheck.ai) [Internet]. 2021 [cited 17 Sep 2021]. Available: <https://www.swarmcheck.ai/home>
  183. Perelman C, Olbrechts-Tyteca L. *The New Rhetoric: A Treatise on Argumentation*. University of Notre Dame Press; 1971.
  184. Wikipedia contributors. Land ethic. In: *Wikipedia, The Free Encyclopedia* [Internet]. 11 Dec 2020 [cited 26 Mar 2021]. Available: [https://en.wikipedia.org/w/index.php?title=Land\\_ethic&oldid=993635012](https://en.wikipedia.org/w/index.php?title=Land_ethic&oldid=993635012)
  185. Ostrom E. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press; 1990.
  186. Boik JC. Science-Driven Societal Transformation, Part I: Worldview. *Sustain Sci Pract Policy*. 2020;12: 6881.
  187. Frey S, Schneider N. Effective Voice: Beyond Exit and Affect in Online Communities. *arXiv [cs.CY]*. 2020. Available: <http://arxiv.org/abs/2009.12470>

188. Schneider N, De Filippi P, Frey S, Tan JZ, Zhang AX. Modular Politics: Toward a Governance Layer for Online Communities. *Proc ACM Hum-Comput Interact.* 2021;5: 1–26.
189. Jhaver S, Frey S, Zhang A. Designing for Multiple Centers of Power: A Taxonomy of Multi-level Governance in Online Social Platforms. *arXiv [cs.HC]*. 2021. Available: <http://arxiv.org/abs/2108.12529>

## Chapter V Works Cited

---

1. Cordes RJ, Friedman DA. Emergent Teams for Complex Threats. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
2. Schoeller F, Miller M, Salomon R, Friston K. Trust as Extended Control: Active Inference and User Feedback During Human-Robot Collaboration. ArXiv. 2021 [cited 21 Sep 2021]. Available: <https://www.semanticscholar.org/paper/7c39921b6050d7be82e18df2a31973f2c3f864ea>
3. Vyatkin A, Metelkin I, Mikhailova A, Cordes RJ, Friedman DA. Active Inference & Behavior Engineering for Teams. 2020. doi:10.5281/zenodo.4021163
4. Guénin--Carlut A. Thinking like a State - Embodied intelligence in the deep history of our collective minds. 2021 [cited 21 Sep 2021]. Available: <https://osf.io/dxnzt/>
5. Collier JL, Collier C. Slavery and the Coming of the Civil War: 1831 - 1861. Blackstone Publishing; 2012. Available: <https://play.google.com/store/books/details?id=8Mt47-uzDH0C>
6. Tzu S. Sun Tzu Art of War. Vij Books India Pvt Ltd; 2012. Available: <https://play.google.com/store/books/details?id=VPSPCQAAQBAJ>
7. Milner NP. Vegetius: Epitome of military science. Liverpool: Translated Texts for Historians. 1993;16: 14.
8. von Moltke H, Hughes DJ. Moltke on the Art of War: Selected Writings. Ballantine Books; 1993.
9. Machiavelli N. Art of War. New edition. University of Chicago Press; 2005. Available: <https://www.amazon.com/Art-War-Niccol%C3%B2-Machiavelli/dp/0226500462>

10. Bin S. Sun Bin: The Art of Warfare: A Translation of the Classic Chinese Work of Philosophy and Strategy (SUNY Series in Chinese Philosophy and Culture). State University of New York Press; 2003. Available: <https://www.amazon.com/Sun-Bin-Translation-Philosophy-Strategy/dp/0791454967>
11. Tse-Tung M. The Art of War by Mao Tse-tung - Special Edition. El Paso Norte Press; 2011. Available: <https://play.google.com/store/books/details?id=2uVIYgEACAAJ>
12. Jomini A-H, Mendell GH, Craighill WP. The Art of War. Courier Corporation; 2007. Available: <https://play.google.com/store/books/details?id=mdl6AwAAQBAJ>
13. Carl Von Clausewitz, Howard M, Paret P, Howard M, Brodie B. On War. Howard M, Paret P, editors. Princeton University Press; 1989.
14. van Creveld M. More on War. Oxford University Press; 2016. Available: <https://play.google.com/store/books/details?id=oWL0DQAAQBAJ>
15. Villacres EJ, Bassford C. Reclaiming the Clausewitzian Trinity. West Point Military Academy; 1995 Jan. Available: <https://apps.dtic.mil/sti/citations/ADA528269>
16. Kelly J, Kilcullen D. Chaos Versus Predictability: A Critique of Effects-Based Operations. Security Challenges. 2006;2: 63–73. Available: <http://www.jstor.org/stable/26458837>
17. Mattis J. Commander's Guidance for Effects-Based Operations. US Joint Forces Command; 2018.
18. Knox BMW, Murray W. The Dynamics of Military Revolution, 1300-2050. Cambridge University Press; 2001. Available: <https://play.google.com/store/books/details?id=zllbUmwXitAC>
19. Krepinevich AF. Cavalry to Computer: The Pattern of Military Revolutions. The National Interest. 1994; 30–42. Available: <http://www.jstor.org/stable/42896863>
20. Knox BMW, Murray W. The Dynamics of Military Revolution, 1300-2050. Cambridge University Press; 2001. Available: <https://play.google.com/store/books/details?id=zllbUmwXitAC>
21. Hoffman FG. Will War's Nature Change in the Seventh Military Revolution? Parameters. 2017;47: 19–31. Available: <https://publications.armywarcollege.edu/pubs/3554.pdf>
22. Lind WS, Nightengale K, Schmitt JF, Sutton JW, Wilson GI. The changing face of war: Into the fourth generation. The Marine Corps Gazette. 1989: 22–26. Available:



<https://www.academia.edu/download/15486651/thechangingfaceofwar-onscreen.pdf>

23. Abbott DH, editor. *The Handbook of Fifth-Generation Warfare (5GW)*. Nimble Books; 2010. Available: <https://www.amazon.com/Handbook-Fifth-Generation-Warfare-5gw/dp/1934840173>
24. Cordes RJ. *The Next Generation of Security: Prioritizing Information Warfare Defense*. 7 Aug 2021 [cited 12 Aug 2021]. Available: <https://www.hstoday.us/subject-matter-areas/intelligence/the-next-generation-of-security-prioritizing-information-warfare-defense/>
25. Kohalyk C. 5GW as Netwar 2.0. In: Abbott DH, editor. *The Handbook of Fifth-Generation Warfare (5GW)*. Nimble Books; 2010. pp. 38–52. Available: <https://www.amazon.com/Handbook-Fifth-Generation-Warfare-5gw/dp/1934840173>
26. Linn BM. *The Echo of Battle*. Harvard University Press; 2009. Available: <https://play.google.com/store/books/details?id=cIKtd4O3MOgC>
27. Lanchester FW. *Aircraft in Warfare: The Dawn of the Fourth Arm*. Constable limited; 1916. Available: <https://play.google.com/store/books/details?id=fIZCAAAIAAJ>
28. Kress M. Lanchester Models for Irregular Warfare. *Sci China Ser A Math*. 2020;8: 737. doi:10.3390/math8050737
29. González E, Villena M. Spatial Lanchester models. *Eur J Oper Res*. 2011;210: 706–715. doi:10.1016/j.ejor.2010.11.009
30. Nicolae F, Cotorcea A, Ristea M, Atodiresei D. Human Reliability Using the Fault Tree Analysis. A Case Study of a Military Accident Investigation. International conference KNOWLEDGE-BASED ORGANIZATION. researchgate.net; 2016. Available: [https://www.researchgate.net/profile/Dinu-Atodiresei/publication/305760875\\_Human\\_Reliability\\_Using\\_the\\_Fault\\_Tree\\_Analysis\\_A\\_Case\\_Study\\_of\\_a\\_Military\\_Accident\\_Investigation/links/58d2be9da6fdccd24d43bc41/Human-Reliability-Using-the-Fault-Tree-Analysis-A-Case-Study-of-a-Military-Accident-Investigation.pdf](https://www.researchgate.net/profile/Dinu-Atodiresei/publication/305760875_Human_Reliability_Using_the_Fault_Tree_Analysis_A_Case_Study_of_a_Military_Accident_Investigation/links/58d2be9da6fdccd24d43bc41/Human-Reliability-Using-the-Fault-Tree-Analysis-A-Case-Study-of-a-Military-Accident-Investigation.pdf)
31. La Band RA, Andrews JD. Phased mission modelling using fault tree analysis. *Proc Inst Mech Eng Part E J Process Mech Eng*. 2004;218: 83–91. doi:10.1243/095440804774134262
32. Cordes RJ, David S, Maan A, Ruiz A, Sapp E, Scannell P, et al. *The Narrative Campaign Field Guide - First Edition*. 1st ed. Cordes RJ, editor. Narrative Strategies Ink; 2021. Available: <https://www.narrative-strategies.com/ncfg>

33. Wagenhals LW, Alex L, Haider S. Planning, Execution, and Assessment of Effects-Based Operations (EBO). GEORGE MASON UNIV FAIRFAX VA CENTER OF EXCELLENCE IN COMMAND CONTROL COMMUNICATIONS AND INTELLIGENCE; 2006 May. Available: <https://apps.dtic.mil/sti/citations/ADA451493>
34. Hause M. The Unified Profile for DoDAF/MODAF (UPDM) enabling systems of systems on many levels. 2010 IEEE International Systems Conference. [ieeexplore.ieee.org](http://ieeexplore.ieee.org); 2010. pp. 426–431. doi:10.1109/SYSTEMS.2010.5482450
35. Handley HAH. Incorporating the NATO human view in the DoDAF 2.0 meta model. Syst Eng Electr. 2012;15: 108–117. doi:10.1002/sys.20206
36. Rumer E. The Primakov (not Gerasimov) doctrine in action. Carnegie Endowment for International Peace; 2019. Available: [https://carnegieendowment.org/files/Rumer\\_PrimakovDoctrine\\_final1.pdf](https://carnegieendowment.org/files/Rumer_PrimakovDoctrine_final1.pdf)
37. Morris V. Grading Gerasimov: Evaluating Russian Nonlinear War Through Modern Chinese Doctrine. Small Wars Journal. 2015. Available: <https://smallwarsjournal.com/jrnl/art/grading-gerasimov-evaluating-russian-nonlinear-war-through-modern-chinese-doctrine>. Accessed 18 Oct 2021.
38. Engstrom J. Systems Confrontation and System Destruction Warfare: How the Chinese People’s Liberation Army Seeks to Wage Modern Warfare. RAND; 2018. Report No.: RR-1708-OSD.
39. Grant T, Kooter B. Network-centric warfare: Its origin and future. Future of C2. academia.edu; 1998. Available: [https://www.academia.edu/download/36514207/1\\_NCW\\_Origin\\_and\\_FutureVice\\_Admiral\\_Arthur\\_K\\_Cebrowski.pdf](https://www.academia.edu/download/36514207/1_NCW_Origin_and_FutureVice_Admiral_Arthur_K_Cebrowski.pdf)
40. Yang A, Curtis N, Abbass HA, Sarker R, Barlow M. WISDOM-II: A network centric model for warfare. Complex Science for a Complex World: Exploring Human Ecosystems with Agents. 2006; 149–173. Available: <https://library.oapen.org/bitstream/handle/20.500.12657/33791/458885.pdf?sequence=1#page=167>
41. Park S-Y, Shin H-Y, Lee T-S, Choi B-W. Design of the Agent-based Network-Centric Warfare Modeling System. Journal of the Korea Society for Simulation. 2010;19: 271–280. Available: <https://www.koreascience.or.kr/article/JAKO201016450102380.page>
42. Kang BG, Choi SH, Kwon SJ, Lee JH, Kim TG. Simulation-Based Optimization on the System-of-Systems Model via Model Transformation and Genetic Algorithm: A Case Study of Network-Centric Warfare. Complexity. 2018;2018. doi:10.1155/2018/4521672

