

Golden Line: Toward What Matters

An Aspirational Thread for Long-Horizon Work

Daniel Ari Friedman

Active Inference Institute

`daniel@activeinference.institute`

ORCID: [0000-0001-6232-9096](https://orcid.org/0000-0001-6232-9096)

DOI: [10.5281/zenodo.21754238](https://doi.org/10.5281/zenodo.21754238)

2026-07-18



Contents

1	Abstract	2
2	Introduction: the work needs a direction	3
3	Relationship to the line set	4
3.1	Note on the name	4
4	Method: aspiration as a directional record	5
4.1	The registry entry	5
4.2	The horizon entry	5
4.3	The staged evaluator	5
4.4	Conservative precedence	5
4.5	Evidence and artifact boundary	5
4.6	The descriptive analysis layer	6
5	Formalism: the evaluator and its invariants	7
5.1	Domain objects	7
5.2	The staged evaluator	7
5.3	Propositions about the evaluator	8
5.3.1	The completeness half of the rule	10
5.3.2	The temporal half of the rule	10
5.4	Structural invariants	13
5.5	The report envelope	13
5.6	Formalism-to-test bindings	13
6	Scholarship: an aspiration is a direction, not a score	17
6.1	A translation, not a synthesis	17
6.2	The shape of an end	18
6.3	Capability and flourishing	18
6.4	Attention, craft, and transfer	18
6.5	Commons, power, and return	19
6.6	Why it must not become a score	19
6.7	From theory to instrument	19
7	Evidence boundary and artifact chain	21
7.1	Hard constraints and soft choices	21
7.2	Statuses as bounded claims	21
7.3	Source to publication	23
8	The Golden Line aspirations	24
8.1	The signal vocabulary in aggregate	24
8.2	The reach of the nine horizons	24
9	Worked records	29
10	Reading a batch: the descriptive layer at work	32
10.1	A worked batch	32
10.2	What intake set aside	33
10.3	Vocabulary and reach as reading context	34
11	Limits and safeguards	35
11.1	Limits of the descriptive analysis layer	35
12	Conclusion	36

1 Abstract

Golden Line is an aspirational thread: a compact, revisable way to say what a piece of work is trying to serve over a long horizon. It asks what is worth reaching toward after Red Line has named the boundary and alongside Black Line's operating discipline. It is deliberately not a compliance score. A directional reading never certifies that work is good, safe, lawful, or complete.

The executable instrument is a small Python package with a versioned aspiration registry of nine entries — four founding aspirations and five further ones — and a staged **progress_report** evaluator. Each aspiration pairs a plain-language thread and a concrete horizon with two short lists: Markers, the observable signs of movement toward it, and Counter-signals, the observable signs of movement away. An observer files a horizon entry, optionally dated, recording which of those were seen. The evaluator screens and normalizes the entry, matches it against the registry, preserves a structured derivation, and returns one of four directional readings: **TOWARD**, **INQUIRY**, **DRIFTING**, or **NOT_OBSERVED**. The report carries the registry digest and the temporal review context, so a reading is never detached from the vocabulary and the date that produced it.

The paper develops the instrument in three registers. A **formal method** section states the domain objects, the staged evaluator, and its decision rule as definitions and propositions that match the code exactly, each bound to the named test that would fail if the code diverged, and records seven structural invariants, each proven by a planted-bad proof-of-detection test. A **scholarship** section situates the founding aspirations — attention before output, usefulness beyond the author, repairable systems, and answerability to human flourishing — in a lineage running from practical wisdom and goods internal to a practice through the capability approach, reflective practice, repair, commons governance, technical power, and the hazards of turning an aspiration into a metric. Thirteen code-derived **figures** render the registry, its signal vocabulary, the decision path, the staged pipeline, the evidence-state matrix, the horizon bands, and a loop for the founding four. Six of the thirteen replay the evaluator rather than diagram it: one entry ageing past the currentness boundary, the whole registry ageing at once across 126 dated readings, every marker subset of every aspiration, the twelve fields of the returned record under six evidence conditions, a worked batch of six entries, and a precedence panel in which complete markers plus one counter-signal still return **DRIFTING** for every aspiration.

Golden Line is standalone and non-redundant with its companions. Its vocabulary is horizon, thread, marker, counter-signal, and direction. The absence of an observation is not a failure finding; it is an invitation to look again, or to leave the claim honestly open.

2 Introduction: the work needs a direction

Every method can become technically competent while losing contact with the reason it exists. A test suite stays green, a pipeline keeps shipping, and yet the question of what the work is *for* quietly falls out of view. Golden Line names that danger without pretending to dissolve it with a single principle. It provides a small, revisable set of directions that can orient a decision, a collaboration, or a long project, and a way to ask, at a stated horizon, whether the work is visibly moving toward them.

The idea has many ancestors, and the instrument inherits from all of them without settling their disagreements. Aristotle treats practical judgment as deliberation about a good life rather than the mechanical application of a rule, and makes an *end* (*telos*) the thing that gives an activity its point [Aristotle, 350 BCE]. Confucian teaching connects learning to the cultivation of conduct within relationships, so that skill and character grow together rather than apart [Confucius, 500 BCE]. Ibn Khaldun's account of cooperation, power, and social change reminds us that purposes are carried by institutions and shared cohesion, not by individual intention alone [Khaldun, 1377]. Closer to the present, MacIntyre's distinction between goods internal to a practice and its external rewards helps explain why useful work must remain more than a successful artifact [MacIntyre, 1981]. The scholarship section extends this lineage through capabilities, reflective practice, repair, commons governance, and the politics of technical systems. These references are prompts for comparison and calibration, not authorities the registry claims to reconcile.

Golden Line's role is deliberately limited, and the limit is the point. It cannot authorize work that Red Line refuses. It cannot certify that work was done well, which is Black Line's job. It cannot fill in what White Line marks as absent. The next section places each of those instruments; here it is enough to say that Golden Line absorbs none of their work. What it adds is a direction held lightly, stated in the open, and revisable when the evidence or the values change.

The paper proceeds from prose to formalism to lineage: the method section describes the staged evaluator in words, the formal section restates it as definitions and propositions faithful to the code, and the scholarship section grounds the founding aspirations in their sources. The remaining sections walk the nine-entry registry, worked records, the evidence boundary, and honest limits.

The manuscript is only half of it. Golden Line is operated day to day through the `daf-golden-line` skill in DAF's private daf-skills toolchain: this paper states what the instrument means and why, and the skill is how a horizon entry is actually filed.

The paper is organised as follows. [Section sec. 4](#) defines the method and evidence protocol. [Section sec. 5](#) states the evaluator formally. [Section sec. 8](#) presents the four aspirations. [Section sec. 9](#) walks through worked examples, and [Section sec. 11](#) closes with limits and epistemic boundaries.

3 Relationship to the line set

Golden Line is the aspirational work in the four-line set, whose declaration lives in the companion work `line_set`. Red Line is the personal security boundary and explicit No document. Black Line is positive operating discipline. White Line records absence, restraint, and negative space. Golden Line does not copy their registries or evaluators, and a directional finding is never a compliance verdict.

A fifth work, `line_set`, is a thin reader that declares the set and checks that no two lines gave the same spelling to different things; it adds no substantive instrument, and Golden Line does not import, depend on, or defer to it.

3.1 Note on the name

The name is not incidental. *Golden Line* openly echoes the yellowing stage of the alchemical *magnum opus*, **citrinitas**, the dawning that classical schemes place between the whitening and the reddening. The echo is meant in Carl Jung’s symbolic-psychological register, where the stages of the opus are read as a map of individuation, and in that register only: it is a metaphor for a practice turning toward what it is for, never an empirical, mystical, or causal claim about matter, minds, or this instrument. The working order of the four-line set refuse, then method, then aspire, then absence, is functional and chosen for how the instruments are actually used, and deliberately does not reenact the opus’s canonical sequence (nigredo → albedo → citrinitas → rubedo). Citrinitas here simply marks Golden Line’s place in the set: the yellowing where a disciplined practice lifts its attention from how the work is done to what the work is worth. The grounding citation is Jung’s reading of the opus stages as figures for individuation rather than as laboratory chemistry [Jung, 1953], and it is carried here rather than deferred, so this paper does not depend on another document to say what its own name means. The set-level version of the framing — the two orders, the caveats, and the non-overlap contract — is declared in the companion work `line_set`; this paper points there for the set, not for its own citation.

4 Method: aspiration as a directional record

4.1 The registry entry

An aspiration has six fields: an identifier, a plain-language title, a thread that explains its direction, a horizon at which the direction becomes visible, and two short lists. **Markers** are observable signs that the work is moving in the declared direction. **Counter-signals** are observable signs that it is drifting away from it. Neither list is exhaustive, and neither is a questionnaire; they are the small, concrete anchors that let a direction be discussed rather than merely admired. The registry ships nine such entries: four founding aspirations and five further ones that fill out the long-horizon picture.

4.2 The horizon entry

An observer records movement by filing a **horizon entry**: the aspiration identifier, the set of markers actually observed, the set of counter-signals actually observed, a free note, and an optional ISO observation date. The entry describes a *record* of work at a moment, not a person and not a project as a whole.

4.3 The staged evaluator

`progress_report` reads a batch of horizon entries against the versioned registry in three stages, and every finding it returns carries a full reasons trail plus structured derivation fields so no reading arrives unexplained.

1. **Intake screening.** Each incoming entry is checked for the expected record shape and against the known aspiration identifiers. An entry for an unknown aspiration is set aside with a note; a second valid entry for an aspiration already seen is set aside, and the first entry stands. JSON-like token lists are normalized to sets. Malformed or hostile input is *recorded*, never raised. A sloppy record cannot crash the report or vanish silently.
2. **Matching.** For each aspiration, the observed markers are intersected with the declared markers, and the observed counter-signals with the declared counter-signals. Tokens the observer supplied that the registry never declared are ignored and noted; they cannot smuggle in a status.
3. **Decision.** With optional temporal review folded in, the matched sets and temporal quality determine one of four directional readings. There is no numeric score or aggregate grade.

The four readings are:

- **TOWARD** means every declared marker is observed, no counter-signal is recorded, and (if temporal review is enabled) currentness is auditable: the observation date is present, parseable, and within the declared window.
- **INQUIRY** means the aspiration is named but the evidence is partial, empty, or stale, so the direction remains honestly open.
- **DRIFTING** means a declared counter-signal is present. It describes the record, not the person or project as a whole.
- **NOT_OBSERVED** means no valid horizon entry was admitted for that aspiration. An entry can be present in the input but fail intake because it is malformed, unknown, or a duplicate set aside by the first-valid-entry rule.

The finding also preserves the note, observed date, declared counter-signals, and undeclared tokens that were ignored. This makes the status machine-readable without requiring a downstream reader to reverse-engineer human-readable reason strings.

4.4 Conservative precedence

The precedence is intentionally conservative. A counter-signal is surfaced before any markers are counted, and staleness never erases a recorded drift: a declared counter-signal yields **DRIFTING** even when the same observation is old. Partial, stale, or currentness-un-auditable positive evidence reverts to **INQUIRY** rather than being rounded up to **TOWARD**. This mirrors a satisficing stance: the evaluator looks for *good enough and current* evidence of direction and refuses to over-read thin signals [Simon, 1956].