43. Grant T, Kooter B. Comparing ooda & other models as operational view c2 architecture topic: C4isr/c2 architecture. ICCRTS2005, Jun. 2005. Available: [https://www.researchgate.net/profile/Tim-Grant-9/publication/237674561\\_Comparing\\_OODA\\_other\\_models\\_as\\_Operational\\_View\\_C2\\_Architecture/links/584dbdde08aeb9892526444d/Comparing-OODA-other-models-as-Operational-View-C2-Architecture.pdf](https://www.researchgate.net/profile/Tim-Grant-9/publication/237674561_Comparing_OODA_other_models_as_Operational_View_C2_Architecture/links/584dbdde08aeb9892526444d/Comparing-OODA-other-models-as-Operational-View-C2-Architecture.pdf)
44. Mees W, Debatty T. An attempt at defining cyberdefense situation awareness in the context of command & control. 2015 [cited 18 Oct 2021]. doi:10.1109/ICMCIS.2015.7158674
45. Endsley MR. A taxonomy of situation awareness errors. Human factors in aviation operations. 1995;3: 287–292. Available: [https://www.researchgate.net/profile/Mica-Endsley/publication/285731357\\_A\\_taxonomy\\_of\\_situation\\_awareness\\_errors\\_human\\_factors\\_in\\_aviation\\_operations/links/58322a6f08ae138f1c07a4e3/A-taxonomy-of-situation-awareness-errors-human-factors-in-aviation-operations.pdf](https://www.researchgate.net/profile/Mica-Endsley/publication/285731357_A_taxonomy_of_situation_awareness_errors_human_factors_in_aviation_operations/links/58322a6f08ae138f1c07a4e3/A-taxonomy-of-situation-awareness-errors-human-factors-in-aviation-operations.pdf)
46. Plehn MT. Control warfare: Inside the OODA Loop. AIR UNI MAXWELL AFB AL SCHOOL OF ADVANCED AIRPOWER STUDIES; 2000 Jun. Available: <https://apps.dtic.mil/sti/citations/ADA391774>
47. Hammond GT. Reflections on the Legacy of John Boyd. Contemporary Security Policy. 2013;34: 600–602. doi:10.1080/13523260.2013.842297
48. Kalloniatis A, Rowe C, La P, Holder A, Bennier J, Mitchell B. Network Analysis of Decision Loops in Operational Command and Control Arrangements. Data and Decision Sciences in Action. Springer International Publishing; 2018. pp. 343–355. doi:10.1007/978-3-319-55914-8\_25
49. Boyd J. A discourse on winning and losing. Hammond GT, editor. Air University Press; 2018. Available: [https://www.airuniversity.af.edu/Portals/10/AUPress/Books/B\\_0151\\_Boyd\\_Discourse\\_Winning\\_Losing.pdf](https://www.airuniversity.af.edu/Portals/10/AUPress/Books/B_0151_Boyd_Discourse_Winning_Losing.pdf)
50. Vettorello M, Eisenbart B, Ranscombe C. Toward Better Design-Related Decision Making: A Proposal of an Advanced OODA Loop. Proceedings of the Design Society: International Conference on Engineering Design. 2019;1: 2387–2396. doi:10.1017/dsi.2019.245
51. Silvester J, Angelin L. Introducing intents to the OODA-loop. Procedia Comput Sci. 2019;159: 878–883. doi:10.1016/j.procs.2019.09.247

52. McManus N, Haddad AN. Incident Records: Understanding the Past to Prevent Future Hazardous Energy Incidents. *Prof Saf.* 2014;59: 34–43. Available: <https://onepetro.org/PS/article-abstract/59/12/34/33209>
53. Pawson R, Wong G, Owen L. Known Knowns, Known Unknowns, Unknown Unknowns: The Predicament of Evidence-Based Policy. *American Journal of Evaluation.* 2011;32: 518–546. doi:10.1177/1098214011403831
54. Alles M. Governance in the age of unknown unknowns. *International Journal of Disclosure and Governance.* 2009;6: 85–88. doi:10.1057/jdg.2009.2
55. Birkemo GA. Is Norwegian long term defence planning risk based? *ffi-publikasjoner.archive...*; 2013. Available: <https://ffi-publikasjoner.archive.knowledgearc.net/handle/20.500.12242/929>
56. Žižek S. Philosophy, the “unknown knowns,” and the public use of reason. *Topoi.* 2006;25: 137–142. doi:10.1007/s11245-006-0021-2
57. Jackson R. Unknown knowns: the subjugated knowledge of terrorism studies. *Critical Studies on Terrorism.* 2012;5: 11–29. doi:10.1080/17539153.2012.659907
58. Sarewitz D. Unknown Knowns. *Issues Sci Technol.* 2020;37: 18–19. Available: <https://issues.org/wp-content/uploads/2020/09/18-19-Sarewitz-Editors-Journal-Fall-2020-ISSUES.pdf>
59. Logan DC. Known knowns, known unknowns, unknown unknowns and the propagation of scientific enquiry. *Journal of experimental botany.* *academic.oup.com*; 2009. pp. 712–714. doi:10.1093/jxb/erp043
60. Loxdale HD, Davis BJ, Davis RA. Known knowns and unknowns in biology. *Biol J Linn Soc Lond.* 2016;117: 386–398. doi:10.1111/bij.12646
61. Bassford C. Christopher Bassford: Tiptoe Through the Trinity. In: <https://www.clausewitzstudies.org/> [Internet]. 2016 [cited 28 Sep 2021]. Available: <https://www.clausewitzstudies.org/mobile/trinity8.htm>
62. Nielsen JN. The Generational Warfare Model. In: [geopolicraticus.wordpress.com](http://geopolicraticus.wordpress.com) [Internet]. 27 Oct 2010 [cited 22 Sep 2021]. Available: <https://geopolicraticus.wordpress.com/2010/10/26/the-generational-warfare-model/>
63. Alexander C. *A Pattern Language: Towns, Buildings, Construction.* Oxford University Press; 1977. Available: <https://play.google.com/store/books/details?id=mW7RCwAAQBAJ>

64. Mascarenhas M, Cordes RJ, Friedman DA. Digital Rhetorical Ecosystem Analysis: Sensemaking of Digital Memetic Discourse. Zenodo. 2021. doi:10.5281/zenodo.5573947
65. Naylor RT. Economic Warfare: Sanctions, Embargo Busting, and Their Human Cost. UPNE; 2001. Available: <https://play.google.com/store/books/details?id=fkcGKb8jLXMC>
66. Clemens J. An analysis of economic warfare. Am Econ Rev. 2013. Available: <https://www.aeaweb.org/articles?id=10.1257/aer.103.3.523>
67. Tang M. Huawei Versus the United States? The Geopolitics of Exterritorial Internet Infrastructure. Int J Commun Syst. 2020;14: 22. Available: <https://ijoc.org/index.php/ijoc/article/view/12624>
68. Cheng D. Winning without fighting: Chinese legal warfare. Backgrounder. 2012. Available: [http://thf\\_media.s3.amazonaws.com/2012/pdf/bg2692.pdf](http://thf_media.s3.amazonaws.com/2012/pdf/bg2692.pdf)
69. Kotani T. Freedom of navigation and the US-Japan alliance: Addressing the threat of legal warfare. US-Japan Papers. 2011; 1–6. Available: <https://www.jcie.org/researchpdfs/USJapanPapers/Kotani.pdf>
70. Atlantic Council GeoTech Center. Standardizing the future: How can the United States navigate the geopolitics of international technology standards? In: Atlanticcouncil.org [Internet]. 14 Oct 2021 [cited 21 Oct 2021]. Available: <https://www.atlanticcouncil.org/in-depth-research-reports/report/standardizing-the-future-how-can-the-united-states-navigate-the-geopolitics-of-international-technology-standards/>
71. Atlantic Council GeoTech Center. Report of the Commission on the Geopolitical Impacts of New Technologies and Data - GeoTech Commission Report. 2021. Report No.: v51j.
72. Caputo A, Marzi G, Maley J, Silic M. Ten years of conflict management research 2007-2017: An update on themes, concepts and relationships. International Journal of Conflict Management. 2018;30: 87–110. doi:10.1108/IJCMA-06-2018-0078
73. Kimbrough EO, Sheremeta RM. Theories of conflict and war. J Econ Behav Organ. 2019;159: 384–387. doi:10.1016/j.jebo.2019.02.007
74. Walton RE, McKersie RB. A Behavioral Theory of Labor Negotiations: An Analysis of a Social Interaction System. Cornell University Press; 1991. Available: <https://play.google.com/store/books/details?id=dW02zPVX9rQC>

75. Burchill F. Walton and McKersie, A Behavioral Theory of Labor Negotiations (1965). *Historical Studies in Industrial Relations*. 1999; 137–168. doi:10.3828/hsir.1999.8.7
76. Werbach K. Supercommons: Toward a Unified Theory of Wireless Communication. 2003. doi:10.2139/ssrn.456020
77. Prainsack B. Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society*. 2019;6: 2053951719829773. doi:10.1177/2053951719829773
78. nobelprize.org. Elinor Ostrom Facts. In: nobelprize.org [Internet]. 2021 [cited 4 Nov 2021]. Available: <https://www.nobelprize.org/prizes/economic-sciences/2009/ostrom/facts/>
79. Hess C, Ostrom E. Understanding Knowledge as a Commons: From Theory to Practice. MIT Press; 2011. Available: <https://play.google.com/store/books/details?id=5UCCkgAACAAJ>
80. Friedman D, Tschantz A, Ramstead MJD, Friston K, Constant A. Active inferants: The basis for an active inference framework for ant colony behavior. *Front Behav Neurosci*. 2021;15: 126. doi:10.3389/fnbeh.2021.647732
81. Da Costa L, Parr T, Sajid N, Veselic S, Neacsu V, Friston K. Active inference on discrete state-spaces: A synthesis. *J Math Psychol*. 2020;99: 102447. doi:10.1016/j.jmp.2020.102447
82. Parr T, Pezzulo G, Friston KJ. Active Inference: The Free Energy Principle in Mind, Brain, and Behavior. 2022.
83. Linson A, Clark A, Ramamoorthy S, Friston K. The Active Inference Approach to Ecological Perception: General Information Dynamics for Natural and Artificial Embodied Cognition. *Frontiers in Robotics and AI*. 2018;5: 21. doi:10.3389/frobt.2018.00021
84. Gallagher S, Allen M. Active inference, enactivism and the hermeneutics of social cognition. *Synthese*. 2018;195: 2627–2648. doi:10.1007/s11229-016-1269-8
85. Fox S. Accessing Active Inference Theory through Its Implicit and Deliberative Practice in Human Organizations. *Entropy* . 2021;23: 1521. doi:10.3390/e23111521
86. Millidge B. Combining active inference and hierarchical predictive coding: A tutorial introduction and case study. *PsyArXiv*. 2019. doi:10.31234/osf.io/kf6wc

87. Active Inference Lab. ActInfLab ModelStream #001.1: “A Step-by-Step Tutorial on Active Inference.” Youtube; 15 Jan 2021 [cited 11 Aug 2021]. Available: <https://www.youtube.com/watch?v=H5AolqFl2Nw>
88. Smith R, Friston K, Whyte C. A Step-by-Step Tutorial on Active Inference and its Application to Empirical Data. 2021. doi:10.31234/osf.io/b4jm6
89. Millidge B, Seth A, Buckley CL. Predictive Coding: a Theoretical and Experimental Review. arXiv [cs.AI]. 2021. Available: <http://arxiv.org/abs/2107.12979>
90. ActInfLab. [cited 16 Nov 2021]. Available: <https://www.activeinference.org/>
91. Rubin S, Parr T, Da Costa L, Friston K. Future climates: Markov blankets and active inference in the biosphere. *J R Soc Interface*. 2020;17: 20200503. doi:10.1098/rsif.2020.0503
92. Ramstead MJD, Kirchhoff MD, Constant A, Friston KJ. Multiscale integration: beyond internalism and externalism. *Synthese*. 2021;198: 41–70. doi:10.1007/s11229-019-02115-x
93. Hipólito I, Ramstead MJD, Convertino L, Bhat A, Friston K, Parr T. Markov blankets in the brain. *Neurosci Biobehav Rev*. 2021;125: 88–97. doi:10.1016/j.neubiorev.2021.02.003
94. Kirchhoff M, Parr T, Palacios E, Friston K, Kiverstein J. The Markov blankets of life: autonomy, active inference and the free energy principle. *J R Soc Interface*. 2018;15. doi:10.1098/rsif.2017.0792
95. Sims M. How to count biological minds: symbiosis, the free energy principle, and reciprocal multiscale integration. *Synthese*. 2020. doi:10.1007/s11229-020-02876-w
96. Fox S. Active Inference: Applicability to Different Types of Social Organization Explained through Reference to Industrial Engineering and Quality Management. *Entropy* . 2021;23: 198. doi:10.3390/e23020198
97. Sajid N, Ball PJ, Parr T, Friston KJ. Active inference: demystified and compared. arXiv [cs.AI]. 2019. Available: <http://arxiv.org/abs/1909.10863>
98. Friston K, Fortier M, Friedman DA. Of woodlice and men. *ALIUS Bulletin*. 2018;2: 17. Available: [https://www.aliusresearch.org/uploads/9/1/6/0/91600416/alius\\_bulletin\\_n%C2%B02\\_2018\\_.pdf#page=27](https://www.aliusresearch.org/uploads/9/1/6/0/91600416/alius_bulletin_n%C2%B02_2018_.pdf#page=27)
99. Ramstead MJD, Kirchhoff MD, Friston KJ. A tale of two densities: active inference is enactive inference. *Adapt Behav*. 2019;28: 1059712319862774. doi:10.1177/1059712319862774