The result is a directional report. It is not an evaluator of worth, a certification, or a substitute for the Red Line, Black Line, or White Line instruments.

4.5 Evidence and artifact boundary

The machine has four layers: the versioned registry declares the vocabulary; an admitted entry records a local observation; the evaluator derives a bounded reading; and the generated registries and figures preserve what source was used. No layer is allowed to add a claim that the earlier layer did not carry. The report's registry version, registry digest, review date, and staleness threshold therefore travel with the findings. The local artifact gate checks that the generated JSON, SVG, PNG, and manuscript figure labels still agree before a sibling template render is attempted.

4.6 The descriptive analysis layer

A small analysis module sits beside the evaluator and is deliberately weaker than it: every helper is pure, deterministic, and read-only, and none can change what a reading means. Four helpers are provided.

- **signal_inventory** tallies the declared markers and counter-signals across a registry — for the shipped registry, 18 markers and 9 counter-signals, with every token distinct — describing the vocabulary the evaluator can match, never its fulfilment.
- **horizon_distribution** groups aspirations into four declared temporal-reach bands (immediate, recurring cycle, at handoff, open-ended). The band map lives in the analysis layer, not the registry contract, and an unclassified horizon raises an error so a registry change must revisit the map deliberately.
- **temporal_currentness_sweep** replays one fully-marked entry through the public **progress_report** across a range of observation ages, exposing the exclusive staleness boundary (current at age 90, stale at 91 for a 90-day window) as an observed trajectory rather than a claim.
- **report_overview** regroups an existing report by status and totals its ignored tokens, temporal flags, and intake notes, so a batch can be characterized without re-parsing prose reasons.

The analysis layer is also where the project’s experiment plan is grounded. The repository ships a **domain_profile.yaml** naming the validation gates: structural invariants, evidence grounding, artifact chain, render validation, and publication readiness. It also ships an **experiment_plan.yaml** whose three conditions — the source-registry baseline, the malformed-input guard, and the temporal-currentness guard — map onto the evaluator’s intake stage and temporal review. The plan’s expected figures are the ones the deterministic builder produces: five through this analysis layer’s helpers (the temporal-currentness sweep, the currentness lattice, the signal inventory, the horizon bands, and the batch reading overview) and eight straight from the registry, the evaluator, and the status contracts. The protocol compares reproducible contract outcomes only. None of these summaries is evidence that any aspiration is true, and none is a score.

5 Formalism: the evaluator and its invariants

The formalism restates the evaluator the method section describes; every result below tracks the code that implements it. This section states the implemented semantics. Every result below describes what the code in `golden_line` actually does, and the reasons and transitions named here are the ones the evaluator produces. The formalism *describes* the instrument; it does not extend or idealize it. Numbering is assigned by the renderer in document order, so no number is written in the source and none can go stale.

5.1 Domain objects

Definition 1 (Aspiration). An aspiration a is the six-tuple

$$a = (\text{id}, \text{title}, \text{thread}, \text{horizon}, M, C),$$

where `id`, `title`, `thread`, `horizon` are text fields and M (Markers) and C (Counter-signals) are finite sequences of observable tokens. Markers are signs of movement *toward* the aspiration; Counter-signals are signs of movement *away* from it.

Definition 2 (Registry). The registry $R = \langle a_1, \dots, a_n \rangle$ is an ordered tuple of aspirations with $n = 9$: four founding aspirations (attention before output, usefulness beyond the author, repairable systems, answerability to human flourishing) followed by five further ones (durable understanding, teachable craft, honest uncertainty, unhurried questions, improvements returned to the commons). Write $\text{ids}(R)$ for the set of identifiers appearing in R .

Definition 3 (Horizon entry). A horizon entry is $e = (\text{aid}, O, K, \text{note}, d_{\text{obs}})$, where `aid` names an aspiration, O is the set of observed markers, K is the set of observed counter-signals, `note` is free text, and d_{obs} is an optional ISO date.

Definition 4 (Status codomain). The directional readings form the four-element set

$$\Sigma = \{\text{TOWARD}, \text{INQUIRY}, \text{DRIFTING}, \text{NOT_OBSERVED}\},$$

copied verbatim from the `HorizonStatus` enumeration. These are the *only* statuses the evaluator can emit.

Definition 5 (Finding). A finding is

$$f = (\text{aid}, \sigma, \text{reasons}, \text{observed}, \text{unmet}, \text{countered}, \text{ign}_M, \text{ign}_C, \text{note}, d_{\text{obs}}, \text{stale}, \text{date_issue})$$

with $\sigma \in \Sigma$ (Definition 4). The structured fields list the declared markers observed and unmet, the declared counter-signals, the undeclared marker tokens ign_M and counter-signal tokens ign_C that were ignored, the original note and date, and two temporal quality flags. Ignored markers and ignored counter-signals are kept apart rather than pooled, because a reader checking why a token did nothing needs to know which vocabulary it failed to match. Human-readable **reasons** remain a parallel explanation, not the only source of derivation.

5.2 The staged evaluator

Definition 6 (Report function). The evaluator has signature

$$\text{progress_report} : (E, R, \text{as_of}, \tau) \rightarrow \text{HorizonReport},$$

where E is a batch of horizon entries, R defaults to the registry of Definition 2, `as_of` is an optional review date, and τ (`stale_after_days`) is an optional non-negative integer staleness threshold; `as_of` and τ are keyword-only. It runs three stages: intake screening, matching, and decision. The returned report also identifies the registry version/digest and the review context.

Definition 7 (Intake screening). Screening builds an accepted map $A : \text{ids}(R) \rightarrow E$ by a first-wins rule. Iterating E , a record with the wrong type or malformed fields is set aside with an intake note; an entry whose `aid` $\notin \text{ids}(R)$ is set aside with an unknown-id note; an entry whose `aid` is already in A is set aside with a duplicate note and the first valid entry stands; otherwise the entry is admitted. Iterable text-token fields are normalized to sets. Screening never raises on malformed or unexpected input. It is a pure function of the pair $(E, \text{ids}(R))$: records are tested in a fixed order — shape, then registry membership, then prior admission — and iteration follows the order of E with indices counted from 1 in the intake notes.

Definition 8 (Matching). For aspiration a (Definition 1) with admitted entry $e = A(a.\text{id})$ (Definition 3), define

$$\text{observed}(a, e) = O \cap M, \quad \text{unmet}(a, e) = M \setminus O, \quad \text{countered}(a, e) = K \cap C.$$

Undeclared tokens $O \setminus M$ and $K \setminus C$ are ignored and recorded in the reasons; they can never determine a status.

Definition 9 (Temporal review). Given review date d and threshold τ , temporal review is enabled exactly when τ is set. If τ is unset, temporal metadata does not affect the status. If τ is set, a missing or unparseable d_{obs} sets `date_issue` to true: the

record is not fatal, but it cannot support a currentness claim. With a parseable date, $\text{stale}(e, d, \tau)$ holds iff $d_{\text{obs}} > d$ (a future observation) or $d - d_{\text{obs}} > \tau$ days.

Definition 10 (Decision rule). The status $\sigma(a, e)$ is determined by the first matching clause, in order:

1. if $e = \perp$ (no admitted entry): NOT_OBSERVED;
2. else if $\text{countered}(a, e) \neq \emptyset$: DRIFTING;
3. else if $\text{observed}(a, e) = \emptyset$: INQUIRY;
4. else if $\text{unmet}(a, e) \neq \emptyset$: INQUIRY;
5. else if $\text{stale}(e, d, \tau)$ or date_issue : INQUIRY;
6. otherwise: TOWARD.

5.3 Propositions about the evaluator

Each proposition below follows from [Definition 7](#) through [Definition 10](#) and is exercised by the package’s test suite; the binding table at the end of this section names the verifying test for every result. Each proposition is a statement about code behavior — never about the world, safety, or persons.

Proposition 1 (Totality). For every aspiration $a \in R$ the evaluator emits exactly one finding ([Definition 5](#)), so a report contains $n = 9$ findings, one per registry entry, whether or not any entry was filed. The report’s **counts** therefore partition the findings across Σ .

Proposition 2 (Counter-signal precedence). If a declared counter-signal is recorded, meaning $\text{countered}(a, e) \neq \emptyset$ ([Definition 8](#)), the status is DRIFTING, regardless of how many markers were observed and regardless of staleness. Clause 2 precedes clauses 3–6, so a positive signal can never launder a recorded drift.

Proposition 3 (Exactness and currency of TOWARD). $\sigma(a, e) = \text{TOWARD}$ iff $\text{observed}(a, e) = M$ (every declared marker seen), $\text{countered}(a, e) = \emptyset$ ([Definition 8](#)), and neither $\text{stale}(e, d, \tau)$ nor date_issue holds. Any missing marker, any counter-signal, stale observation, or unparseable reviewed date reverts the reading to INQUIRY or DRIFTING. Counter-signal precedence is as stated in [Proposition 2](#). There is no partial credit: an entry recording every marker but one reads exactly as an entry recording none.

Proposition 4 (Temporal uncertainty reopens, never condemns). A fully-marked but stale or date-un-auditable observation with no counter-signal yields INQUIRY (clause 5), not DRIFTING. Age or malformed temporal metadata reopens a question; it does not manufacture drift.

Proposition 5 (Absence is not negation). NOT_OBSERVED arises solely from the absence of an admitted entry (clause 1). The admission rule of [Proposition 6](#) determines which entries are admitted; when none is, the reading is NOT_OBSERVED regardless of what records may have been submitted. It asserts nothing about whether the aspiration is being served; it reports only that no valid record was available to this report. The intake notes preserve why submitted records were set aside when that distinction matters.

Proposition 6 (First valid entry stands; the reading is deterministic). For each identifier, screening admits the first valid entry bearing it in the order of E : the first valid entry bearing an identifier stands, and every later entry with that identifier is set aside with a duplicate note. For an identifier with no valid entry, the reading is NOT_OBSERVED ([Proposition 5](#)). Exchanging the positions of two entries that share an identifier can therefore change the finding, but nothing else about the call can: replaying an identical batch with identical as_of and τ reproduces the identical report, digest for digest. Determinism is a property of the code path; it does not make the underlying records true.

Proposition 7 (Intake screening determinism). The intake screening function ([Definition 7](#)) is a pure function of the pair $(E, \text{ids}(R))$: records are tested in a fixed order — shape, then registry membership, then prior admission — and iteration follows the order of E with indices counted from 1 in the intake notes. For the identical batch E and the identical known $\text{ids}(R)$, screening produces the identical accepted map A and the identical intake notes every time. Malformed input is set aside deterministically: a record with the wrong type or malformed fields always gets a malformed note, an entry whose $\text{aid} \notin \text{ids}(R)$ always gets an unknown-id note, and an entry whose aid is already in A always gets a duplicate note. Screening never raises on malformed or unexpected input — it is a pure computation, and its determinism is separate from the first-wins semantics of [Proposition 6](#), which covers which entry stands when multiple valid entries share an identifier. Intake screening determinism covers the screening function’s behavior on all input, including malformed records.

Proposition 8 (The currentness boundary is exclusive). Staleness uses the strict comparison $d - d_{\text{obs}} > \tau$ of [Definition 9](#), so a fully-marked, counter-signal-free entry with $\tau = 90$ reads TOWARD at age exactly 90 days and reverts from age 91 days: age τ is the last current age and $\tau + 1$ the first stale one. The threshold is the caller’s review-cadence choice; nothing in the evaluator endorses any particular number of days as a natural constant.

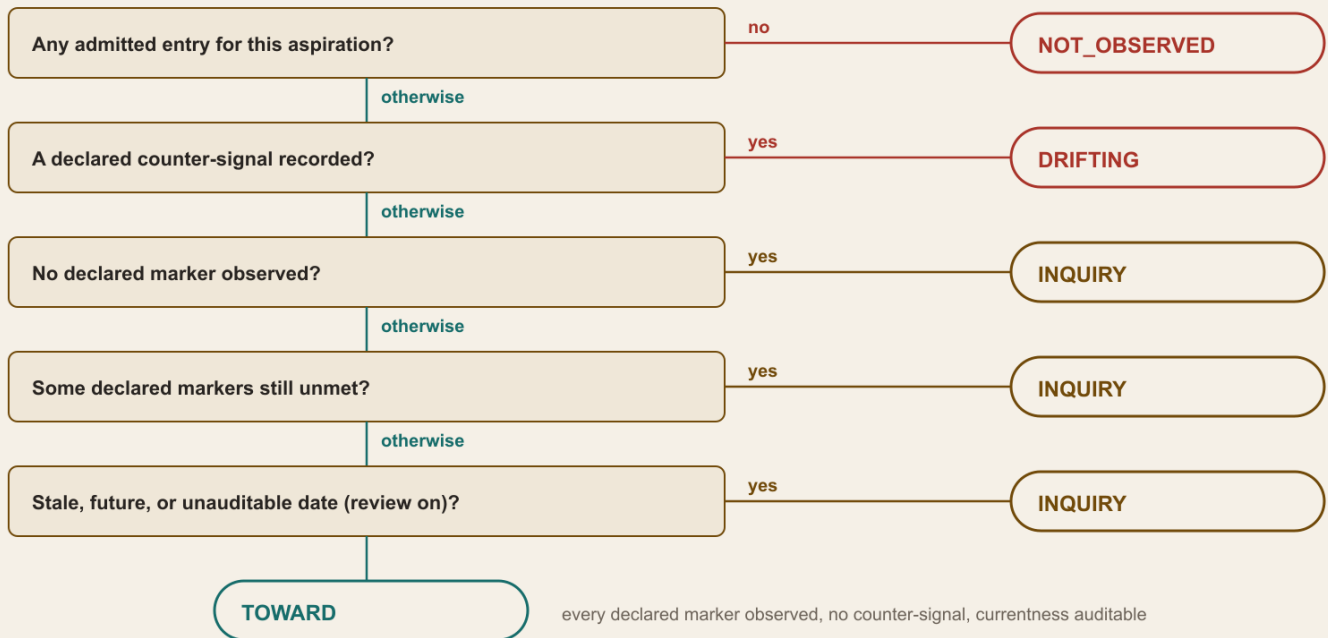
The ordered clauses and the four readings they terminate in are drawn from the `HorizonStatus` enumeration — the decision rule’s codomain, not the registry — in [fig. 1](#).

THE HORIZON DECISION PATH

How one entry becomes one directional reading

Clauses are tested in order; the first that matches decides.

A counter-signal outranks every marker.



SOURCE-DRIVEN SCHEMATIC · statuses quoted from the HorizonStatus enumeration · NOT A COMPLIANCE VERDICT

Figure 1: The evaluator's ordered decision rule. Counter-signal precedence is explicit; TOWARD requires complete markers and auditable currentness when temporal review is enabled. The path reads a record; it is not a compliance verdict.

5.3.1 The completeness half of the rule

The exactness clause of [Proposition 3](#) is the one a reader is most likely to soften into “mostly there”. [fig. 2](#) refuses that reading by running it: every subset of every aspiration’s declared markers is filed as an entry and evaluated, and the panel prints the status the evaluator returned alongside the number of markers left unmet. Only the complete subset reads **TOWARD**; a single missing marker holds the direction open at **INQUIRY**, exactly as an empty entry does. The panel characterizes the clause order, not any observed practice.

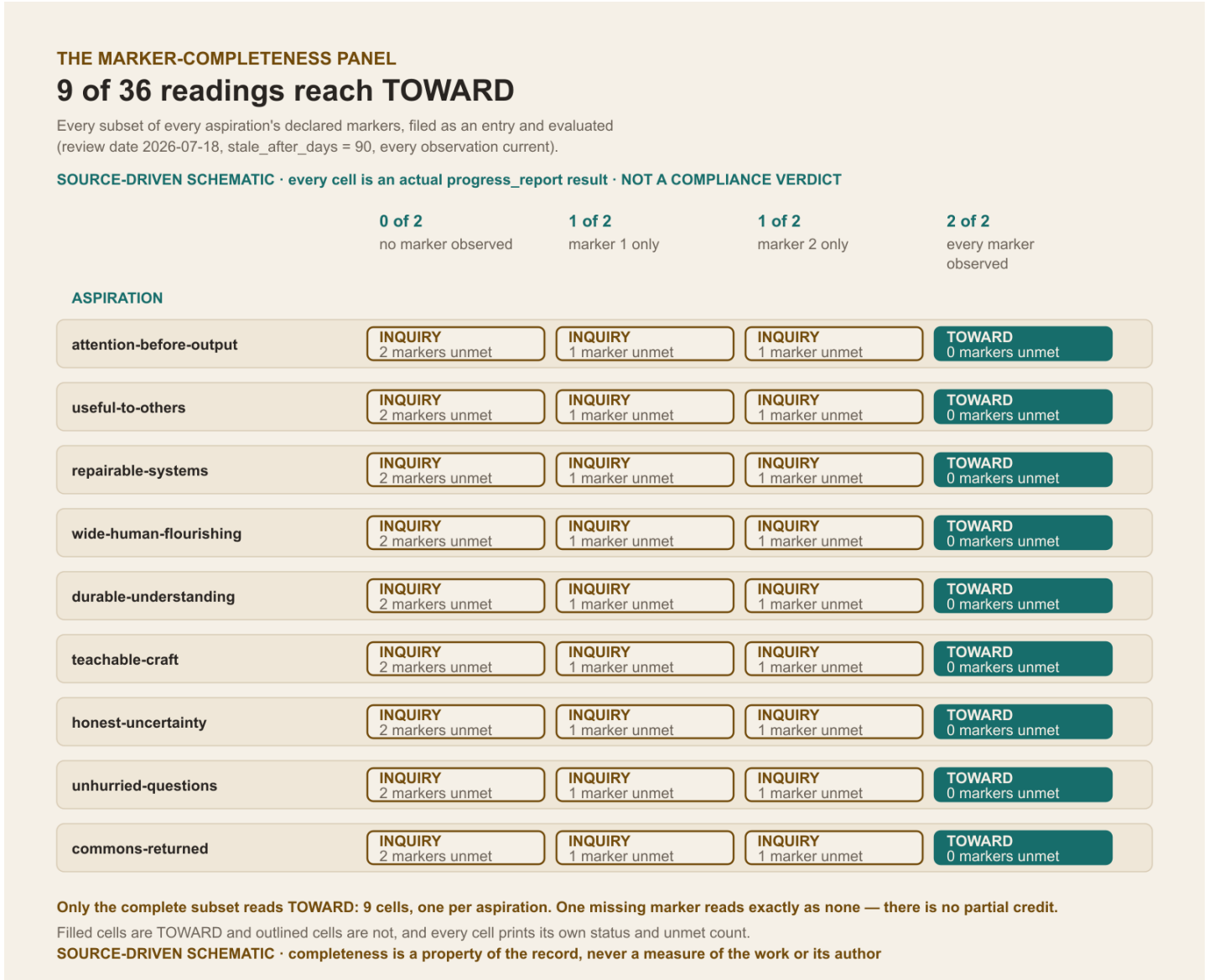


Figure 2: Every observed-marker subset for every aspiration, filed as an entry and evaluated. Each cell prints the status `progress_report` returned and the count of markers still unmet, so the grid reads without colour; only the complete subset reaches **TOWARD**, and one missing marker reads exactly as none. Marker completeness is a property of the record, never a measure of the work.

5.3.2 The temporal half of the rule

[Proposition 4](#) has a picture. The `temporal_currentness_sweep` analysis helper replays one fully-marked, counter-signal-free entry through the real evaluator at a range of observation ages with $\tau = 90$: the reading holds **TOWARD** through age 90 exactly (the boundary is exclusive), reverts to **INQUIRY** from age 91, and a future-dated observation likewise reads **INQUIRY**. Every cell in [fig. 3](#) is an actual `progress_report` result, not an illustration of one; the sweep observes the evaluator and never re-implements it.

Proposition 9 (Sweep delegation and codomain). The sweep computes no status of its own: its only status-producing call is `progress_report` ([Definition 6](#)), and each `SweepPoint` copies that call’s finding. This is an architectural property, not an empirical

THE TEMPORAL CURRENTNESS SWEEP

How one reading ages under temporal review

A fully-marked entry for 'attention-before-output' replayed through the real evaluator at each observation age (review date 2026-07-18, stale_after_days = 90).

SOURCE-DRIVEN SCHEMATIC · every cell is an actual progress_report result · NOT A COMPLIANCE VERDICT

INQUIRY -7 d (future) 2026-07-25 stale	TOWARD 0 d 2026-07-18 current	TOWARD 15 d 2026-07-03 current	TOWARD 30 d 2026-06-18 current	TOWARD 45 d 2026-06-03 current	TOWARD 60 d 2026-05-19 current	TOWARD 75 d 2026-05-04 current
exclusive boundary: age 90 is current, age 91 is stale						
TOWARD 89 d 2026-04-20 current	TOWARD 90 d 2026-04-19 current	INQUIRY 91 d 2026-04-18 stale	INQUIRY 105 d 2026-04-04 stale	INQUIRY 120 d 2026-03-20 stale	INQUIRY 150 d 2026-02-18 stale	INQUIRY 180 d 2026-01-19 stale

TOWARD **INQUIRY**

READING RULE age or a future date reopens the question (INQUIRY); it never manufactures drift

A stale reading is an invitation to look again, not evidence of failure.

Figure 3: A fully-marked entry replayed through the real evaluator at 14 observation ages with stale_after_days = 90. Every cell names the status progress_report actually returned, so nothing here is carried by colour alone: TOWARD through age 90, INQUIRY from age 91 and for future-dated observations. Age reopens the question; it never manufactures drift.

one — comparing a sweep point against a second `progress_report` call would compare the same code path with itself, so the verifying test instead asserts that `analysis.py` constructs no `HorizonStatus` anywhere. Because sweep entries are fully marked, counter-signal-free, and bear a known identifier, only clauses 5 and 6 of [Definition 10](#) can fire: `DRIFTING` and `NOT_OBSERVED` are unreachable in a sweep; that half *is* falsifiable and is tested by enumeration. The sweep characterizes the instrument’s temporal behavior; it does not describe any observed practice.

[Proposition 8](#) states the boundary and the sweep shows it for one registry entry. Whether the boundary belongs to the *evaluator* rather than to that one entry is a separate question, and one that a single row cannot answer.