100. Lobo L, Heras-Escribano M, Travieso D. The History and Philosophy of Ecological Psychology. *Front Psychol.* 2018;9: 2228. doi:10.3389/fpsyg.2018.02228
101. Friston K. The free-energy principle: a unified brain theory? *Nat Rev Neurosci.* 2010;11: 127–138. doi:10.1038/nrn2787
102. Friston K, Heins C, Ueltzhöffer K, Da Costa L, Parr T. Stochastic Chaos and Markov Blankets. *Entropy* . 2021;23. doi:10.3390/e23091220
103. Parr T, Friston KJ. Generalised free energy and active inference. *Biol Cybern.* 2019;113: 495–513. doi:10.1007/s00422-019-00805-w
104. Ramstead MJD, Constant A, Badcock PB, Friston KJ. Variational ecology and the physics of sentient systems. *Phys Life Rev.* 2019. doi:10.1016/j.plrev.2018.12.002
105. Friston K. A free energy principle for a particular physics. *arXiv [q-bio.NC]*. 2019. Available: <http://arxiv.org/abs/1906.10184>
106. Bolis D, Schilbach L. “Through others we become ourselves”: The dialectics of predictive coding and active inference. 2019. doi:10.31234/osf.io/6uwyn
107. Veissière SPL, Constant A, Ramstead MJD, Friston KJ, Kirmayer LJ. Thinking through other minds: A variational approach to cognition and culture. *Behav Brain Sci.* 2019;43: 1–97. doi:10.1017/S0140525X19001213
108. Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier; 2011. Available: [https://play.google.com/store/books/details?id=G\\_qdEsDlcp0C](https://play.google.com/store/books/details?id=G_qdEsDlcp0C)
109. Friston K, Mattout J, Kilner J. Action understanding and active inference. *Biol Cybern.* 2011;104: 137–160. doi:10.1007/s00422-011-0424-z
110. Butz MV, Bilkey D, Humaidan D, Knott A, Otte S. Learning, planning, and control in a monolithic neural event inference architecture. *Neural Netw.* 2019;117: 135–144. doi:10.1016/j.neunet.2019.05.001
111. Ramstead MJD, Veissière SPL, Kirmayer LJ. Cultural Affordances: Scaffolding Local Worlds Through Shared Intentionality and Regimes of Attention. *Front Psychol.* 2016;7: 1090. doi:10.3389/fpsyg.2016.01090
112. Constant A, Ramstead MJD, Veissière SPL, Friston K. Regimes of Expectations: An Active Inference Model of Social Conformity and Human Decision Making. *Front Psychol.* 2019;10: 679. doi:10.3389/fpsyg.2019.00679
113. Tschantz A, Seth AK, Buckley CL. Learning action-oriented models through active inference. *PLoS Comput Biol.* 2020;16: e1007805. doi:10.1371/journal.pcbi.1007805



114. Kaufmann R, Gupta P, Taylor J. An Active Inference Model of Collective Intelligence. *Entropy* . 2021;23. doi:10.3390/e23070830
115. Kittrie OF. *Lawfare: Law as a Weapon of War*. Oxford University Press; 2016. Available: <https://play.google.com/store/books/details?id=r1jhCgAAQBAJ>
116. Dunlap CJ Jr. *Lawfare 101: A Primer*. 2017 [cited 8 Jun 2021]. Available: [https://scholarship.law.duke.edu/faculty\\_scholarship/3742/](https://scholarship.law.duke.edu/faculty_scholarship/3742/)
117. Lee S. China's "Three Warfares": Origins, Applications, and Organizations. *Journal of Strategic Studies*. 2014;37: 198–221. doi:10.1080/01402390.2013.870071
118. Isomura T, Parr T, Friston K. Bayesian Filtering with Multiple Internal Models: Toward a Theory of Social Intelligence. *Neural Comput*. 2019; 1–42. doi:10.1162/neco\_a\_01239
119. Albarracin M, Constant A, Friston K, Ramstead M. A variational approach to scripts. 2020. Available: <https://psyarxiv.com/67zy4/download>
120. Bouizegarene N, Ramstead M, Constant A, Friston K, Kirmayer L. Narrative as active inference. 2020. doi:10.31234/osf.io/47ub6
121. cogsec.org. What is Cognitive Security? In: cogsec.org [Internet]. 2021 [cited 22 Nov 2021]. Available: <https://www.cogsec.org/what-is-cognitive-security>
122. Bezuidenhout M, Mouton F, Venter HS. Social engineering attack detection model: SEADM. 2010 Information Security for South Africa. [ieeexplore.ieee.org](http://ieeexplore.ieee.org); 2010. pp. 1–8. doi:10.1109/ISSA.2010.5588500
123. Cordes RJ, Applegate-Swanson S, Friedman DA, Knight VB, Mikhailova A. Narrative Information Management. Zenodo. COGSEC; 2021. doi:10.5281/zenodo.5573287
124. Wood W. Attitude change: persuasion and social influence. *Annu Rev Psychol*. 2000;51: 539–570. doi:10.1146/annurev.psych.51.1.539
125. Maslow A, Lewis KJ. Maslow's hierarchy of needs. Salenger Incorporated. 1987;14: 987–990. Available: <https://www.researchhistory.org/2012/06/16/maslows-hierarchy-of-needs/>
126. Austin NJE, Rankov NB. *Exploratio: Military and Political Intelligence in the Roman World from the Second Punic War to the Battle of Adrianople*. Psychology Press; 1998. Available: <https://play.google.com/store/books/details?id=NqleloHqezUC>
127. Bray DA. *Knowledge Ecosystems: A Theoretical Lens for Organizations Confronting Hyperturbulent Environments*. 2007. Available: <https://papers.ssrn.com/abstract=984600>

128. Friedman DA, Cordes RJ. Infinite Games for Infinite Teams. DARPA; 2020 Jul.
129. cadCAD – A Python package for designing, testing and validating complex systems through simulation. [cited 26 Mar 2020]. Available: <https://cadcad.org/>
130. ForneyLab.jl: Julia package for automatically generating Bayesian inference algorithms through message passing on Forney-style factor graphs. Github; Available: <https://github.com/biaslab/ForneyLab.jl>
131. infer-actively. Github; Available: <https://github.com/infer-actively>

## Chapter VI Works Cited

---

1. Weaver W. Science and complexity. *Am Sci.* 1948;36: 536–544. Available: <https://www.ncbi.nlm.nih.gov/pubmed/18882675>
2. Gershenson C. The Implications of Interactions for Science and Philosophy. *arXiv [nlin.AO]*. 2013. pp. 781–790. doi:10.1007/s10699-012-9305-8
3. Wright C, Bechtel W. - Mechanisms and psychological explanation. In: Thagard P, editor. *Philosophy of Psychology and Cognitive Science*. Amsterdam: North-Holland; 2007. pp. 31–79. doi:10.1016/B978-044451540-7/50019-0
4. Chan ME, Arvey RD. Meta-Analysis and the Development of Knowledge. *Perspect Psychol Sci.* 2012;7: 79–92. doi:10.1177/1745691611429355
5. Lawrence MG, Williams S, Nanz P, Renn O. Characteristics, potentials, and challenges of transdisciplinary research. *One Earth.* 2022;5: 44–61. doi:10.1016/j.oneear.2021.12.010
6. Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The extent and consequences of p-hacking in science. *PLoS Biol.* 2015;13: e1002106. doi:10.1371/journal.pbio.1002106
7. Rivera H. Fake Peer Review and Inappropriate Authorship Are Real Evils. *J Korean Med Sci.* 2019;34: e6. doi:10.3346/jkms.2019.34.e6
8. Scott JC. The Trouble with the View from Above. 2010 [cited 30 Jan 2022]. Available: <https://www.cato-unbound.org/2010/09/08/james-c-scott/trouble-view-above/>
9. Fanelli D. “Positive” results increase down the Hierarchy of the Sciences. *PLoS One.* 2010;5: e10068. doi:10.1371/journal.pone.0010068

10. Vazire S. Implications of the Credibility Revolution for Productivity, Creativity, and Progress. *Perspect Psychol Sci.* 2018;13: 411–417.  
doi:10.1177/1745691617751884
11. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A manifesto for reproducible science. *Nature Human Behaviour.* 2017;1: 1–9. doi:10.1038/s41562-016-0021
12. Reproducibility and reliability of biomedical research. [cited 26 Feb 2022]. Available: <https://acmedsci.ac.uk/policy/policy-projects/reproducibility-and-reliability-of-biomedical-research>
13. Hills TT. The Dark Side of Information Proliferation. *Perspect Psychol Sci.* 2018; 1745691618803647. doi:10.1177/1745691618803647
14. Barker Bausell R. *The Problem with Science: The Reproducibility Crisis and What to do About It.* Oxford University Press; 2021. Available: <https://play.google.com/store/books/details?id=oHEWEAAAQBAJ>
15. Bar-Yam Y. Multiscale variety in complex systems. *Complexity.* 2004;9: 37–45. doi:10.1002/cplx.20014
16. Bruni LE, Giorgi F. Towards a heterarchical approach to biology and cognition. *Prog Biophys Mol Biol.* 2015;119: 481–492. doi:10.1016/j.pbiomolbio.2015.07.005
17. Feinerman O, Korman A. Individual versus collective cognition in social insects. *J Exp Biol.* 2017;220: 73–82. doi:10.1242/jeb.143891
18. Rossi F, Bandyopadhyay S, Wolf M, Pavone M. Review of Multi-Agent Algorithms for Collective Behavior: a Structural Taxonomy. *IFAC-PapersOnLine.* 2018;51: 112–117. doi:10.1016/j.ifacol.2018.07.097
19. Mohanty D. Ethereum Architecture. In: Mohanty D, editor. *Ethereum for Architects and Developers: With Case Studies and Code Samples in Solidity.* Berkeley, CA: Apress; 2018. pp. 37–54. doi:10.1007/978-1-4842-4075-5\_2
20. Nabben K. Decentralised Autonomous Organisations (DAOs) as Data Trusts: A General-purpose Data Governance Framework for Decentralised Data Ownership, Storage, and Utilisation. 2021. doi:10.2139/ssrn.4009205
21. Schneider N, De Filippi P, Frey S, Tan JZ, Zhang AX. Modular Politics: Toward a Governance Layer for Online Communities. *Proc ACM Hum-Comput Interact.* 2021;5: 1–26. doi:10.1145/3449090
22. Gencer AE, Basu S, Eyal I, van Renesse R, Sirer EG. Decentralization in Bitcoin and Ethereum Networks. *arXiv [cs.CR].* 2018. Available: <http://arxiv.org/abs/1801.03998>

23. Fields C, Levin M. How Do Living Systems Create Meaning? *Philosophies*. 2020;5: 36. doi:10.3390/philosophies5040036
24. Veissière SPL, Constant A, Ramstead MJD, Friston KJ, Kirmayer LJ. Thinking through other minds: A variational approach to cognition and culture. *Behav Brain Sci*. 2019;43: 1–97. doi:10.1017/S0140525X19001213
25. Albarracin M, Demekas D, Ramstead M, Heins C. Epistemic Communities under Active Inference. 2022. Available: <https://www.preprints.org/manuscript/202201.0124>
26. Bruineberg J, Kiverstein J, Rietveld E. The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese*. 2018;195: 2417–2444. doi:10.1007/s11229-016-1239-1
27. Constant A, Clark A, Friston KJ. Representation Wars: Enacting an Armistice Through Active Inference. *Front Psychol*. 2020;11: 598733. doi:10.3389/fpsyg.2020.598733
28. Hipólito I. Cognition Without Neural Representation: Dynamics of a Complex System. *Front Psychol*. 2021;12: 643276. doi:10.3389/fpsyg.2021.643276
29. Guénin--Carlut A. Thinking like a State - Embodied intelligence in the deep history of our collective minds. 2021 [cited 25 Jan 2022]. doi:10.31219/osf.io/dxnzt
30. Buterin V. The meaning of decentralization. *Medium com*. 2017;6.
31. Nabben K. Is a DAO a Panopticon? Algorithmic Governance as Creating and Mitigating Vulnerabilities in “Decentralised Autonomous Organisations.” 2021. doi:10.2139/ssrn.3907693
32. Kiayias A, Lazos P. SoK: Blockchain Governance. *arXiv [cs.CR]*. 2022. Available: <http://arxiv.org/abs/2201.07188>
33. Graber J. Web3 is Self-Certifying. In: *Medium [Internet]*. 23 Dec 2021 [cited 26 Feb 2022]. Available: <https://jaygraber.medium.com/web3-is-self-certifying-9dad77fd8d81>
34. Benet J. IPFS - Content Addressed, Versioned, P2P File System. *arXiv [cs.NI]*. 2014. Available: <http://arxiv.org/abs/1407.3561>
35. IPFS powers the distributed web. [cited 15 Feb 2022]. Available: <https://ipfs.io/>
36. Tenorio-Fornés Á, Tirador EP, Sánchez-Ruiz AA, Hassan S. Decentralizing science: Towards an interoperable open peer review ecosystem using blockchain. *Inf Process Manag*. 2021;58: 102724. doi:10.1016/j.ipm.2021.102724

37. Hamburg S. A Guide to DeSci, the Latest Web3 Movement. In: Future [Internet]. 9 Feb 2022 [cited 25 Feb 2022]. Available: <https://future.a16z.com/what-is-decentralized-science-aka-desci/>
38. Cordes RJ, Applegate-Swanson S, Friedman DA, Knight VB, Mikhailova A. Narrative Information Management. Zenodo. COGSEC; 2021. doi:10.5281/zenodo.5573287
39. From Theory to Practice. The MIT Press; 2007. Available: <http://www.jstor.org/stable/j.ctt5hhdf6>
40. Wilson DS, Ostrom E, Cox ME. Generalizing the core design principles for the efficacy of groups. *J Econ Behav Organ*. 2013;90: S21–S32. doi:10.1016/j.jebo.2012.12.010
41. Paul W.B. Atkins, David Sloan Wilson, Steven C. Hayes. Prosocial. 2019. Available: <https://www.newharbinger.com/9781684030248/prosocial/>
42. Wilson DS, Philip MM, MacDonald IF, Atkins PWB, Kniffin KM. Core design principles for nurturing organization-level selection. *Sci Rep*. 2020;10: 13989. doi:10.1038/s41598-020-70632-8
43. Beck J. What is Web3? Here Are Some Ways To Explain It To A Friend. In: ConsenSys [Internet]. 12 Jan 2022 [cited 22 Feb 2022]. Available: <https://consensys.net/blog/blockchain-explained/what-is-web3-here-are-some-ways-to-explain-it-to-a-friend/>
44. Edelman G. The Father of Web3 Wants You to Trust Less. *Wired*. 29 Nov 2021. Available: <https://www.wired.com/story/web3-gavin-wood-interview/>. Accessed 26 Feb 2022.
45. ÐApps: What Web 3.0 Looks Like. [cited 22 Feb 2022]. Available: <https://gavwood.com/dappsweb3.html>
46. Kim HM, Laskowski M, Nan N. A First Step in the Co-Evolution of Blockchain and Ontologies: Towards Engineering an Ontology of Governance at the Blockchain Protocol Level. *arXiv [cs.CY]*. 2018. Available: <http://arxiv.org/abs/1801.02027>
47. Graeber D. Debt : The First 5000 Years. In: Melville House Books [Internet]. 2011 [cited 18 Feb 2022]. Available: <https://www.mhpbooks.com/books/debt/>
48. clepp. Molecule’s biopharma IP-NFTs — A technical description. In: Molecule Blog [Internet]. 6 Aug 2021 [cited 15 Feb 2022]. Available: <https://medium.com/molecule-blog/molecules-biopharma-ipnfts-a-technical-description-4dcfc6bf77f8>

49. Bartoletti M, Pompianu L. An empirical analysis of smart contracts: platforms, applications, and design patterns. arXiv [cs.CR]. 2017. Available: <http://arxiv.org/abs/1703.06322>
50. Khan SN, Loukil F, Ghedira-Guegan C, Benkhelifa E, Bani-Hani A. Blockchain smart contracts: Applications, challenges, and future trends. Peer Peer Netw Appl. 2021; 1–25. doi:10.1007/s12083-021-01127-0
51. Shorish J, Stephenson M, Zargham M. A Practical Theory Of Fungibility. WU Vienna University of Economics and Business; 2021 May. Report No.: 8137. Available: <https://ideas.repec.org/p/wiw/wus051/8137.html>
52. Nandwani A, Gupta M, Thakur N. Proof-of-Participation: Implementation of Proof-of-Stake Through Proof-of-Work. International Conference on Innovative Computing and Communications. Springer Singapore; 2019. pp. 17–24. doi:10.1007/978-981-13-2324-9\_3
53. poapxyz. POAP. [cited 1 Mar 2022]. Available: <https://www.poap.xyz/>
54. Peters G, Panayi E, Chapelle A. Trends in Cryptocurrencies and Blockchain Technologies: A Monetary Theory and Regulation Perspective. 2015. Available: <https://papers.ssrn.com/abstract=3084011>
55. Carson KA. Exodus. 2021. Available: <https://theanarchistlibrary.org/library/kevin-a-carson-exodus?v=1624008858>
56. Carter I. Positive and Negative Liberty. Zalta EN, editor. The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University; 2022. Available: <https://plato.stanford.edu/archives/spr2022/entries/liberty-positive-negative/>
57. Duniter. In: Duniter [Internet]. [cited 1 Mar 2022]. Available: <https://duniter.org/>
58. Analysis D. GRIDCOIN WHITE PAPER. Available: <https://gridcoin.us/assets/docs/whitepaper.pdf>
59. McConaghy T. Towards a Practice of Token Engineering. In: Ocean Protocol [Internet]. 1 Mar 2018 [cited 19 Jan 2022]. Available: <https://blog.oceanprotocol.com/towards-a-practice-of-token-engineering-b02feeff7ca>
60. Tools for the Web3 Data Economy. In: Ocean Protocol [Internet]. [cited 22 Jan 2022]. Available: <https://oceanprotocol.com/>
61. Smart contract research forum. In: Smart Contract Research Forum [Internet]. [cited 15 Feb 2022]. Available: <https://www.smartcontractresearch.org/>

62. DeSci Foundation. Why we need to fundamentally rethink scientific publishing. In: Medium [Internet]. 1 Dec 2021 [cited 15 Feb 2022]. Available: <https://desci.medium.com/why-we-need-to-fundamentally-rethink-scientific-publishing-43f2ae39af76>
63. Trovò B, Massari N. Ants-Review: A Privacy-Oriented Protocol for Incentivized Open Peer Reviews on Ethereum. Euro-Par 2020: Parallel Processing Workshops. Springer International Publishing; 2021. pp. 18–29. doi:10.1007/978-3-030-71593-9\_2
64. Avital M. Peer Review: Toward a Blockchain-enabled Market-based Ecosystem. Communications of the Association for Information Systems. 2018;42: 28. doi:10.17705/1CAIS.04228
65. van Rossum J. The blockchain and its potential for science and academic publishing. Inf Serv Use. 2018;38: 95–98. doi:10.3233/isu-180003
66. The ResearchCoin Whitepaper. In: ResearchHub [Internet]. [cited 15 Feb 2022]. Available: <https://www.researchhub.com/paper/819400/the-researchcoin-whitepaper>
67. Tenorio-Fornés A, Jacynycz V, Llop-Vila D, Sánchez-Ruiz A, Hassan S. Towards a Decentralized Process for Scientific Publication and Peer Review using Blockchain and IPFS. 2019. doi:10.24251/HICSS.2019.560
68. Jan Z, Third A, Ibanez L-D, Bachler M, Simperl E, Domingue J. ScienceMiles: Digital Currency for Researchers. Companion Proceedings of the The Web Conference 2018. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee; 2018. pp. 1183–1186. doi:10.1145/3184558.3191556
69. Ethereum. Ethereum Whitepaper. In: ethereum.org [Internet]. [cited 14 Feb 2022]. Available: <https://ethereum.org/en/whitepaper/>
70. Williams S, Berman L, Raybould I, Uemlianin I. Arweave: A protocol for economically sustainable information permanence. [cited 15 Feb 2022]. Available: <https://www.arweave.org/yellow-paper.pdf>
71. Explanation of DAICOs. In: Ethereum Research [Internet]. 6 Jan 2018 [cited 30 Jan 2022]. Available: <https://ethresear.ch/t/explanation-of-daicos/465>
72. Science Fund. [cited 28 Feb 2022]. Available: <https://www.sciencefund.io/>
73. El Damaty S. Rich in Data, Poor in Wisdom: Science Needs a Decentralized Data Commons. In: OpSci [Internet]. 4 Dec 2021 [cited 27 Feb 2022]. Available:



- <https://pulse.opsci.io/rich-in-data-poor-in-wisdom-science-needs-a-decentralized-data-commons-98c7ffdb56a1>
74. Ocean Protocol Foundation with BigchainDB GmbH. Ocean Protocol: Tools for the Web Data Economy. Available: <https://oceanprotocol.com/tech-whitepaper.pdf>
  75. Datum. Datum-WhitePaper.pdf. Available: <https://datum.org/assets/Datum-WhitePaper.pdf>
  76. Dataspace. datapace\_whitepaper.pdf. Available: [https://www.datapace.io/datapace\\_whitepaper.pdf](https://www.datapace.io/datapace_whitepaper.pdf)
  77. The Problem of Data: Data Management and Curation Practices Among University Researchers. In: CLIR [Internet]. 13 Dec 2016 [cited 18 Feb 2022]. Available: <https://www.clir.org/pubs/reports/pub154/problem-of-data/>
  78. Gorgolewski KJ, Auer T, Calhoun VD, Craddock RC, Das S, Duff EP, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci Data*. 2016;3: 160044. doi:10.1038/sdata.2016.44
  79. Gregory K, Groth P, Scharnhorst A, Wyatt S. Lost or found? Discovering data needed for research. *Harvard Data Science Review*. 2020. doi:10.1162/99608f92.e38165eb
  80. Kitano H. Nobel Turing Challenge: creating the engine for scientific discovery. *NPJ Syst Biol Appl*. 2021;7: 29. doi:10.1038/s41540-021-00189-3
  81. The Turing AI scientist grand challenge. In: The Alan Turing Institute [Internet]. [cited 16 Feb 2022]. Available: <https://www.turing.ac.uk/research/research-projects/turing-ai-scientist-grand-challenge>
  82. SEBoK. [cited 24 Jan 2022]. Available: [https://sebokwiki.org/wiki/Guide\\_to\\_the\\_Systems\\_Engineering\\_Body\\_of\\_Knowledge\\_\(SEBoK\)](https://sebokwiki.org/wiki/Guide_to_the_Systems_Engineering_Body_of_Knowledge_(SEBoK))
  83. Vyatkin A, Metelkin I, Mikhailova A, Cordes RJ, Friedman DA. Active Inference & Behavior Engineering for Teams. 2020. doi:10.5281/zenodo.4021163
  84. Cordes RJ, Friedman DA, Maron M. Reimagining Maps. 2020. doi:10.5281/zenodo.4170026
  85. ISO/IEC/IEEE 15288:2015. In: ISO [Internet]. 2020 [cited 23 Feb 2022]. Available: <https://www.iso.org/standard/63711.html>
  86. Essence 1.2 - SEMAT. [cited 23 Feb 2022]. Available: <http://semat.org/essence-1.2>

87. Weightman G. The History of the Bar Code. In: smithsonianmag.com [Internet]. 23 Sep 2015 [cited 28 Jan 2022]. Available: <https://www.smithsonianmag.com/innovation/history-bar-code-180956704/>
88. Silverman L. Barcodes: A Brief History. In: corp.trackabout.com [Internet]. 19 Nov 2015 [cited 28 Jan 2022]. Available: <https://corp.trackabout.com/blog/barcodes-brief-history>
89. David S, Cordes RJ, Friedman DA. Active Inference in Modeling Conflict. 2021. doi:10.5281/zenodo.5759807
90. Dikopoulou A, Mihiotis A. The contribution of records management to good governance. *The TQM Journal*. 2012;24: 123–141. doi:10.1108/17542731211215071
91. ISO. ISO 15489 - Information and documentation - Records Management. 2001.
92. Cordes RJ. The Pipeline. *The Pipeline: EdTech, Knowledge Management, and NATSEC*. COGSEC; 2021.
93. Cox RJ, Wallace DA, Wallace D. Archives and the Public Good: Accountability and Records in Modern Society. Greenwood Publishing Group; 2002. Available: <https://play.google.com/store/books/details?id=sH5BaIWT7o0C>
94. Tumiel J. Spinning up in active inference and the free energy principle. In: jared tumiel [Internet]. 14 Oct 2020 [cited 21 Jan 2022]. Available: <https://jaredtumiel.github.io/blog/2020/10/14/spinning-up-in-ai.html>
95. Friston K, Heins C, Ueltzhöffer K, Da Costa L, Parr T. Stochastic Chaos and Markov Blankets. *Entropy*. 2021;23. doi:10.3390/e23091220
96. Parr T, Pezzulo G, Friston KJ. Active Inference: The Free Energy Principle in Mind, Brain, and Behavior. MIT Press; 2022. Available: [https://scholar.google.ca/scholar?cluster=1067966391259043561&hl=en&as\\_sdt=0,5&sciodt=0,5](https://scholar.google.ca/scholar?cluster=1067966391259043561&hl=en&as_sdt=0,5&sciodt=0,5)
97. Sims M. How to count biological minds: symbiosis, the free energy principle, and reciprocal multiscale integration. *Synthese*. 2020. doi:10.1007/s11229-020-02876-w
98. Ramstead MJD, Kirchhoff MD, Constant A, Friston KJ. Multiscale integration: beyond internalism and externalism. *Synthese*. 2021;198: 41–70. doi:10.1007/s11229-019-02115-x
99. Friston K. A free energy principle for a particular physics. *arXiv [q-bio.NC]*. 2019. Available: <http://arxiv.org/abs/1906.10184>