Proposition 10 (The currentness boundary is uniform across the registry). Replaying the sweep ([Proposition 9](#)) for every $a \in R$ at the same review date and threshold yields the same `TOWARD` $\rightarrow$ `INQUIRY` transition age for every aspiration, because [Definition 9](#) reads only d_{obs} , d , and τ — no field of a enters the staleness comparison. The sweep confirms that the temporal-inquiry behaviour stated in [Proposition 4](#) and the exclusive boundary stated in [Proposition 8](#) are properties of the evaluator’s code path, not of any particular aspiration. The lattice in [fig. 4](#) is that replay: $9 \times 14 = 126$ executed `progress_report` calls, with each row’s own transition age printed beside it and a count of rows that differ. Uniformity here is a fact about the code path; it says nothing about how any aspiration is actually served.

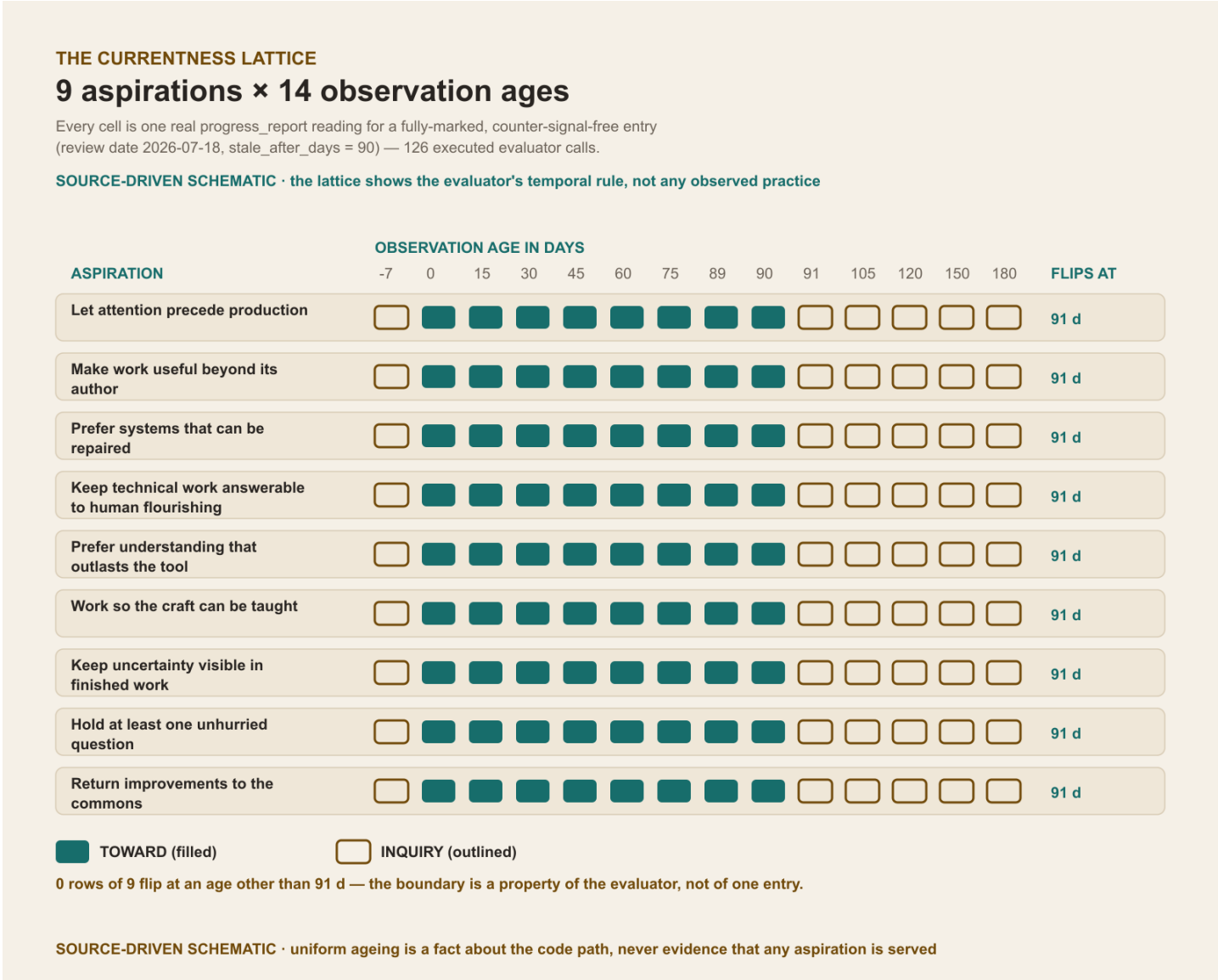


Figure 4: The same currentness replay run across the whole registry: 9 aspiration rows by 14 observation ages, 126 executed `progress_report` calls at `stale_after_days = 90`. Filled cells are `TOWARD` and outlined cells are `INQUIRY`, so the grid reads without colour; the `FLIPS AT` column gives each row’s own transition age and the footer counts rows that differ. Uniform ageing is a fact about the evaluator’s code path, not evidence that any aspiration is being served.

5.4 Structural invariants

The evaluator reads a registry; a malformed registry would make its readings meaningless. Seven pure-compute structural checks validate the *shape* of the registry, independent of any horizon entry. Let $I_1, \dots, I_7$ be their predicates; `registry_sound(R)` returns $\bigwedge_{k=1}^7 I_k(R)$.

- I_1 **distinct identifiers**: no two aspirations share an id, so findings are unambiguous.
- I_2 **fields populated**: id, title, thread, and horizon all carry content.
- I_3 **reachable statuses**: each aspiration has at least one marker and at least one counter-signal, so both TOWARD and DRIFTING are reachable for it.
- I_4 **signal text**: every marker and counter-signal is a non-blank string.
- I_5 **signal uniqueness**: no marker or counter-signal repeats within one declaration.
- I_6 **signal disjointness**: $M \cap C = \emptyset$ within each aspiration, so no token reads as movement toward and away at once.
- I_7 **digest stability**: the canonical registry serializes, and its SHA-256 digest is independent of registry order.

Proposition 11 (Digest order-independence). `canonical_registry` serializes aspirations sorted by identifier, so `registry_digest(R) = registry_digest( $\pi(R)$ )` for every permutation π of R , and the digest is exactly 64 lowercase hexadecimal characters (SHA-256). Two readers comparing digests are comparing registry content, never the order in which they happened to list it — and a matching digest attests only sameness of content, not soundness or merit.

Proposition 12 (Proof of detection). For each invariant I_k , the test suite asserts both that I_k *passes* on the real registry and that I_k *fails* on a deliberately planted-bad registry constructed to violate exactly that check. A green check that had never seen a bad input would not count as evidence, so each invariant is paired with the counter-example that gives it meaning.

These digests and checks are review and drift-detection instruments for humans comparing registry revisions. They carry no safety, warranty, or attestation semantics of any kind. The detection proof of [Proposition 12](#) establishes that the structural invariants can fail — each has a planted counter-example — and therefore a green check is a positive result, not a silent absence. [Proposition 11](#) gives the digest its order-independence; detection gives the invariants their falsifiability.

5.5 The report envelope

A reader holding reports from several independent instruments needs one uniform way to say “this instrument, about this subject, at this review moment, said this — and here is the pointer to its complete native report.” The envelope is that data contract and nothing more. For this instrument the usual worry — that a selected status becomes a safe projection mistaken for the whole state — takes a specific form: Golden Line has *no* single overall verdict, so the only honest transportable status is the complete ordered vector of per-aspiration readings. The envelope carries exactly that vector, in report order, and invents no summary above it; compressing nine directional readings into one word would manufacture precisely the aggregate virtue score this instrument refuses everywhere else.

Definition 11 (Report envelope). The report envelope is the frozen record $v = (\text{schema_version}, \text{line_id}, \text{subject_id}, \text{review_date}, \text{native_status}, \text{report_ref}, \text{scope_and_nonclaims})$ with exactly those ten fields, in order, exported under the schema string `line.report-envelope/1.0`. `native_status` is the complete ordered sequence of (aspiration_id, status) pairs from the report’s findings — this line’s own vocabulary, one pair per registry aspiration ([Proposition 1](#)), never a summary. `report_ref` is the SHA-256 of the complete canonical report, so the envelope points at the full derivation — reasons trails, matched and ignored tokens, temporal flags, intake notes — rather than copying or restating any of it. `scope_and_nonclaims` carries the instrument’s transportable non-claims inside the record itself, so a stored envelope cannot quietly outgrow what the instrument was allowed to say. Sibling instruments export the same shape by publishing the same schema string, never by importing one another, and envelopes from different lines must not be compared, ranked, averaged, or merged on `native_status`.

Proposition 13 (The envelope points, never reinterprets). For every report r , `report_envelope(r)` satisfies `envelope_matches_report(envelope, r)`: the digest pointer, the review date, the registry version and digest ([Proposition 11](#)), and the per-aspiration readings all agree with the report they were exported from, and editing any of them afterwards makes the check return false. The envelope ([Definition 11](#)) adds no field the report does not determine except the caller-supplied `subject_id` and `source_snapshot_refs`, which the evaluator stores and does not verify. A matching envelope attests that the pair was archived unedited; it says nothing about the truth of the report or the merit of anything the report read.

5.6 Formalism-to-test bindings

Every definition and proposition above is verified by named tests in the package’s suite; the tables below bind each result to the test that would fail if the code stopped satisfying it. Each row is keyed on the block’s *label*, not on its number, and the label renders as the number the reader sees, so inserting a result rennumbers the prose and the table together and can never split them.

Two binding tests police the tables. `tests/test_formalism_bindings.py::test_binding_tables_bind_every_declared_block` fails if the set of row labels stops matching the set of labels declared in this section, and `tests/test_formalism_bindings.py::test_every_binding_row_names_an_existing_test` fails per row if any row’s verifying-test cell names no test or names one that does not exist. Neither checks that a named test is a *good* test, only that every declared block is bound to one that exists. The boundary column restates what each result does *not* claim.