100. Kuchling F, Friston K, Georgiev G, Levin M. Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems. *Phys Life Rev.* 2019. doi:10.1016/j.plrev.2019.06.001
101. Fields C, Levin M. Scale-Free Biology: Integrating Evolutionary and Developmental Thinking. *Bioessays.* 2020;42: e1900228. doi:10.1002/bies.201900228
102. Friston K. Active inference and agency. *Cognitive neuroscience.* 2014. pp. 119–121. doi:10.1080/17588928.2014.905517
103. Pezzulo G, Rigoli F, Friston K. Active Inference, homeostatic regulation and adaptive behavioural control. *Prog Neurobiol.* 2015;134: 17–35. doi:10.1016/j.pneurobio.2015.09.001
104. Hipólito I, Baltieri M, Friston K, Ramstead MJD. Embodied skillful performance: where the action is. *Synthese.* 2021;199: 4457–4481. doi:10.1007/s11229-020-02986-5
105. Marković D, Goschke T, Kiebel SJ. Meta-control of the exploration-exploitation dilemma emerges from probabilistic inference over a hierarchy of time scales. *Cogn Affect Behav Neurosci.* 2021;21: 509–533. doi:10.3758/s13415-020-00837-x
106. Da Costa L, Parr T, Sajid N, Veselic S, Neacsu V, Friston K. Active inference on discrete state-spaces: A synthesis. *J Math Psychol.* 2020;99: 102447. doi:10.1016/j.jmp.2020.102447
107. Vasil J, Badcock PB, Constant A, Friston K, Ramstead MJD. A World Unto Itself: Human Communication as Active Inference. *Front Psychol.* 2020;11: 417. doi:10.3389/fpsyg.2020.00417
108. Schoeller F, Miller M, Salomon R, Friston K. Trust as Extended Control: Active Inference and User Feedback During Human-Robot Collaboration. *ArXiv.* 2021 [cited 21 Sep 2021]. Available: <https://www.semanticscholar.org/paper/7c39921b6050d7be82e18df2a31973f2c3f864ea>
109. Heyes C. Précis of Cognitive Gadgets: The Cultural Evolution of Thinking. *Behav Brain Sci.* 2019;42. doi:10.1017/S0140525X18002145
110. Muthukrishna M, Doebeli M, Chudek M, Henrich J. The Cultural Brain Hypothesis: How culture drives brain expansion, sociality, and life history. *PLoS Comput Biol.* 2018;14: e1006504. doi:10.1371/journal.pcbi.1006504
111. Graeber D, Wengrow D. The dawn of everything -A new history of humanity. 2021 [cited 25 Feb 2022]. Available: <https://uujec.org/Portals/0/The%20Dawn%20of%20Everything.pdf>

112. Doyle MJ, Marsh L. Stigmergy 3.0: From ants to economies. *Cogn Syst Res*. 2013;21: 1–6. doi:10.1016/j.cogsys.2012.06.001
113. Correia L, Sebastião AM, Santana P. On the role of stigmergy in cognition. *Progress in Artificial Intelligence*. 2017;6: 79–86. doi:10.1007/s13748-016-0107-z
114. Ramstead MJD, Constant A, Badcock PB, Friston KJ. Variational ecology and the physics of sentient systems. *Phys Life Rev*. 2019. doi:10.1016/j.pprev.2018.12.002
115. Constant A, Ramstead MJD, Veissière SPL, Campbell JO, Friston KJ. A variational approach to niche construction. *J R Soc Interface*. 2018;15: 20170685. doi:10.1098/rsif.2017.0685
116. Friedman D, Tschantz A, Ramstead MJD, Friston K, Constant A. Active inferants: The basis for an active inference framework for ant colony behavior. *Front Behav Neurosci*. 2021;15: 126. doi:10.3389/fnbeh.2021.647732
117. Guénin--Carlut A. Cognitive agency in sociocultural evolution. 2022. doi:10.31219/osf.io/x7yr4
118. Coelho FC, Brandão A. Decentralising scientific publishing: can the blockchain improve science communication? *Mem Inst Oswaldo Cruz*. 2019;114: e190257. doi:10.1590/0074-02760190257
119. Brian Arthur W. Economics in Nouns and Verbs. *arXiv [econ.GN]*. 2021. Available: <http://arxiv.org/abs/2104.01868>
120. Fields C, Friston K, Glazebrook JF, Levin M. A free energy principle for generic quantum systems. *arXiv [quant-ph]*. 2021. Available: <http://arxiv.org/abs/2112.15242>
121. Boik JC. Science-Driven Societal Transformation, Part I: Worldview. *Sustain Sci Pract Policy*. 2020;12: 6881. doi:10.3390/su12176881
122. Rubin S, Parr T, Da Costa L, Friston K. Future climates: Markov blankets and active inference in the biosphere. *J R Soc Interface*. 2020;17: 20200503. doi:10.1098/rsif.2020.0503
123. Friston KJ, Parr T, de Vries B. The graphical brain: Belief propagation and active inference. *Netw Neurosci*. 2017;1: 381–414. doi:10.1162/NETN\_a\_00018
124. van de Laar TW, de Vries B. Simulating Active Inference Processes by Message Passing. *Frontiers in Robotics and AI*. 2019;6: 20. doi:10.3389/frobt.2019.00020
125. ForneyLab.jl: Julia package for automatically generating Bayesian inference algorithms through message passing on Forney-style factor graphs. Github; Available: <https://github.com/biaslab/ForneyLab.jl>

126. Friston K, Douglass DS, Iennaco de Vasconcelos ML, Sillett S, Sganzerla L, Tickles D, et al. Transcript of: Karl Friston, 1st Applied Active Inference Symposium, Active Inference Lab, June 21, 2021. 2021. doi:10.5281/zenodo.5797041
127. Friston K, Da Costa L, Sajid N, Heins C, Ueltzhöffer K, Pavliotis GA, et al. The free energy principle made simpler but not too simple. arXiv [cond-mat.stat-mech]. 2022. Available: <http://arxiv.org/abs/2201.06387>
128. Kirchhoff M, Parr T, Palacios E, Friston K, Kiverstein J. The Markov blankets of life: autonomy, active inference and the free energy principle. J R Soc Interface. 2018;15: 20170792. doi:10.1098/rsif.2017.0792
129. cadCAD – A Python package for designing, testing and validating complex systems through simulation. [cited 15 Feb 2022]. Available: <https://cadcad.org/>
130. darcspice: EVM agent-based token simulator 🧠🔗. Github; Available: <https://github.com/opscientia/darcspice>
131. CRediT - Contributor Roles Taxonomy. In: CASRAI [Internet]. 14 Sep 2019 [cited 28 Feb 2022]. Available: <https://casrai.org/credit/>

## Chapter VII Works Cited

---

1. Alberts, David Stephen. Power to the Edge: Command and Control in the Information Age. CCRP. 2003.
2. Alexander, Scott. Meditations on Moloch. Slate Star Codex.
3. Ashby, W.R. Requisite Variety and its Implications for the Control of Complex Systems. Springer. 1991.
4. Bach, Joscha. Principles of Synthetic Intelligence Building Blocks for an Architecture of Motivated Cognition
5. Bejan, Adrian. Freedom and Evolution: Hierarchy in Nature, Society and Science. Springer. 2020.
6. Buterin, Vitalik. Ethereum. Ethereum Whitepaper. In: ethereum.org. Available:<https://ethereum.org/en/whitepaper>
7. Bostrom, Nick. Global Catastrophic Risk. Oxford University press. 2008.
8. Coase, Ronald. The Nature of the Firm. Economica, 1937
9. Dawkins, R. The Selfish Gene. Oxford University Press. 1989.
10. Duettmann, Allison et al. Gaming the Future: Technologies for Intelligent Voluntary Cooperation. Substack. 2022.
11. Elliot, M.A. Stigmergic collaboration: a theoretical framework for mass collaboration. PhD thesis, University of Melbourne. 2007.
12. Friedman, Daniel et al. An Active Inference Ontology for Decentralised Science: from Situated Sensemaking to the Epistemic Commons. Zenodo. 2020.
13. Froehlich, Sara. Notes on Non-Cooperative Game Theory and the Core. Maths, McGill. 2019.

14. Friston, Karl. A Free Energy Principle for a Particular Physics. arXiv. 2019.
15. Friedman, David. Legal Systems Very Different from Ours. 2012.
16. Guénin Carlut, Avel. Thinking Like a State - Embodied Intelligence in the Deep History of Our Collective Minds. OSF Preprints. 2021.
17. Hagens, Nate. Economics for the future – Beyond the superorganism. Ecological Economics, Vol 169. 2020.
18. Huxley, Aldous. Brave New World Revisited. Harper. 1958.
19. Read, Leonard. I, Pencil
20. Skinner, B.F. Beyond Freedom and Dignity. Hackett. 1971.
21. Schelling, Thomas. The Strategy of Conflict.
22. Tainter, Joseph. The Collapse of Complex Societies. Cambridge University Press. 1988.
23. Nielsen, Micheals & Gowers, Timothy. Massively collaborative mathematics. Nature. Vol 461. 2009.
24. Popper, Karl. The Open Society and Its Enemies. Princeton University Press, 1971.
25. Minsky, Marvin. Society of mind. Simon and Schuster. 1988.
26. Miller, Mark and Drexler, Erik. Markets and Computation: Agoric Open Systems.
27. Minsky, Marvin. The Society of Mind. Simon & Schuster. 1986
28. Taleb, Nassim. Antifragile. Penguin Books. 2012.
29. Szabo, Nick. computational contracts
30. Wilson, E.O. The Evolution of Eusociality. Nature. 2012.
31. Xuesen, Y. Jingyuan & D. Ruwei. A new discipline of science: The study of open complex giant systems and its methodology. Journal of Systems Engineering and Electronics. 1993.

## Chapter VIII Works Cited

---

1. University of Washington IRSIRI. NSF Convergence Accelerator: Verified Information Exchange Environments Call for Submissions. 2021 Nov.
2. Milgrom PR, Roberts J. Economics, Organization, and Management. Prentice-Hall; 1992.
3. McMillan J. Reinventing the Bazaar: A Natural History of Markets. W. W. Norton & Company; 2003.
4. Karachiwalla R, Pinkow F. Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. *Creat Innov Manag*. 2021;30: 563–584.
5. Rosenbaum EF. What is a Market? On the Methodology of a Contested Concept. *Rev Soc Econ*. 2000;58: 455–482.
6. Simmel G. The philosophy of money. Routledge; 2004.
7. Nietzsche FW, Hollingdale RJ. On the Genealogy of Morals. Vintage Books; 1989.
8. Kaplan HS, Schniter E, Smith VL, Wilson BJ. Risk and the evolution of human exchange. *Proc Biol Sci*. 2012;279: 2930–2935.
9. Hill KR, Walker RS, Bozicević M, Eder J, Headland T, Hewlett B, et al. Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science*. 2011;331: 1286–1289.
10. Isaac G. The food-sharing behavior of protohuman hominids. *Sci Am*. 1978;238: 90–108.
11. Orlan A. The Empire of Value: A New Foundation for Economics. MIT Press; 2014.



12. Cibulski L, Wascher CAF, Weiss BM, Kotrschal K. Familiarity with the experimenter influences the performance of Common ravens (*Corvus corax*) and Carrion crows (*Corvus corone corone*) in cognitive tasks. *Behav Processes*. 2014;103: 129–137.
13. Brosnan SF, De Waal FBM. Monkeys reject unequal pay. *Nature*. 2003;425: 297–299.
14. Hojo MK, Pierce NE, Tsuji K. Lycaenid Caterpillar Secretions Manipulate Attendant Ant Behavior. *Curr Biol*. 2015;25: 2260–2264.
15. Noë R, Hammerstein P. Biological markets. *Trends Ecol Evol*. 1995;10: 336–339.
16. McDonald M. *Market Segmentation: How to Do It and How to Profit from It*. John Wiley & Sons; 2012.
17. World Bank. *World development report 1996 from plan to market: From plan to market*. New York, NY: Oxford University Press; 1996.
18. Ioannides YM. Evolution of Trading Structures. In: Arthur WB, Durlauf SN, Lane DA, editors. *The Economy as an Evolving Complex System II*. Santa Fe Institute; 1996.
19. Zadeh LA. Fuzzy sets. *Information and Control*. 1965;8: 338–353.
20. Zimmermann H-J. Fuzzy set theory. *Wiley Interdiscip Rev Comput Stat*. 2010;2: 317–332.
21. Maiers J, Sherif YS. Applications of fuzzy set theory. *IEEE Trans Syst Man Cybern*. 1985;SMC-15: 175–189.
22. Ragin CC. *Fuzzy-Set Social Science*. University of Chicago Press; 2000.
23. Ankei Y. Fish as “Primitive Money”: Barter Markets of the Songola. *Senri Ethnol Stud*. 1984. Available: [https://minpaku.repo.nii.ac.jp/?action=repository\\_action\\_common\\_download&item\\_id=3324&item\\_no=1&attribute\\_id=18&file\\_no=1](https://minpaku.repo.nii.ac.jp/?action=repository_action_common_download&item_id=3324&item_no=1&attribute_id=18&file_no=1)
24. Haklai M. Visualising Roman Institutional Environments for Exchange as a Complex System. In: Verboven K, editor. *Complexity Economics: Building a New Approach to Ancient Economic History*. Cham: Springer International Publishing; 2021. pp. 125–159.
25. David S, Cordes RJ, Friedman DA. *Active Inference in Modeling Conflict*. 2021. doi:10.5281/zenodo.5759807
26. Cordes RJ, Applegate-Swanson S, Friedman DA, Knight VB, Mikhailova A. *Narrative Information Management*. Zenodo. COGSEC; 2021. doi:10.5281/zenodo.5573287

27. Way MJ. Mutualism Between Ants and Honeydew-Producing Homoptera. *Annu Rev Entomol.* 1963;8: 307–344.
28. Fullbrook E. *Pluralist Economics*. Bloomsbury Academic; 2008.
29. Verboven K, editor. *Complexity Economics: Building a New Approach to Ancient Economic History*. Springer Nature; 2020.
30. Lee F. *Heterodox Economics. The Long Term View*. 2008;7.
31. Colander D, Föllmer H, Haas A, Goldberg MD, Juselius K, Kirman A, et al. The Financial Crisis and the Systemic Failure of Academic Economics. Kiel Institute for the World Economy; 2009. doi:10.2139/ssrn.1355882
32. Colander D. Pluralism and Heterodox Economics: Suggestions for an “Inside the Mainstream” Heterodoxy. Middlebury College, Department of Economics; 2007 Oct. Report No.: 0724. Available: <https://ideas.repec.org/p/mdl/mdlpap/0724.html>
33. Savageau MA. Growth of complex systems can be related to the properties of their underlying determinants. *Proc Natl Acad Sci U S A.* 1979;76: 5413–5417.
34. Gordon DM. The ecology of collective behavior. *PLoS Biol.* 2014;12: e1001805.
35. Gell-Mann M. *The Quark and the Jaguar: Adventures in the Simple and the Complex*. Macmillan; 1995.
36. Cordes RJ, Friedman DA. Emergent Teams for Complex Threats. *The Great Preset: Remote Teams and Operational Art*. COGSEC; 2020.
37. Gell-Mann M. Plectics: The study of simplicity and complexity. *Europhys News.* 2009;33: 17–20.
38. Page SE. A complexity perspective on institutional design. *Politics, Philosophy & Economics.* 2012;11: 5–25.
39. Arthur WB, Durlauf SN, Lane DA, editors. *The Economy as an Evolving Complex System II*. Santa Fe Institute; 1996.
40. Lankenau SE. Smoke’em if you got’em: Cigarette black markets in US prisons and jails. *Prison J.* 2001;81: 142–161.
41. Hammerstein P. *Genetic and Cultural Evolution of Cooperation*. MIT Press; 2003.
42. Smith MH. 3. Weaving Wealth: Cloth and Trade in Viking Age and Medieval Iceland. *Textiles and the Medieval Economy: Production, Trade and Consumption of Textiles, 8th-16th Centuries*. 2014; 23–40.