Definition	Statement essence	Verifying test	Boundary
Definition 1	six fields, two of them token sequences	<code>tests/test_formalism_bindings.py::test_aspiration_tuple_matches_the_dataclasses</code>	names the fields, not what a marker is worth
Definition 2	nine entries, four founding then five further	<code>tests/test_formalism_bindings.py::test_registry_definition_matches_the_source_tuple</code>	a versioned design choice, not a canon
Definition 3	five fields, the observation date optional	<code>tests/test_formalism_bindings.py::test_horizon_entry_tuple_matches_the_dataclass</code>	records an observation, not a person
Definition 4	exactly the four <code>HorizonStatus</code> values, in enum order	<code>tests/test_formalism_bindings.py::test_status_codomain_matches_the_enumeration</code>	four readings, not a scale from bad to good
Definition 5	twelve fields, ignored markers kept apart from ignored counter-signals	<code>tests/test_formalism_bindings.py::test_finding_tuple_matches_the_dataclass</code>	exposes a derivation, not a justification
Definition 6	four parameters, two keyword-only; three named stages	<code>tests/test_formalism_bindings.py::test_report_function_signature_matches_manuscript</code>	a call shape, not a guarantee about inputs
Definition 7	shape, then membership, then prior admission; notes indexed from 1; never raises	<code>tests/test_formalism_bindings.py::test_intake_screening_tests_records_in_the_stated_order</code> , <code>tests/test_formalism_bindings.py::test_intake_screening_never_raises_on_hostile_input</code> , <code>tests/test_formalism_bindings.py::test_intake_first_wins_and_replay_determinism_match_manuscript</code>	screening judges records, never their authors
Definition 8	the three set operations, and undeclared tokens discarded	<code>tests/test_formalism_bindings.py::test_matching_set_operations_match_manuscript</code>	set membership, not sufficiency of evidence
Definition 9	enabled iff τ is set; missing or unparseable dates flag rather than fail	<code>tests/test_formalism_bindings.py::test_temporal_review_rule_matches_manuscript</code>	a data-quality window, not a decay law
Definition 10	six clauses, first match wins	<code>tests/test_formalism_bindings.py::test_decision_clauses_fire_in_the_stated_order</code>	clause order, not moral order

Proposition	Statement essence	Verifying test	Boundary
Proposition 1	one finding per registry aspiration; counts partition $n = 9$	<code>tests/test_progress.py::test_counts_summary, tests/test_formalism_bindings.py::test_totality_count_matches_manuscript</code>	report shape, not coverage of a life
Proposition 2	a declared counter-signal forces DRIFTING over every later clause	<code>tests/test_progress.py::test_drifting_survives_staleness</code>	flags a recorded signal, not a failing person
Proposition 3	TOWARD iff all markers observed, none countered, currentness auditable	<code>tests/test_progress.py::test_observed_and_unmet_fields_are_populated, tests/test_progress.py::test_fresh_observation_stays_toward, tests/test_figures.py::test_completeness_panel_gives_no_partial_credit</code>	a reading of one record, not an accreditation
Proposition 4	temporal uncertainty yields INQUIRY, never DRIFTING	<code>tests/test_progress.py::test_stale_observation_reverts_toward_to_inquiry, tests/test_progress.py::test_undated_entry_cannot_certify_currentness</code>	age reopens a question, never condemns
Proposition 5	NOT_OBSERVED arises solely from no admitted entry	<code>tests/test_progress.py::test_unknown_aspiration_id_is_noted_not_fatal, tests/test_progress.py::test_empty_entries_yield_all_not_observed</code>	absence of signal is not evidence of drift
Proposition 6	first valid entry per id stands; identical calls reproduce identical reports	<code>tests/test_progress.py::test_duplicate_entries_first_wins_and_is_noted, tests/test_formalism_bindings.py::test_intake_first_wins_and_replay_determinism_match_manuscript</code>	determinism of the reading, not truth of the records
Proposition 7	screening is a pure function of $(E, \text{ids}(R))$; malformed input is set aside deterministically; never raises	<code>tests/test_formalism_bindings.py::test_intake_screening_never_raises_on_hostile_input, tests/test_formalism_bindings.py::test_intake_screening_is_deterministic_on_all_input</code>	determinism of the screening function, not truth of records
Proposition 8	staleness is strict: TOWARD at age τ , stale from $\tau + 1$	<code>tests/test_progress.py::test_observation_exactly_at_staleness_boundary_stays_fresh, tests/test_progress.py::test_observation_one_day_past_staleness_boundary_is_stale, tests/test_formalism_bindings.py::test_currentness_boundary_constants_match_manuscript</code>	the threshold is a review-cadence choice, not a decay law

Proposition	Statement essence	Verifying test	Boundary
Proposition 9	sweep delegates to <code>progress_report</code> and computes no status; codomain is TOWARD/INQUIRY only	<code>tests/test_analysis.py::test_sweep_module_construc</code> <code>ts_no_status_of_its_own</code> , <code>tests/test_analysis.py::test_sweep_never_produces_d</code> <code>rifting_or_not_observed</code> , <code>tests/test_formalism_bindings.py::test_sweep_codoma</code> <code>in_matches_manuscript</code>	characterizes the instrument, not any observed practice
Proposition 10	every aspiration's reading flips at the same observation age	<code>tests/test_figures.py::test_lattice_cells_are_exec</code> <code>uted_evaluator_readings</code> , <code>tests/test_figures.py::test_lattice_reports_a_unifo</code> <code>rm_flip_age</code> , <code>tests/test_figures.py::test_lattice_de</code> <code>viation_count_detects_a_p</code> <code>lanted_outlier</code>	uniformity of the code path, not of any practice
Proposition 11	registry digest is permutation-invariant, 64 lowercase hex characters	<code>tests/test_serialization.py::test_canonical_regist</code> <code>ry_is_deterministic_json</code> , <code>tests/test_golden_line.py</code> <code>::test_registry_is_unique_and_order_independent</code> , <code>tests/test_formalism_bindings.py::test_digest_order_independence_matches_manu</code> <code>script</code>	a drift-review handle, no attestation semantics
Proposition 12	every invariant passes on the real registry and rejects a planted-bad one	<code>tests/test_invariants.py::test_battery_passes_on_r</code> <code>eal_registry</code> , <code>tests/test_invariants.py::test_regist</code> <code>ry_sound_false_on_any_planted_bad</code>	detection proof concerns the checks, not registry merit

The envelope section declares its blocks after the propositions above, so its rows sit in their own table, in the same document order:

Envelope block	Statement essence	Verifying test	Boundary
Definition 11	ten fields, <code>native_status</code> the complete ordered per-aspiration pairs	<code>tests/test_formalism_bindings.py::test_report_enve</code> <code>lope_tuple_matches_the_dataclass</code>	a data contract, not a summary and not a score
Proposition 13	the envelope agrees with its report field for field; any post-export edit is visible	<code>tests/test_formalism_bindings.py::test_envelope_po</code> <code>inter_matches_manuscript</code> , <code>tests/test_report_envelope.py::test_envelope_matches_report_verifies_an_arc</code> <code>hived_pair</code>	archival agreement, not truth of the report

The bindings are themselves code behavior: they show which claims the suite would catch, not that the registry's aspirations are wise or well served.

6 Scholarship: an aspiration is a direction, not a score

Golden Line’s four founding aspirations are not inventions. Each translates a larger argument into the narrow idiom of a versioned registry: attention before production, usefulness beyond the author, repairable systems, and technical work answerable to human flourishing. The translation is deliberately incomplete. Golden Line borrows questions and warnings from these traditions; it does not claim to reconcile them, turn them into a moral canon, or use their authority to certify a project.

6.1 A translation, not a synthesis

The scholarship contributes constraints on how to hold a direction. It does not provide a universal scoring function:

Lineage	What it contributes here	What Golden Line refuses to import
Aristotle and MacIntyre	Work becomes intelligible through ends, practical judgment, and goods internal to a practice [Aristotle, 350 BCE, MacIntyre, 1981].	A fixed rule that settles every particular case.
Sen, Nussbaum, Robeyns, and Jonas	Capability is answerable to real human possibilities and to future life, not valuable merely because it is power [Sen, 1980, 1999, Nussbaum, 2011, Jonas, 1984]. Which capabilities matter is settled by public reasoning inside an open framework, not fixed by a theorist [Sen, 2004, Robeyns, 2017].	The instrument’s author deciding what flourishing means for everyone affected, and any claim that a marker is a functioning.
Schön, Polanyi, Sennett, and Jackson	Skilled work is situated, partly tacit, reflective, and sustained through maintenance and repair [Schön, 1983, Polanyi, 1966, Sennett, 2008, Jackson, 2014].	Expert opacity, heroics, or a productivist bias that treats repair as secondary.
Ostrom and Winner	Shared goods need governance, while technical arrangements can embody power [Ostrom, 1990, Winner, 1980].	Treating “the commons” as frictionless openness or technology as socially neutral.
Campbell, Goodhart, Strathern, Merton, Espeland and Sauder, Manheim and Garrabrant, and Simon	Proxy targets distort conduct through several distinct mechanisms, and bounded agents need judgment rather than aggregate maximization [Campbell, 1979, Goodhart, 1975, Strathern, 1997, Merton, 1948, Espeland and Sauder, 2007, Manheim and Garrabrant, 2018, Simon, 1956].	A numeric Golden Line grade that would recreate the problem it names.

This is the relevant sense in which Golden Line is scholarly: its registry is a small design intervention informed by disagreements among practical philosophy, capability theory, sociology of evaluation, studies of work, and technology critique. None of those sources validates the registry as a universal theory of value. They help specify what the instrument must ask, what it must not pretend to know, and where it can be gamed.

The groupings above are not a claim of agreement. Aristotle and MacIntyre offer accounts of ends, practices, and judgment; Sen’s comparative capability approach does not collapse into Nussbaum’s list-and-threshold proposal, and Sen has said why he declines to supply a fixed list [Sen, 2004]; Jonas extends responsibility toward future life. Ostrom studies institutional arrangements for common-pool resources, not a universal recipe for openness. Winner makes a political argument about artifacts; Jackson gives an interpretive account of repair in a broken world. Goodhart’s original argument concerns monetary-policy targets, not a general law of human behavior. Golden Line cites these works as distinct design resources, not as a single theoretical foundation or empirical validation.

The evidentiary status matters. Most of this section is normative or interpretive scholarship: it clarifies concepts and exposes failure modes. Two citations carry empirical weight of different kinds — Ostrom’s comparative institutional research and Espeland and Sauder’s field study of law-school rankings — while the remaining measurement literature is historical, methodological, and sociological. None of the cited work tests Golden Line’s registry, evaluator, or visual language. The package’s own tests establish

implementation properties such as determinism and precedence; they do not establish that an aspiration is true, that a marker is socially sufficient, or that a directional reading predicts an outcome.

6.2 The shape of an end

The oldest commitment is that an activity has a point beyond its own motion. Aristotle's ethics is organized around the *telos*, the end that makes a practice intelligible, and around *phronesis*, practical wisdom that reads a particular situation rather than applying a fixed rule [Aristotle, 350 BCE]. Golden Line translates that distinction into a horizon and a human reading of the record. A horizon says what the work is trying to serve; it does not predetermine what a good next decision must look like.

MacIntyre sharpens the point for modern practice. The goods *internal* to a craft, the excellences reachable only by doing the work well, differ from the external rewards the same work can earn; a practice is weakened when the latter replace the former [MacIntyre, 1981]. The aspiration *make work useful beyond its author* therefore names more than distribution. It asks whether the work has acquired an excellence that can travel without being severed from the practice that made it trustworthy. The aspiration *prefer systems that can be repaired* makes a parallel demand: the work must remain open to correction rather than preserve the appearance of finishedness.

6.3 Capability and flourishing

Keep technical work answerable to human flourishing draws on the capability approach. Sen's Tanner Lecture put the prior question first — equality of *what?* — and argued that neither utility nor a bundle of primary goods answers it, because what matters is what a person is actually able to be and to do [Sen, 1980]. *Development as Freedom* carries that into policy: development is judged by real freedoms rather than by the accumulation of means [Sen, 1999]. Nussbaum develops a threshold account of central human capabilities, making the question concrete without reducing dignity to one aggregate index [Nussbaum, 2011]. Sen declines to fix such a list, holding that which capabilities matter depends on the purpose of the assessment and has to be settled by public reasoning rather than by the theorist [Sen, 2004]. Jonas extends the horizon: modern technology can alter distant and future life, so responsibility must include people who are not present at the moment of design [Jonas, 1984].

That disagreement about lists is the part of the literature Golden Line uses most directly, because the registry *is* a short list. Sen's objection applies to it without modification: nine entries chosen by one author are a starting point, not a settled account of what matters. So the registry is versioned and digest-identified rather than canonical, and revising it is an ordinary operation rather than an admission of error.

Robeyns marks the limit that matters most here. The capability approach is an open framework rather than a finished theory, and it supplies no evaluation metric by itself; the theories built inside it are what make the further choices about which capabilities count and how they are weighed [Robeyns, 2017]. Golden Line borrows the framework's question and none of those choices, and the transfer stops well short of the theory: a marker is not a functioning, an aspiration is not a capability, and a horizon report says nothing about any person's real freedoms. What the evaluator can do is make "capability treated as its own justification" visible as a counter-signal. It cannot determine whose flourishing is at stake, resolve a conflict among affected people, or convert a local record into a public welfare claim. The code records a direction and its derivation; it does not authorize the person who filed the record to speak for everyone downstream.

6.4 Attention, craft, and transfer

Let attention precede production is a claim about the sequence of judgment, not a celebration of slowness for its own sake. Schön's account of the reflective practitioner shows why competent work cannot be reduced to applying a formula: practitioners often discover and revise the problem while acting, through reflection-in-action [Schön, 1983]. Golden Line's "next decision" horizon is small enough to preserve that situated judgment. It asks for a visible pause before the next output, not for a ritual of documentation.

Polanyi explains why transfer is difficult: experts know more than they can straightforwardly tell [Polanyi, 1966]. Sennett's craftsman gives that difficulty a material and ethical form: quality grows through repeated attention to materials, error, and the desire to do a job well for its own sake [Sennett, 2008]. Jackson's "broken world" account pushes the argument further. Maintenance and repair are not merely costs after the productive act; they are sites where knowledge, creativity, power, and care become visible [Jackson, 2014].

These sources explain why the registry uses both Markers and Counter-signals. The markers declared for *make work useful beyond its author* are "handoff used" and "reader question answered": both test transfer by its effect on another person, without assuming that every piece of tacit knowledge can be compressed into a checklist. The counter-signal declared for *prefer systems that can be repaired* is "defect hidden to preserve appearance", which tests whether repair has been made ordinary. The signal is not a verdict on character; it is a prompt to inspect the conditions under which knowledge travels and failure can be discussed.

Confucian teaching adds the relational dimension: learning is cultivated through conduct with others, not detached from the practice of living together [Confucius, 500 BCE]. Frankl's account of meaning explains why a direction can remain orienting

through difficulty without becoming a guarantee of success [Frankl, 1959]. Ibn Khaldun supplies a political caution: purposes persist when institutions and shared cohesion carry them beyond an individual’s intention [Khaldun, 1377]. Golden Line therefore treats usefulness, teachability, and durable understanding as social properties, not merely private virtues.

6.5 Commons, power, and return

The aspiration to return improvements to the commons needs more precision than a generic call for openness. Ostrom’s work on common-pool resources shows that shared goods are sustained by situated institutions: boundaries, rules, monitoring, graduated response, and the capacity of participants to govern themselves [Ostrom, 1990]. The lesson for Golden Line is not that every artifact should be released without conditions. It is that “returned to the commons” should invite questions about who can use, inspect, adapt, maintain, and contest an improvement, and about which dependencies make that return possible.

Winner’s question, *Do Artifacts Have Politics?*, offers the complementary argument that technical things and systems can embody forms of power and authority [Winner, 1980]. Human flourishing cannot therefore be checked only at the level of a feature or repository. It also requires attention to the institutional arrangement in which the feature operates: who sets the terms, who bears the risk, whose labor remains invisible, and who can refuse. Golden Line does not answer those questions for the affected community. It keeps the questions from disappearing behind the language of capability or contribution.

6.6 Why it must not become a score

The literature on evaluation supplies Golden Line’s strongest negative design constraint, and it is older and more specific than the slogan it usually gets compressed into. Campbell stated the mechanism for social indicators: the more a quantitative indicator is used for social decision-making, the more it is subject to corruption pressures, and the more apt it is to distort the process it was meant to monitor [Campbell, 1979]. Goodhart’s 1975 paper addressed problems of monetary management and was not a universal moral law [Goodhart, 1975]; Strathern’s study of audit culture is where the compressed “measure becomes a target” phrasing entered wide circulation, along with the finding that target-driven measurement can displace the performance it was supposed to represent [Strathern, 1997].

Manheim and Garrabrant separate the failure into four mechanisms — regressional, extremal, causal, and adversarial — and note Campbell’s law arguably has scholarly precedence over the Goodhart formulations [Manheim and Garrabrant, 2018]. The taxonomy shows exactly how far Golden Line’s design helps. Emitting no aggregate removes one target: there is no single Golden Line number for selection pressure to push to an extreme. It does not remove the proxies. Each marker is itself a proxy, and an observer working for “marker present” rather than for the direction reproduces the regressional case one token at a time. Surfacing Counter-signals before Markers and reading missing evidence as inquiry keeps thin evidence from being rounded up. Nothing touches the adversarial case: an observer who wants a TOWARD can file the tokens that produce one, because the evaluator checks tokens against the registry and never the record against the world.

Espeland and Sauder give the empirical version of the same worry. Studying United States law-school rankings, they identify two mechanisms of reactivity — self-fulfilling prophecy and commensuration — by which a public measure changes the conduct it claims to describe [Espeland and Sauder, 2007]. Merton’s earlier account supplies the first in general form: a definition of a situation can evoke behavior that makes the definition come true [Merton, 1948]. Golden Line blocks commensuration by construction, because it emits no comparable number and never aggregates the nine entries into a grade. It does not escape the first. Its scope is also narrower than theirs in a way worth stating: Espeland and Sauder study a public ranking of organizations, while a horizon report is a local record with no audience beyond the people who filed it. Whether a private directional record produces reactivity at all is a question their study does not answer.

Simon supplies the constructive counterpoint. Under bounded rationality, agents work with limited information in structured environments, and the practical task is satisficing — deciding what is good enough for the situation at hand — rather than maximizing an abstract proxy [Simon, 1956]. Golden Line makes that restraint executable. It emits TOWARD, INQUIRY, DRIFTING, or NOT_OBSERVED; it surfaces counter-signals before markers; it treats missing evidence as inquiry; and it refuses to aggregate the nine entries into a single grade. A status is a directional reading of one admitted record, never a measure of a person, a project, or a whole moral life.

6.7 From theory to instrument

The translation can be stated compactly:

- *Telos* becomes a stated horizon, while *phronesis* remains with the reader.
- Capability theory becomes a question about affected lives, and a reason to keep the list of aspirations short, versioned, and open to revision by the people the work touches — not a license for capability to justify itself.
- Reflective practice, tacit knowing, craft, and repair become observable markers and counter-signals about how work is done and handed on.

- Commons scholarship becomes a demand to name governance, reciprocity, and maintenance rather than celebrate openness abstractly.
- Technology critique becomes a reminder that the surrounding institution is part of the object of attention.
- The sociology of targets becomes a hard prohibition on a Golden Line score, and an admission that no arrangement of clauses stops an observer who wants a particular reading from filing one.

The result is intentionally modest. Golden Line is a local instrument for keeping a purpose visible while work changes. Its scholarship makes that modesty more intelligent: a direction is useful only when it remains revisable, situated, answerable to people beyond its author, and honest about what the record cannot show.

7 Evidence boundary and artifact chain

Golden Line does not begin with a score. It begins with a separation of things that are easy to conflate: what the registry declares, what an observer records, what the evaluator can derive, and what a generated artifact can prove about its own provenance (formalized in [Definition 9](#) and [Proposition 8](#)).

7.1 Hard constraints and soft choices

The hard constraints are epistemic and semantic. An observation cannot create a signal that the registry did not declare. Missing evidence cannot become evidence of absence. A status must be reproducible from its inputs. Malformed temporal metadata cannot establish currentness. An aspiration cannot authorize what the Red Line refuses or resolve what the White Line leaves absent.

The soft choices are the nine-entry registry, exact token matching, the first-valid-entry rule for duplicates, the optional staleness window, and the names of the four statuses. They are versioned design decisions, not universal truths. A future registry may revise them, but it must revise the digest, figures, tests, and manuscript together.



Figure 5: The four HorizonStatus readings mapped to bounded evidence conditions: no admitted entry; incomplete, stale, or unauditable evidence; a declared counter-signal; and complete current/auditable evidence. The matrix is not a ranking.

7.2 Statuses as bounded claims

NOT_OBSERVED means that no valid entry was admitted. INQUIRY means that an entry exists but does not support a complete current positive reading. DRIFTING means that a declared counter-signal was recorded and takes precedence over markers. TOWARD means that every declared marker was observed, no declared counter-signal was recorded, and the observation was current when temporal review was enabled. None of these readings is a grade, a diagnosis, a safety claim, or a permission mechanism. [fig. 5](#)

The full reasons trail remains useful to a human reader, but structured fields (`observed`, `unmet`, `countered`, ignored tokens, note, date, and temporal flags) make the derivation directly inspectable. A consumer need not turn explanatory prose back into data before checking what happened.

Every reading returns the same record; which of its fields carry content is decided by what the matching stage found. [fig. 6](#) runs one entry per evidence condition and reports, field by field, what came back. Three fields are carried under every condition — the aspiration id, the status, and the reasons trail — so no reading arrives unexplained, and the remaining nine are carried exactly when the record supplied something for them. An empty field is an absent observation and nothing more.

THE SHAPE OF A FINDING

12 fields × 6 evidence conditions

Every reading returns the same record; the matching stage decides which fields carry content.
One executed progress_report call per column (review date 2026-07-18, stale_after_days = 90).

SOURCE-DRIVEN SCHEMATIC · a derivation trail, never a justification of what was recorded

	NOT_OBSERVED	DRIFTING	INQUIRY	INQUIRY	INQUIRY	TOWARD
	no admitted entry	counter-signal, one undeclared	markers incomplete, one undeclared	complete but stale	complete but undated	complete and current

FINDING FIELD

aspiration_id	carried	carried	carried	carried	carried	carried
status	carried	carried	carried	carried	carried	carried
reasons	carried	carried	carried	carried	carried	carried
observed	empty	carried	carried	carried	carried	carried
unmet	empty	empty	carried	empty	empty	empty
countered	empty	carried	empty	empty	empty	empty
ignored_markers	empty	carried	carried	empty	empty	empty
ignored_counter_signals	empty	carried	carried	empty	empty	empty
note	empty	carried	carried	carried	carried	carried
observed_on	empty	carried	carried	carried	empty	carried
stale	empty	empty	empty	carried	empty	empty
date_issue	empty	empty	empty	empty	carried	empty

3 fields carried under every condition: aspiration_id, status, reasons — no reading arrives unexplained.

Filled cells are carried and outlined cells are empty, and every cell prints which it is.

A carried flag is a recorded condition, not a good result: stale and date_issue are carried exactly when currentness could not be established.

An empty field is an absent observation: the record says what was seen, never what is true.

SOURCE-DRIVEN SCHEMATIC · field contents describe one synthetic record, never a person or a project

Figure 6: The twelve structured fields of a finding against six evidence conditions, one executed progress_report call per column. Filled cells are fields the returned record carries and outlined cells are empty, and every cell prints which it is. A carried temporal flag records a currentness problem, not a good result; an empty field records an absent observation and never asserts that what it names is false.