43. Benson BL. The Spontaneous Evolution of Commercial Law. *South Econ J.* 1989;55: 644–661.
44. Milgrom PR, North DC, Weingast\* BR. The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Econ Polit.* 1990;2: 1–23.
45. Reinert S, Fredona R. Merchants and the Origins of Capitalism. Business School BGIE Unit Working Paper. 2017. doi:10.2139/ssrn.3037173
46. Sgard J. Global economic governance during the middle ages: The jurisdiction of the champagne fairs. *Int Rev Law Econ.* 2015;42: 174–184.
47. Reinecke JE. Trade Jargons and Creole Dialects as Marginal Languages. *Soc Forces.* 1938;17: 107–118.
48. Hancock IF. *Readings in Creole Studies.* John Benjamins Publishing; 1979.
49. Vyatkin A, Metelkin I, Mikhailova A, Cordes RJ, Friedman DA. Active Inference and Behavior Engineering for Teams. In: Friedman DA, Cordes RJ, editors. *The Great Preset: Remote Teams and Operational Art.* COGSEC; 2020.
50. Virant-Doberlet M, Kuhelj A, Polajnar J, Šturm R. Predator-Prey Interactions and Eavesdropping in Vibrational Communication Networks. *Frontiers in Ecology and Evolution.* 2019;7. doi:10.3389/fevo.2019.00203
51. Virant-Doberlet M, Mazzoni V, de Groot M, Polajnar J, Lucchi A, Symondson WOC, et al. Vibrational Communication Networks: Eavesdropping and Biotic Noise. In: Cocroft RB, Gogala M, Hill PSM, Wessel A, editors. *Studying Vibrational Communication.* Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. pp. 93–123.
52. Johnson C. Exploring social markets, partner debt, and mimetic currency in dolphins. *Anim Behav Cogn.* 2016;3: 224–242.
53. Dietler M. Driven by drink: The role of drinking in the political economy and the case of Early Iron Age France. *Journal of Anthropological Archaeology.* 1990;9: 352–406.
54. Dunbar RIM. *Grooming, Gossip, and the Evolution of Language.* Harvard University Press; 1998.
55. Lang M, Xygalatas D, Kavanagh CM, Boccardi N, Halberstadt J, Jackson C, et al. Outgroup threat and the emergence of cohesive groups: A cross-cultural examination. *Group Process Intergroup Relat.* 2021; 136843022110169.
56. McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. *Annu Rev Sociol.* 2001;27: 415–444.

57. Forgas JP, Haselton MG, von Hippel W. Evolution and the Social Mind: Evolutionary Psychology and Social Cognition. Psychology Press; 2011.
58. Williams KD, Forgas JP, von Hippel W, editors. The social outcast: Ostracism, social exclusion, rejection, and bullying. Sydney Symposium of Social Psychology series. 2005;366. Available: <https://psycnet.apa.org/fulltext/2005-13813-000.pdf>
59. Whitehouse H, François P, Savage PE, Currie TE, Feeney KC, Cioni E, et al. Complex societies precede moralizing gods throughout world history. *Nature*. 2019;568: 226–229.
60. Whitehouse H, François P, Savage PE, Currie TE, Feeney KC, Cioni E, et al. Retraction Note: Complex societies precede moralizing gods throughout world history. *Nature*. 2021;595: 320.
61. Haidt J. The Righteous Mind: Why Good People Are Divided by Politics and Religion. Knopf Doubleday Publishing Group; 2012.
62. Holzman RS. SPIN-OFFS, SPLIT-UPS, AND SPLIT-OFFS. *Natl Tax J*. 1952;5: 277–278.
63. Mintz SS. Divisive Corporate Reorganizations: Split-Ups and Split-Offs. *Tax L Rev*. 1950. Available: [https://heinonline.org/hol-cgi-bin/get\\_pdf.cgi?handle=hein.journals/taxlr6&section=31](https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/taxlr6&section=31)
64. Escartín González E, Velasco Morente F, González Abril L. On Sevillian guilds towards the end of the 11th century. *Revista crítica de historia de las relaciones laborales y de la política social*, 7. 2013. Available: <https://idus.us.es/handle/11441/76311>
65. Greif A. Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition. *Am Econ Rev*. 1993;83: 525–548.
66. Greif A. Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders. *J Econ Hist*. 1989;49: 857–882.
67. Porter T, Ronit K. Self-Regulation as Policy Process: The Multiple and Criss-Crossing Stages of Private Rule-Making. *Policy Sci*. 2006;39: 41–72.
68. Kirton JJ, Trebilcock MJ. Hard choices, soft law: Voluntary standards in global trade, environment and social governance. Routledge; 2017.
69. Greif A. Institutions and the Path to the Modern Economy: Lessons from Medieval Trade. Cambridge University Press; 2006.
70. Greif A. The Maghribi traders: a reappraisal? *Econ Hist Rev*. 2012;65: 445–469.

71. Bernstein PL. *Against the Gods: The Remarkable Story of Risk*. John Wiley & Sons; 2012.
72. Wakita S. Efficiency of the Dojima rice futures market in Tokugawa-period Japan. *Journal of Banking & Finance*. 2001;25: 535–554.
73. Hamori S, Hamori N, Anderson DA. An empirical analysis of the efficiency of the Osaka rice market during japan’s Tokugawa era. *J Futures Mark*. 2001;21: 861–874.
74. Hofstadter D. *Godel, Escher, Bach: An eternal golden braid*. Basic Books; 1979.
75. Bernstein A, Hughson E, Weidenmier M. Counterparty Risk and the Establishment of the New York Stock Exchange Clearinghouse. *J Polit Econ*. 2019;127: 689–729.
76. Bray D, Croxson K, Dutton W. *Information markets: Feasibility and performance*. University of Oxford; 2008. Available:  
[https://www.oii.ox.ac.uk/archive/downloads/research/dpsn/Informationmarkets\\_full.pdf](https://www.oii.ox.ac.uk/archive/downloads/research/dpsn/Informationmarkets_full.pdf)
77. Dawkins R. *The Selfish Gene*. Oxford University Press; 2016.
78. Kylafis G, Loreau M. Ecological and evolutionary consequences of niche construction for its agent. *Ecol Lett*. 2008;11: 1072–1081.
79. Bain JS. Conditions of Entry and the Emergence of Monopoly. In: Chamberlin EH, editor. *Monopoly and Competition and their Regulation: Papers and Proceedings of a Conference held by the International Economic Association*. London: Palgrave Macmillan UK; 1954. pp. 215–241.
80. Demmig-Adams B, Cohu CM, Adams WW III. Dealing with the hazards of harnessing sunlight. *Nature Education*. 2012;4: 18.
81. Tainter J. *The Collapse of Complex Societies*. Cambridge University Press; 1988.
82. Hopkins K, Beard M, Beard R in C. *The Colosseum*. Harvard University Press; 2005.
83. Knapp AB, Manning SW. Crisis in Context: The End of the Late Bronze Age in the Eastern Mediterranean. *Am J Archaeol*. 2016;120: 99–149.
84. Kindleberger CP, Aliber RZ, Solow R. *Manias, Panics and Crashes*. Wiley; 2005.
85. Dalrymple W. The East India Company: the original corporate raiders. *Guardian*. 2015;4.

86. Clegg S. The East India Company: The First Modern Multinational? Multinational Corporations and Organization Theory: Post Millennium Perspectives. Emerald Publishing Limited; 2017. pp. 43–67.
87. Hess C, Ostrom E. Understanding Knowledge as a Commons: From Theory to Practice. MIT Press; 2011.
88. Wall D. The Commons in History: Culture, Conflict, and Ecology. MIT Press; 2014.
89. Ranelagh J. A Short History of Ireland. Cambridge University Press; 2012.
90. Simms K. The Poetic Brehon Lawyers of Early Sixteenth-Century Ireland. *Ériu*. 2007;57: 121–132.
91. Hore HF. Irish Brehons and Their Laws. *Ulster Journal of Archaeology*. 1857;5: 36–52.
92. Westropp HM. On the Tribal System and Land-Tenure in Ireland, under the Brehon Laws. *The Journal of the Ethnological Society of London* (1869-1870). 1870;2: 342–351.

## Chapter IX Works Cited

---

1. Cordes RJ, David S, Maan A, Ruiz A, Sapp E, Scannell P, et al. The Narrative Campaign Field Guide - First Edition. 1st ed. Cordes RJ, editor. Narrative Strategies Ink; 2021.
2. Disarm Foundation. DISARM Foundation. In: disarm.foundation [Internet]. [cited 5 Apr 2022]. Available: <https://www.disarm.foundation/>
3. COGSEC-Collaborative, DISARM Foundation. AMITT: Adversarial Misinformation and Influence Tactics and Techniques. Github; 2022. Available: <https://github.com/cogsec-collaborative/AMITT>
4. Varsos C, Flouris G, Bitsaki M, Fasoulakis M. A Study of Misinformation Games. PRICAI 2021: Trends in Artificial Intelligence. Springer International Publishing; 2021. pp. 76–87.
5. Koops MA. Misinformation and assessment uncertainty in the ecology of information use. mspace.lib.umanitoba.ca. 1998. Available: <https://mspace.lib.umanitoba.ca/handle/1993/1722>
6. Stowe MK, Turlings TC, Loughrin JH, Lewis WJ, Tumlinson JH. The chemistry of eavesdropping, alarm, and deceit. Proc Natl Acad Sci U S A. 1995;92: 23–28.
7. Virant-Doberlet M, Mazzoni V, de Groot M, Polajnar J, Lucchi A, Symondson WOC, et al. Vibrational Communication Networks: Eavesdropping and Biotic Noise. In: Cocroft RB, Gogala M, Hill PSM, Wessel A, editors. Studying Vibrational Communication. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. pp. 93–123.
8. Hanley JE, Orbell J, Morikawa T. The cost of misinformation in deadly conflicts: Hawk-Dove games and suicidal terrorism. Politics Life Sci. 2002;21: 13–18.
9. Cabral LMB. Umbrella branding with imperfect observability and moral hazard. Int J Ind Organiz. 2009;27: 206–213.

10. Glowacki A. From narratives of violence to narratives of peace: The renunciation of violence as a discursive phenomenon. 2013. Available: <https://search.proquest.com/openview/12b2e7d7d9476024a27c66140ecff86/1?pq-origsite=gscholar&cbl=18750>
11. Mascarenhas M, Cordes RJ, Friedman DA. Digital Rhetorical Ecosystem Analysis: Sensemaking of Digital Memetic Discourse. Zenodo. 2021. doi:10.5281/zenodo.5573947
12. Meadow CT, Yuan W. Measuring the impact of information: Defining the concepts. *Inf Process Manag.* 1997;33: 697–714.
13. Davis JJ. *Advertising Research: Theory & Practice* (2nd Edition). 2nd ed. Pearson; 2019.
14. David S, Cordes RJ, Friedman DA. Active Inference in Modeling Conflict. 2021. doi:10.5281/zenodo.5759807
15. Borkovsky RN, Goldfarb A, Haviv AM, Moorthy S. Measuring and Understanding Brand Value in a Dynamic Model of Brand Management. *Marketing Science.* 2017;36: 471–499.
16. Kamakura WA, Russell GJ. Measuring brand value with scanner data. *International Journal of Research in Marketing.* 1993;10: 9–22.
17. Milgrom P. Auctions and Bidding: A Primer. *J Econ Perspect.* 1989;3: 3–22.
18. Teschner F, Stathel S, Weinhardt C. A Prediction Market for Macro-Economic Variables. 2011 44th Hawaii International Conference on System Sciences. [ieeexplore.ieee.org](http://ieeexplore.ieee.org); 2011. pp. 1–9.
19. McAfee RP, McMillan J. Auctions and Bidding. *J Econ Lit.* 1987;25: 699–738.
20. Vickrey W. Counterspeculation, Auctions, and Competitive Sealed Tenders. *J Finance.* 1961;16: 8–37.
21. Rothkopf MH, Teisberg TJ, Kahn EP. Why Are Vickrey Auctions Rare? *J Polit Econ.* 1990;98: 94–109.
22. Cordes RJ, Applegate-Swanson S, Friedman DA, Knight VB, Mikhailova A. Narrative Information Management. Zenodo. COGSEC; 2021. doi:10.5281/zenodo.5573287



## Chapter X Works Cited

---

1. Roose K. A QAnon “Digital Soldier” Marches On, Undeterred by Theory’s Unraveling. The New York Times. 17 Jan 2021. Available: <https://www.nytimes.com/2021/01/17/technology/qanon-meme-queen.html>. Accessed 14 Jul 2022.
2. Castaño Díaz. Defining and characterizing the concept of Internet Meme. CES Psicol. 2013. Available: [http://www.scielo.org.co/scielo.php?pid=S2011-30802013000200007&script=sci\\_arttext&lng=en](http://www.scielo.org.co/scielo.php?pid=S2011-30802013000200007&script=sci_arttext&lng=en)
3. Molina. What makes an internet meme a meme? Six essential characteristics. Handbook of visual communication. 2020. doi:10.4324/9780429491115-35/makes-internet-meme-meme-six-essential-characteristics-maria-molina
4. Dawkins R. The Selfish Gene. Oxford University Press; 2016.
5. Blackmore S. The Meme Machine. OUP Oxford; 2000.
6. Dennett DC. Darwin’s Dangerous Idea: Evolution and the Meanings of Life. Simon and Schuster; 1996.
7. Heylighen F. Selection criteria for the evolution of knowledge. Proc 13th Int Congress on Cybernetics. Association Internat. de Cybernétique; 1993. pp. 524–528.
8. Moreno-Almeida C. Memes as snapshots of participation: The role of digital amateur activists in authoritarian regimes. New Media & Society. 2021;23: 1545–1566.
9. de Saint Laurent C, Glăveanu VP, Literat I. Internet Memes as Partial Stories: Identifying Political Narratives in Coronavirus Memes. Social Media + Society. 2021;7: 2056305121988932.

10. Niebuurt JT. Internet Memes: Leaflet Propaganda of the Digital Age. *Frontiers in Communication*. 2021;5. doi:10.3389/fcomm.2020.547065
11. Mascarenhas M. Memes as Quasi-Argument. *Local Theories of Argument*. 2021. pp. 385–390. doi:10.4324/9781003149026-65
12. Observatory, DiResta, Cryst, Masterson, Small, Miller, et al. Memes, Magnets and Microchips: Narrative dynamics around COVID-19 vaccines. *Virality Project*. 2022. Available: <https://cyber.fsi.stanford.edu/io/publication/memes-magnets-and-microchips-narrative-dynamics-around-covid-19-vaccines>
13. Toulmin SE. *The Uses of Argument*. Cambridge University Press; 1958.
14. Ling C, AbuHilal I, Blackburn J, De Cristofaro E, Zannettou S, Stringhini G. Dissecting the Meme Magic: Understanding Indicators of Virality in Image Memes. *Proc ACM Hum-Comput Interact*. 2021;5: 1–24.
15. Mascarenhas M, Cordes RJ, Friedman DA. Digital Rhetorical Ecosystem Analysis: Sensemaking of Digital Memetic Discourse. *Zenodo*. 2021. doi:10.5281/zenodo.5573947
16. Chagas V, Freire F, Rios D, Magalhães D. Political memes and the politics of memes: A methodological proposal for content analysis of online political memes. *First Monday*. 2019. Available: <https://firstmonday.org/ojs/index.php/fm/article/view/7264/7731>
17. Zakem V, McBride MK, Hammerberg K. Exploring the utility of memes for US government influence campaigns. *Center for Naval Analyses Arlington United States*; 2018. Available: <https://apps.dtic.mil/sti/citations/AD1052398>
18. Cordes RJ, David S, Maan A, Ruiz A, Sapp E, Scannell P, et al. *The Narrative Campaign Field Guide - First Edition*. 1st ed. Cordes RJ, editor. *Narrative Strategies Ink*; 2021.
19. Cordes RJ. The Next Generation of Security: Prioritizing Information Warfare Defense. 7 Aug 2021 [cited 12 Aug 2021]. Available: <https://www.hstoday.us/subject-matter-areas/intelligence/the-next-generation-of-security-prioritizing-information-warfare-defense/>
20. Cordes RJ. Games with serious consequences: Extremist movements and kayfabe - Atlantic Council. In: [atlanticcouncil.org](https://atlanticcouncil.org) [Internet]. 19 May 2021 [cited 3 Jun 2021]. Available: <https://www.atlanticcouncil.org/blogs/geotech-cues/games-with-serious-consequences-extremist-movements-and-kayfabe-2/>
21. Friedman DA, Cordes RJ. Infinite Games for Infinite Teams. *The Great Preset: Remote Teams and Operational Art*. COGSEC; 2020.

22. Taecharungroj V, Nueangjamnong P. Humour 2.0: Styles and Types of Humour and Virality of Memes on Facebook. *Journal of Creative Communications*. 2015;10: 288–302.
23. Guenther L, Ruhrmann G, Bischoff J, Penzel T, Weber A. Strategic Framing and Social Media Engagement: Analyzing Memes Posted by the German Identitarian Movement on Facebook. *Social Media + Society*. 2020;6: 2056305119898777.
24. Tindale CW. Replicating Reasons: Arguments, Memes, and the Cognitive Environment. *Philosophy & Rhetoric*. 2017;50: 566–588.
25. Ceccarelli L. Polysemy: Multiple meanings in rhetorical criticism. *Q J Speech*. 1998;84: 395–415.
26. Wikipedia contributors. Condensing Wonka. In: Wikipedia, The Free Encyclopedia [Internet]. 1 Jun 2021. Available: [https://en.wikipedia.org/w/index.php?title=Condensing\\_Wonka&oldid=1026354342](https://en.wikipedia.org/w/index.php?title=Condensing_Wonka&oldid=1026354342)
27. Petty RE, Cacioppo JT. The Elaboration Likelihood Model of Persuasion. In: Berkowitz L, editor. *Advances in Experimental Social Psychology*. Academic Press; 1986. pp. 123–205.
28. Karlova N. Misinformation and disinformation in online games: an exploratory investigation of possible cues. *digital.lib.washington.edu*. 2018. Available: <http://digital.lib.washington.edu/researchworks/handle/1773/42416>
29. Hemsley, Snyder. Dimensions of visual misinformation in the emerging media landscape. *Misinformation and mass audiences*. 2018. Available: [https://books.google.com/books?hl=en&lr=&id=jUw7DwAAQBAJ&oi=fnd&pg=PA91&dq=misinformation+artifacts&ots=Cv4UnjEdKZ&sig=DtirCALUY0-oCZ\\_ROpqb8ZYBqQc](https://books.google.com/books?hl=en&lr=&id=jUw7DwAAQBAJ&oi=fnd&pg=PA91&dq=misinformation+artifacts&ots=Cv4UnjEdKZ&sig=DtirCALUY0-oCZ_ROpqb8ZYBqQc)
30. Cordes RJ, Friedman DA. Emergent Teams for Complex Threats. *The Great Preset: Remote Teams and Operational Art*. COGSEC; 2020. pp. 1–15.
31. Hürsch WL, Lopes CV. Separation of Concerns. 1995 [cited 24 Aug 2020]. doi:10.1.1.29.5223
32. Cordes RJ, Friedman DA, Maron M. Reimagining Maps. {National GeoSpatial Intelligence Agency Polyplexus Incubator}; 2020 Oct. doi:10.5281/zenodo.4170026
33. Karachiwalla R, Pinkow F. Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. *Creat Innov Manag*. 2021;30: 563–584.

34. Friedman D, Applegate-Swanson S, Choudhury A, Cordes RJ, El Damaty S, Guénin-Carlut A, et al. An Active Inference Ontology for Decentralized Science: from Situated Sensemaking to the Epistemic Commons. 2022. doi:10.5281/zenodo.6320575
35. Jamal U. Facebook is reversing its ban on posts praising Ukraine's far-right Azov Battalion, report says. Business Insider. 25 Feb 2022. Available: <https://www.businessinsider.com/facebook-reverses-ban-praise-ukraine-far-right-forces-2022-2>. Accessed 19 May 2022.
36. Cordes RJ, Friedman DA. The Facilitator's Catechism. In: Friedman DA, Cordes RJ, editors. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
37. StoneToss. Fast Friends Image Meme. In: stonetoss.com [Internet]. 24 Mar 2022 [cited 18 May 2022]. Available: <https://stonetoss.com/comic/fast-friends/>
38. Ukraine and Putin RN. In: 9GAG [Internet]. 1 Mar 2022 [cited 23 May 2022]. Available: <https://9gag.com/gag/a21vjMe>
39. van Brugen I. Russia Arrests Multiple People for Holding Up Blank Signs. In: Newsweek [Internet]. 14 Mar 2022 [cited 23 May 2022]. Available: <https://www.newsweek.com/russia-ukraine-war-invasion-protests-police-arrest-activists-holding-blank-signs-paper-1687603>
40. Heuer RJ. Psychology of Intelligence Analysis. Center for the Study of Intelligence; 1999.
41. Jervis R. Perception and Misperception in International Politics. Princeton University Press; 2017.
42. Potamias RA, Siolas G, Stafylopatis A-G. A transformer-based approach to irony and sarcasm detection. Neural Comput Appl. 2020;32: 17309–17320.
43. Koutlis C, Schinas M, Papadopoulos S. MemeTector: Enforcing deep focus for meme detection. arXiv [cs.CV]. 2022. Available: <http://arxiv.org/abs/2205.13268>
44. Sinha A, Patekar P, Mamidi R. Unsupervised Approach for Monitoring Satire on Social Media. Proceedings of the 11th Forum for Information Retrieval Evaluation. New York, NY, USA: Association for Computing Machinery; 2019. pp. 36–41.
45. Chen E. 30% of Google's Emotions Dataset is Mislabeled. 14 Jul 2022 [cited 14 Jul 2022]. Available: <https://www.surgehq.ai/blog/30-percent-of-googles-reddit-emotions-dataset-is-mislabeled>
46. Wei P, He F, Zou Y. Content semantic image analysis and storage method based on intelligent computing of machine learning annotation. Neural Comput Appl. 2020;32: 1813–1822.

47. Henning C, Ewerth R. Estimating the information gap between textual and visual representations. *International Journal of Multimedia Information Retrieval*. 2018;7: 43–56.
48. Otto C, Springstein M, Anand A, Ewerth R. Characterization and classification of semantic image-text relations. *International Journal of Multimedia Information Retrieval*. 2020;9: 31–45.
49. Blair AM. *Too Much to Know*. Yale University Press; 2010.
50. Mangold C. A survey and classification of semantic search approaches. *Int J Metadata Semant Ontol*. 2007;2: 23–34.
51. Waltz E. *Knowledge Management in the Intelligence Enterprise*. Artech House; 2003.
52. Kent S. *Strategic Intelligence for American World Policy*. Princeton University Press; 1949.
53. Tamari R, Friedman D, Fischer W, Hebert L, Shahaf D. From Users to (Sense)Makers: On the Pivotal Role of Stigmergic Social Annotation in the Quest for Collective Sensemaking. *arXiv [cs.SI]*. 2022. Available: <http://arxiv.org/abs/2205.06345>
54. Gleich DF. PageRank Beyond the Web. *SIAM Rev*. 2015;57: 321–363.
55. Cordes RJ, Applegate-Swanson S, Friedman DA, Knight VB, Mikhailova A. Narrative Information Management. In: Cordes RJ, Friedman DA, editors. *Narrative Information Ecosystems: Conflict and Trust on the Endless Frontier*. COGSEC; 2021. pp. 1–64.
56. Kashcha A. Map of Reddit. In: [anvaka.github.io](https://anvaka.github.io) [Internet]. [cited 16 Jun 2022]. Available: <https://anvaka.github.io/map-of-reddit/?x=137263.0138067139&y=331978.27245969005&z=967321.673408511>
57. Search by Functional Annotation. In: [genomehubs.org](https://genomehubs.org) [Internet]. [cited 21 Jul 2022]. Available: <https://genomehubs.org/communities/tutorials/search-by-functional-annotation/>
58. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13: 2498–2504.
59. Adobe. *One Platform to Rule Them All: 5 reasons to choose a hybrid platform that combines B2B and B2C capabilities*. 2019 Feb.