7.3 Source to publication

The local release chain is:

```
registry + status enum
  ↓
progress_report + invariant battery
  ↓
deterministic SVG, PNG, and JSON registries
  ↓
check_artifacts.py
  ↓
template PDF/HTML render and release audit
```

`check_artifacts.py` verifies the registry version and digest, the expected figure inventory, SVG/PNG pairs, generator identity, and figure-label citations in the manuscript. It is a consistency gate, not independent truth verification. A rendered artifact still needs visual and publication validation in the sibling template checkout.

8 The Golden Line aspirations

The registry opens with four founding aspirations. They are broad enough to travel across research and engineering, but each is paired with a concrete horizon so that it can be tested against a record rather than merely admired.

1. **Let attention precede production.** Make enough room to see what the work is actually doing at the next decision. *Horizon:* the next decision. *Counter-signal:* automatic output without review.
2. **Make work useful beyond its author.** Leave knowledge, tools, and explanations that another person can carry to a collaborator or public reader. *Horizon:* a collaborator or public reader. *Counter-signal:* private cleverness without transfer.
3. **Prefer systems that can be repaired.** Expose failure early and make correction ordinary in the revision cycle. *Horizon:* the revision cycle. *Counter-signal:* a defect hidden to preserve appearance.
4. **Keep technical work answerable to human flourishing.** Treat capability as a means whose value depends on the lives around it over a long horizon. *Horizon:* the long horizon. *Counter-signal:* capability treated as its own justification.

Five further entries extend the founding four in the versioned registry, filling out the long-horizon picture: **durable understanding** that outlasts the tool, **teachable craft** that can be handed to the next learner, **honest uncertainty** kept visible at the same prominence as the claim, at least one **unhurried question** measured in years rather than sprints, and improvements **returned to the commons** they came from. Each follows the same shape: a horizon, observable markers, and counter-signals ([Definition 1](#)), with the same reading rule ([Definition 10](#)). Together the nine entries and their signal structure are laid out in [fig. 8](#).

The order is not a ranking. The aspirations can conflict: usefulness beyond the author can pull against protecting an unhurried question; returning everything to the commons can pull against an obligation to a specific collaborator, and an honest record may need to show that conflict rather than resolve it. The companion White Line is the proper place to record what the registry cannot see or what should not be claimed.

The founding four are best read as a loop of correction, drawn in [fig. 7](#). Attention protects perception before output; usefulness tests whether what was learned can travel; repair keeps failure from becoming identity; answerability to human flourishing asks what the capability is for. No point is a maturity level. A project can move around this loop, lose one of its conditions, or find that two aspirations pull against each other. The useful question is therefore not *How high are we?* but *What does this record make visible, and what remains unasked?*

8.1 The signal vocabulary in aggregate

Read as a whole, the registry declares a deliberately small vocabulary: 18 markers and 9 counter-signals across the nine aspirations — two markers and one counter-signal per entry, with all 18 marker tokens and all 9 counter-signal tokens distinct across the registry. The `signal_inventory` helper in the analysis layer derives these counts from the live source, and [fig. 9](#) draws them one block per declared token. The distinctness matters: because no token is shared between aspirations, an observed marker can never accidentally support two directions at once. The blocks are vocabulary the evaluator can match, never observations and never points.

8.2 The reach of the nine horizons

The nine horizon phrases can also be grouped by *when* their direction becomes visible. The analysis layer declares four interpretive temporal-reach bands — immediate (1 aspiration, at the next decision), recurring cycle (2, at revision or tool turnover), at handoff (4, when the work reaches another person), and open-ended (2, over years) — and `horizon_distribution` places every registry entry into exactly one band, refusing loudly if a future registry adds a horizon the map does not classify. [fig. 10](#) shows the grouping. The bands are a reading aid declared outside the registry contract: band order widens reach, and a wider horizon is not a higher rank.

The aspirations are exercised in [the worked examples](#) and [the batch reading](#).

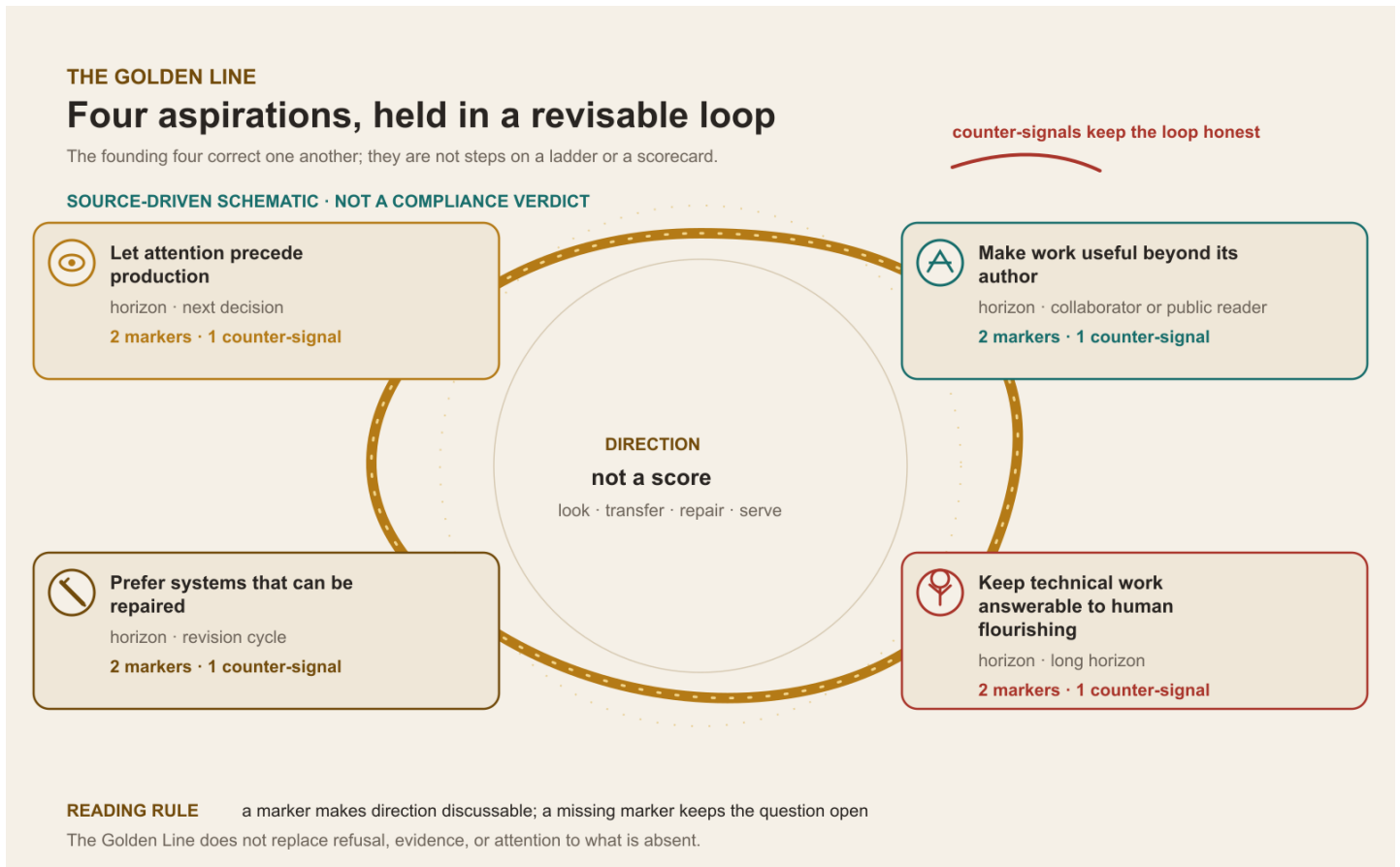


Figure 7: The four founding aspirations held in a revisable loop. Titles, horizons, and marker/counter-signal counts are source-derived; the connecting loop and icons are interpretive. The loop is direction, not a score or ranking.

THE ASPIRATION REGISTRY

9 versioned aspirations, drawn from source

Each row shows a horizon and the declared markers and counter-signals that make its readings reachable. No cell is a grade.

FOUNDING · 4 entries

Let attention precede production

horizon: next decision

2 markers

1 counter-signal

Make work useful beyond its author

horizon: collaborator or public reader

2 markers

1 counter-signal

Prefer systems that can be repaired

horizon: revision cycle

2 markers

1 counter-signal

Keep technical work answerable to human flourishing

horizon: long horizon

2 markers

1 counter-signal

FURTHER · 5 entries

Prefer understanding that outlasts the tool

horizon: tool turnover

2 markers

1 counter-signal

Work so the craft can be taught

horizon: next learner

2 markers

1 counter-signal

Keep uncertainty visible in finished work

horizon: public claim

2 markers

1 counter-signal

Hold at least one unhurried question

horizon: multi-year inquiry

2 markers

1 counter-signal

Return improvements to the commons

horizon: shared pool

2 markers

1 counter-signal

SOURCE-DRIVEN SCHEMATIC · marker and counter-signal counts describe reachability, not performance

Figure 8: The full nine-entry aspiration registry: four founding aspirations and five further entries, drawn from the versioned source. Each row shows its horizon and declared marker/counter-signal counts; the layout is a taxonomy for review, not a performance scale.

THE SIGNAL INVENTORY

18 declared markers, 9 counter-signals

One block per declared token, aggregated from the versioned registry.

The blocks are vocabulary the evaluator can match, not observations and not points.

UNIQUENESS · 18 of 18 marker tokens distinct · 9 of 9 counter-signal tokens distinct

Let attention precede production horizon: next decision	☒ ☒	☐
Make work useful beyond its author horizon: collaborator or public reader	☒ ☒	☐
Prefer systems that can be repaired horizon: revision cycle	☒ ☒	☐
Keep technical work answerable to human flourishing horizon: long horizon	☒ ☒	☐
Prefer understanding that outlasts the tool horizon: tool turnover	☒ ☒	☐
Work so the craft can be taught horizon: next learner	☒ ☒	☐
Keep uncertainty visible in finished work horizon: public claim	☒ ☒	☐
Hold at least one unhurried question horizon: multi-year inquiry	☒ ☒	☐
Return improvements to the commons horizon: shared pool	☒ ☒	☐

☒ declared marker (a sign of movement toward)

☐ declared counter-signal (a sign of drift)

SOURCE-DRIVEN SCHEMATIC · counts describe the declared vocabulary, never fulfilment or performance

Figure 9: The aggregate declared-signal vocabulary of the registry, one unit block per token: filled blocks for markers, outlined blocks for counter-signals. The counts describe what the evaluator can match, never fulfilment or performance.

THE HORIZON-BAND DISTRIBUTION

9 horizons, 4 reaches of visibility

An interpretive grouping declared in the analysis layer: each aspiration is placed by when its direction becomes visible. Band order widens reach; it does not rank importance.