60. Solarwinds. "One Platform to Rule Them All": How SolarWinds provided IT visibility across the VA's enterprise. In: Federal News Network [Internet]. 30 Oct 2019 [cited 24 Jun 2022]. Available: <https://federalnewsnetwork.com/innovation-in-government-success-stories/2019/10/one-platform-to-rule-them-all-how-solarwinds-provided-it-visibility-across-the-vas-enterprise/>
61. Dunn J. One Platform To Rule Them All: How N2N Makes Life Easier for Smaller Institutions. In: illuminateapp.com [Internet]. 7 Jun 2019 [cited 24 Jun 2022]. Available: <https://illuminateapp.com/web/EnlightenEd/one-platform-to-rule-them-all-how-n2n-makes-life-easier-for-smaller-institutions/>
62. Reed J. One app to replace them all? In: diginomica [Internet]. 3 Feb 2021 [cited 24 Jun 2022]. Available: <https://diginomica.com/one-app-replace-them-all-clickups-ceo-future-productivity-and-how-clickup-addresses-proliferation>
63. Mattr. Mattr - Building an Open and Interoperable Ecosystem for Digital Trust. In: Mattr.global [Internet]. 2021 [cited 29 Jun 2022]. Available: <https://mattr.global/approach/>
64. Norman D. The Design of Everyday Things: Revised and Expanded Edition. Basic Books; 2013.
65. Gamez-Diaz A, Fernandez P, Ruiz-Cortes A. An Analysis of RESTful APIs Offerings in the Industry. Service-Oriented Computing. Springer International Publishing; 2017. pp. 589–604.
66. Soomro, Awan. Challenges and Future of Enterprise Application Integration. Int J Comput Appl. 2012. Available: <http://www.ijcaonline.org/archives/volume42/number7/5708-7762>
67. Cestari RH, Ducos S, Exposito E. iPaaS in Agriculture 4.0: An Industrial Case. 2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE). ieeexplore.ieee.org; 2020. pp. 48–53.
68. Gartner. Definition of Integration Platform as a Service (iPaaS). In: Gartner Information Technology Glossary [Internet]. [cited 30 Jun 2022]. Available: <https://www.gartner.com/en/information-technology/glossary/information-platform-as-a-service-ipaas>
69. Ortiz A, Waldron R, Taylor H. The Beginner's Guide to General Automation Platforms. 2019.
70. Frantz RZ, Corchuelo R, Basto-Fernandes V, Rosa-Sequeira F, Roos-Frantz F, L. Arjona J. A cloud-based integration platform for enterprise application integration: A Model-Driven Engineering approach. Softw Pract Exp. 2020. doi:10.1002/spe.2916

71. Rahmati A, Fernandes E, Jung J, Prakash A. IFTTT vs. Zapier: A Comparative Study of Trigger-Action Programming Frameworks. arXiv [cs.CR]. 2017. Available: <http://arxiv.org/abs/1709.02788>
72. Ovadia S. Automate the Internet With “If This Then That” (IFTTT). Behav Soc Sci Librar. 2014;33: 208–211.
73. Singh, Iqbal, Singh, Kumar. Smart garden with iot based plant monitoring system. IEEE Solid-State Circuits Soc Newsl. 2020. Available: [https://www.researchgate.net/profile/Surya-Singh-2/publication/353878999\\_Smart\\_Garden\\_with\\_IoT\\_based\\_Plant\\_Monitoring\\_System/links/611632f00c2bfa282a3f72fb/Smart-Garden-with-IoT-based-Plant-Monitoring-System.pdf](https://www.researchgate.net/profile/Surya-Singh-2/publication/353878999_Smart_Garden_with_IoT_based_Plant_Monitoring_System/links/611632f00c2bfa282a3f72fb/Smart-Garden-with-IoT-based-Plant-Monitoring-System.pdf)
74. DOMO. DOMO Data Integration. In: Domo.com [Internet]. [cited 6 Jul 2022]. Available: [https://www.domo.com/solution/data-integration?utm\\_source=google&utm\\_medium=paidsearch&campid=7015w000000vIPfAAI&gcreative=474206076414&gdevice=c&gnetwork=g&gkeyword=application%20interface&gplacement=&gmatchtype=p&gtarget=&gadposition=&s\\_kwcid=AL15964!3!474206076414!p!!g!!application%20interface&gclid=Cj0KCQjw5ZSWBhCVARIsALERCvyz8XHgFUQc-ylxkHj8NvRiT3BoOkbAbdEp0L\\_-gX8k6GWClO5Jt3kaAghiEALw\\_wcB](https://www.domo.com/solution/data-integration?utm_source=google&utm_medium=paidsearch&campid=7015w000000vIPfAAI&gcreative=474206076414&gdevice=c&gnetwork=g&gkeyword=application%20interface&gplacement=&gmatchtype=p&gtarget=&gadposition=&s_kwcid=AL15964!3!474206076414!p!!g!!application%20interface&gclid=Cj0KCQjw5ZSWBhCVARIsALERCvyz8XHgFUQc-ylxkHj8NvRiT3BoOkbAbdEp0L_-gX8k6GWClO5Jt3kaAghiEALw_wcB)
75. MuleSoft. What is an API Economy. In: MuleSoft [Internet]. [cited 21 Jul 2022]. Available: <https://www.mulesoft.com/resources/api/what-is-an-api-economy>
76. KnowYourMeme. About Know Your Meme. In: Know Your Meme [Internet]. [cited 6 Jul 2022]. Available: <https://knowyourmeme.com/about>
77. LiterallyMedia. Literally – Home of Legendary Brands. In: Literally Media [Internet]. [cited 6 Jul 2022]. Available: <https://literally.media/>
78. David S, Cordes RJ, Friedman DA. Active Inference in Modeling Conflict. 2021. doi:10.5281/zenodo.5750935
79. Netskope Marketing. Netskope Cloud Report - August 2019. 1 Aug 2019 [cited 30 Jun 2022]. Available: <https://resources.netskope.com/cloud-security-infographics/netkope-cloud-report-august-2019-2>
80. Ismail K. How To Manage Digital Workplace Tool Overload. In: CMSWire.com [Internet]. 1 Nov 2017 [cited 30 Jun 2022]. Available: <https://www.cmswire.com/digital-workplace/how-to-manage-digital-workplace-tool-overload/>
81. Eide N, Bolden-Barrett V. App overload wastes 32 days of employee productivity each year. In: ciodive.com [Internet]. 7 Mar 2018 [cited 30 Jun 2022]. Available:

<https://www.ciodive.com/news/app-overload-wastes-32-days-of-employee-productivity-each-year/518520/>

82. Kilpatrick J, Brown J, Flynn R, Addicoat A, Mitchell P. Deloitte Global 2021 Chief Procurement Officer Survey. In: Deloitte Insights [Internet]. Deloitte; 23 Apr 2021 [cited 30 Jun 2022]. Available: <https://www2.deloitte.com/global/en/insights/topics/operations/chief-procurement-officer-cpo-survey.html>



## Chapter XI Works Consulted

---

In addition to the works included in this volume:

1. Xia H, Østerlund C, McKernan B. TRACE: A stigmergic crowdsourcing platform for intelligence analysis. of the 52nd .... 2019. Available: <https://scholarspace.manoa.hawaii.edu/handle/10125/59484>
2. Suhartono D, Gema AP, Winton S, David T, Fanany MI, Arymurthy AM. Argument annotation and analysis using deep learning with attention mechanism in Bahasa Indonesia. *Journal of Big Data*. 2020;7: 1–18.
3. Karachiwalla R, Pinkow F. Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. *Creat Innov Manag*. 2021;30: 563–584.
4. Goodnight GT. Complex Cases and Legitimation Inference: Extending the Toulmin Model to Deliberative Argument in Controversy. In: Hitchcock D, Verheij B, editors. *Arguing on the Toulmin Model: New Essays in Argument Analysis and Evaluation*. Dordrecht: Springer Netherlands; 2006. pp. 39–48.
5. Hitchcock D, Verheij B, editors. *Arguing on the Toulmin Model: New Essays in Argument Analysis and Evaluation*. Springer, Dordrecht; 2006.
6. Weinstein M. A Metamathematical Extension of the Toulmin Agenda. In: Hitchcock D, Verheij B, editors. *Arguing on the Toulmin Model: New Essays in Argument Analysis and Evaluation*. Dordrecht: Springer Netherlands; 2006. pp. 49–69.
7. Bianchini M, Gori M, Scarselli F. Inside PageRank. *ACM Trans Internet Technol*. 2005;5: 92–128.
8. Stab C, Gurevych I. Annotating Argument Components and Relations in Persuasive Essays. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Dublin, Ireland:

Dublin City University and Association for Computational Linguistics; 2014. pp. 1501–1510.

9. Coonradt CA. The Game of Work. Gibbs Smith; 2007.
10. Friedman D, Applegate-Swanson S, Choudhury A, Cordes RJ, El Damaty S, Guénin-Carlut A, et al. An Active Inference Ontology for Decentralized Science: from Situated Sensemaking to the Epistemic Commons. 2022. doi:10.5281/zenodo.6320575
11. Vyatkin A, Metelkin I, Mikhailova A, Cordes RJ, Friedman DA. Active Inference & Behavior Engineering for Teams. Active Inference Lab; 2020 Sep. doi:10.5281/zenodo.4021163
12. Afuah A, Tucci CL. Crowdsourcing As a Solution to Distant Search. AMRO. 2012;37: 355–375.
13. Ayaburi EW, Lee J, Maasberg M. Understanding Crowdsourcing Contest Fitness Strategic Decision Factors and Performance: An Expectation-Confirmation Theory Perspective. Inf Syst Front. 2020;22: 1227–1240.
14. Tamari R, Friedman D, Fischer W, Hebert L, Shahaf D. From Users to (Sense)Makers: On the Pivotal Role of Stigmergic Social Annotation in the Quest for Collective Sensemaking. arXiv [cs.SI]. 2022. Available: <https://dl.acm.org/doi/abs/10.1145/3511095.3536361>
15. He J, van Ossenbruggen J, de Vries AP. Fish4label: accomplishing an expert task without expert knowledge. Proceedings of the 10th Conference on Open Research Areas in Information Retrieval. groups.inf.ed.ac.uk; 2013. pp. 211–212.
16. Cordes RJ, Friedman DA. Narrative Information Ecosystems: Conflict and Trust on the Endless Frontier. COGSEC; 2021.
17. Friedman DA, Cordes RJ, editors. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
18. Baxter G, Sommerville I. Socio-technical systems: From design methods to systems engineering. Interact Comput. 2011;23: 4–17.
19. Zadeh LA. Fuzzy sets. Information and Control. 1965;8: 338–353.
20. Maiers J, Sherif YS. Applications of fuzzy set theory. IEEE Trans Syst Man Cybern. 1985;SMC-15: 175–189.
21. McMillan J. Reinventing the Bazaar: A Natural History of Markets. W. W. Norton & Company; 2003.

22. Milgrom PR, Roberts J. Economics, Organization, and Management. Prentice-Hall; 1992.
23. Milgrom PR, North DC, Weingast\* BR. The role of institutions in the revival of trade: The law merchant, private judges, and the champagne fairs. *Econ Polit.* 1990;2: 1–23.
24. Milgrom P. Auctions and Bidding: A Primer. *J Econ Perspect.* 1989;3: 3–22.
25. Glowacki A. From narratives of violence to narratives of peace: The renunciation of violence as a discursive phenomenon. 2013. Available: <https://search.proquest.com/openview/12b2e7d7d9476024a27c66140ecff86/1?pq-origsite=gscholar&cbl=18750>
26. Cordes RJ, David S, Maan A, Ruiz A, Sapp E, Scannell P, et al. The Narrative Campaign Field Guide - First Edition. 1st ed. Cordes RJ, editor. Narrative Strategies Ink; 2021.
27. Sharma V, You I, Jayakody DNK, Atiquzzaman M. Cooperative trust relaying and privacy preservation via edge-crowdsourcing in social Internet of Things. *Future Gener Comput Syst.* 2019;92: 758–776.
28. citizensciencegames.com. List of Citizen Science Games. In: Citizen Science Games [Internet]. 2018 [cited 17 Jun 2022]. Available: <https://citizensciencegames.com/games/>
29. Dikopoulou A, Mihiotis A. The contribution of records management to good governance. *The TQM Journal.* 2012;24: 123–141.
30. Waldispühl J, Szantner A, Knight R, Caisse S, Pitchford R. Leveling up citizen science. *Nat Biotechnol.* 2020;38: 1124–1126.
31. Friedman DA, Cordes RJ. Infinite Games for Infinite Teams. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020.
32. Cordes RJ. Games with serious impacts: The next generation of serious games - Atlantic Council. In: atlanticcouncil.org [Internet]. 21 May 2021 [cited 3 Jun 2021]. Available: <https://www.atlanticcouncil.org/blogs/geotech-cues/games-with-serious-impacts-the-next-generation/>
33. Friedman JA, Zeckhauser R. Assessing Uncertainty in Intelligence. *Intell Natl Sec.* 2012;27: 824–847.
34. Cordes RJ, Friedman DA. Emergent Teams for Complex Threats. The Great Preset: Remote Teams and Operational Art. COGSEC; 2020. pp. 1–15.

35. Waltz E. Quantitative Intelligence Analysis: Applied Analytic Models, Simulations, and Games. Illustrated edition. Rowman & Littlefield Publishers; 2014.
36. Waltz E. Knowledge Management in the Intelligence Enterprise. Artech House; 2003.