SOURCE-DRIVEN SCHEMATIC · the bands are a reading aid, not registry contract and not a maturity ladder

IMMEDIATE

1 aspiration

Let attention precede production

"next decision"

RECURRING CYCLE

2 aspirations

Prefer systems that can be repaired

"revision cycle"

Prefer understanding that outlasts the tool

"tool turnover"

AT HANDOFF

4 aspirations

Make work useful beyond its author

"collaborator or public reader"

Work so the craft can be taught

"next learner"

Keep uncertainty visible in finished work

"public claim"

Return improvements to the commons

"shared pool"

OPEN-ENDED

2 aspirations

Keep technical work answerable to human flourishing

"long horizon"

Hold at least one unhurried question

"multi-year inquiry"

SOURCE-DRIVEN SCHEMATIC · a wider horizon is not a higher rank; every band is revisited in its own time

Figure 10: The nine aspiration horizons grouped into four temporal-reach bands — immediate, recurring cycle, at handoff, and open-ended. An interpretive reading aid, not a maturity ladder and not part of the registry contract.

9 Worked records

The model is intentionally small enough to inspect at the command line. A single horizon entry against one aspiration produces a full report over the whole registry:

```
from golden_line import HorizonEntry, progress_report

report = progress_report([
    HorizonEntry(
        aspiration_id="repairable-systems",
        observed_markers=frozenset({"failure named", "revision attempted"}),
        note="The failure was recorded before the release note was written.",
    )
])
```

That entry yields TOWARD for `repairable-systems` and NOT_OBSERVED for the other eight registry items, because no horizon note was recorded for them. The readings move exactly as the decision rule prescribes:

- If the same entry also records the counter-signal `defect hidden to preserve appearance`, the finding becomes DRIFTING. Counter-signal precedence ([Proposition 2](#)) means this holds even if both markers are still present; a positive signal cannot launder a recorded drift.
- If only one of the two declared markers is present, the finding becomes INQUIRY with an `unmet` list naming the marker still outstanding, so the direction is recorded as open rather than reached — no partial credit ([Proposition 3](#)).
- If both markers are present but the observation carries an old, future, or missing/unparseable `observed_on` date and the caller passes `stale_after_days`, the finding reverts to INQUIRY: temporal uncertainty reopens the question without inventing drift ([Proposition 4](#)).
- An entry naming an aspiration that is not in the registry is set aside during intake ([Definition 7](#)) and reported in the intake notes; it never crashes the report and never silently disappears.
- A malformed record, such as one whose marker field is `None`, is also set aside and named in the intake notes ([Definition 7](#)). If a dated positive record is supplied with temporal review enabled but its date is not ISO-parseable, the finding is INQUIRY with `date_issue=True` ([Proposition 4](#)): malformed time metadata cannot certify currentness.

The first of those bullets is the one most worth distrusting, because it is the rule a reader is most likely to assume has an exception. It does not. Running all nine aspirations through four evidence conditions — complete markers alone, complete markers with one declared counter-signal, a counter-signal with no markers, and a counter-signal on an observation four hundred days old — returns DRIFTING in every cell where a counter-signal is present, including all nine cells where every declared marker is also present. [fig. 12](#) is that run: each cell in it is a `progress_report` return value rather than a box drawn to illustrate one.

These distinctions keep a positive signal from laundering a counter-signal ([Proposition 2](#)), and keep the absence of evidence from masquerading as either success or failure ([Proposition 5](#)). The path from an entry to its reading is shown in [fig. 1](#), and the three evaluator stages an entry passes through — including the set-aside branch that carries malformed, unknown-id, and duplicate records into the intake notes — are shown in [fig. 11](#), while the bounded meaning of all four readings is summarized in [fig. 5](#).

A single entry shows the decision rule in isolation. The next section runs a whole batch — positive, drifting, partial, malformed, and duplicate entries at once — through the same evaluator and the descriptive analysis layer, so the gap between what was submitted and what was read becomes fully accountable.

THE STAGED EVALUATION PIPELINE

progress_report runs three stages

Malformed or unexpected input is recorded, never raised;
every finding carries a full reasons trail.

1 · INTAKE SCREENING

One entry per known id (first wins). Malformed records, unknown ids, and duplicates are set aside into intake notes.

2 · MATCHING

Observed markers and counter-signals are intersected with the declared sets. Undeclared tokens are ignored and noted.

3 · DECISION

The matched sets, with optional temporal review, select one of the four directional readings.

OUTCOME: 9 findings — one per aspiration, each a direction, never a grade.

SOURCE-DRIVEN SCHEMATIC · set-aside notes are actual progress_report intake notes · NOT A COMPLIANCE VERDICT

SET ASIDE · 3 kinds → intake_notes

notes quoted verbatim from the evaluator

MALFORMED RECORD

entry #1 was set aside: expected a HorizonEntry record

UNKNOWN ID

entry for unknown aspiration
'not-a-real-id' was set aside

DUPLICATE ENTRY

duplicate entry for
'attention-before-output' was set aside;
the first entry stands

set aside, named, and counted — never silent

UNDECLARED TOKENS

ignored and noted; they can never
determine a status

Figure 11: A batch of horizon entries passes through intake screening, signal matching, and decision. Records that fail intake — malformed, unknown-id, or duplicate — are set aside into visible intake notes (quoted verbatim from a real progress_report replay), undeclared tokens are ignored and noted, and the output is one bounded finding per registry aspiration. Findings are directional readings, never grades.

COUNTER-SIGNAL PRECEDENCE

9 of 9 aspirations still read DRIFTING

Each cell is one real progress_report call on a synthetic entry for that aspiration
(review date 2026-07-18, stale_after_days = 90).

SOURCE-DRIVEN SCHEMATIC · a replayed clause ordering, never a judgement of the work or its author

	complete markers, no counter-signal	complete markers, one counter-signal	no markers, one counter-signal	complete markers, one counter-signal, 400 d old
ASPIRATION				
attention-before-output	TOWARD	DRIFTING	DRIFTING	DRIFTING
useful-to-others	TOWARD	DRIFTING	DRIFTING	DRIFTING
repairable-systems	TOWARD	DRIFTING	DRIFTING	DRIFTING
wide-human-flourishing	TOWARD	DRIFTING	DRIFTING	DRIFTING
durable-understanding	TOWARD	DRIFTING	DRIFTING	DRIFTING
teachable-craft	TOWARD	DRIFTING	DRIFTING	DRIFTING
honest-uncertainty	TOWARD	DRIFTING	DRIFTING	DRIFTING
unhurried-questions	TOWARD	DRIFTING	DRIFTING	DRIFTING
commons-returned	TOWARD	DRIFTING	DRIFTING	DRIFTING

Complete markers plus one declared counter-signal returns DRIFTING for 9 of 9 aspirations; at an observation 400 days old it still does, for 9.

Staleness reopens a positive reading; it never erases a recorded drift, and drift never becomes a verdict on a person.

SOURCE-DRIVEN SCHEMATIC · filled cells are DRIFTING, outlined cells are not · NOT A COMPLIANCE VERDICT

Figure 12: Counter-signal precedence replayed rather than drawn: for each of the nine aspirations, four progress_report calls at review date 2026-07-18 with stale_after_days = 90. Complete markers with no counter-signal read TOWARD; adding one declared counter-signal returns DRIFTING even though every marker is still present, and it still returns DRIFTING when the same observation is stale. Filled cells are DRIFTING and outlined cells are not, so the panel reads without colour. The panel characterizes clause order; it does not grade any work or person.

10 Reading a batch: the descriptive layer at work

A single horizon entry is easy to read by eye. A real review is rarely a single entry: it is a batch of observations filed against several aspirations at once, some positive, some drifting, some partial, and a few malformed. The descriptive analysis layer exists to characterize such a batch without adding any claim the evaluator did not already make. This section runs one batch end to end so the helpers can be seen doing exactly — and only — what the formalism permits.

10.1 A worked batch

Consider six submitted horizon entries evaluated against the shipped nine-entry registry with `progress_report`:

```
from golden_line import HorizonEntry, progress_report, report_overview

entries = [
    HorizonEntry("attention-before-output",
        observed_markers=frozenset({"question revisited", "context named"})),
    HorizonEntry("repairable-systems",
        observed_markers=frozenset({"failure named", "revision attempted"}),
        counter_signals=frozenset({"defect hidden to preserve appearance"})),
    HorizonEntry("useful-to-others",
        observed_markers=frozenset({"handoff used"})),
    HorizonEntry("honest-uncertainty",
        observed_markers=frozenset({"limit stated beside claim",
                                     "confidence qualified in print",
                                     "extra token that is undeclared"})),
    HorizonEntry("not-a-real-id",
        observed_markers=frozenset({"whatever"})),
    HorizonEntry("attention-before-output",
        observed_markers=frozenset({"question revisited"})),
]
report = progress_report(entries)
overview = report_overview(report)
```

Six entries were submitted, but the report still contains exactly nine findings — one per registry aspiration — because totality ([Proposition 1](#)) does not depend on what was filed. `report.counts()` partitions those nine findings as **two TOWARD, one INQUIRY, one DRIFTING, and five NOT_OBSERVED**. The `report_overview` helper regroups the same findings by status and totals the intake so the batch can be described without re-parsing prose:

Reading	Count	Aspirations
TOWARD	2	attention-before-output, honest-uncertainty
INQUIRY	1	useful-to-others
DRIFTING	1	repairable-systems
NOT_OBSERVED	5	wide-human-flourishing, durable-understanding, teachable-craft, unhurried-questions, commons-returned

Each reading follows the decision rule of the formalism, and each illustrates one of its guarantees:

- `attention-before-output` recorded both of its declared markers and no counter-signal, so it reads TOWARD.
- `repairable-systems` recorded **both** declared markers *and* a declared counter-signal. Counter-signal precedence ([Proposition 2](#)) surfaces the drift first: the reading is DRIFTING, and the two present markers cannot launder it.
- `useful-to-others` recorded only one of its two declared markers, so the direction is held open as INQUIRY with the second marker still unmet — not rounded up and not condemned ([Proposition 3](#)).
- `honest-uncertainty` recorded both declared markers plus one token, "extra token that is undeclared", that the registry never declares. The undeclared token is ignored, counted once in the overview's `ignored_marker_total`, and changes nothing: because both declared markers were present and no counter-signal was recorded, the reading is TOWARD. An observer-supplied token can never smuggle in a status ([Definition 8](#)).
- The five aspirations with no admitted entry read NOT_OBSERVED. This is not evidence that those directions are absent or failing; it reports only that no valid record reached this report ([Proposition 5](#)).

10.2 What intake set aside

Two of the six submitted entries never became findings, and the overview makes the reason legible without any narrative: `intake_note_count` is **2**. The `report.intake_notes` are, verbatim:

- "entry for unknown aspiration 'not-a-real-id' was set aside"
- "duplicate entry for 'attention-before-output' was set aside; the first entry stands"

The unknown identifier was screened out during intake rather than crashing the report; the duplicate `attention-before-output` entry — a weaker record with only one marker — was set aside by the first-valid-entry rule, and the first, complete entry stands. Neither disappeared silently: both are named in the intake notes so the gap between six submissions and four incorporated readings is fully accounted for. For this batch the overview's `ignored_marker_total` is **1**, its `ignored_counter_signal_total`, `stale_count`, and `date_issue_count` are all **0** (temporal review was left disabled here), so the entire delta between raw input and final readings is visible in five small integers.

The whole batch is drawn in fig. 13: the figure builder replays exactly this batch through `progress_report` and `report_overview`, so every count, grouping, and quoted intake note in the picture is the evaluator's actual output for the code block above, not an illustration of it.

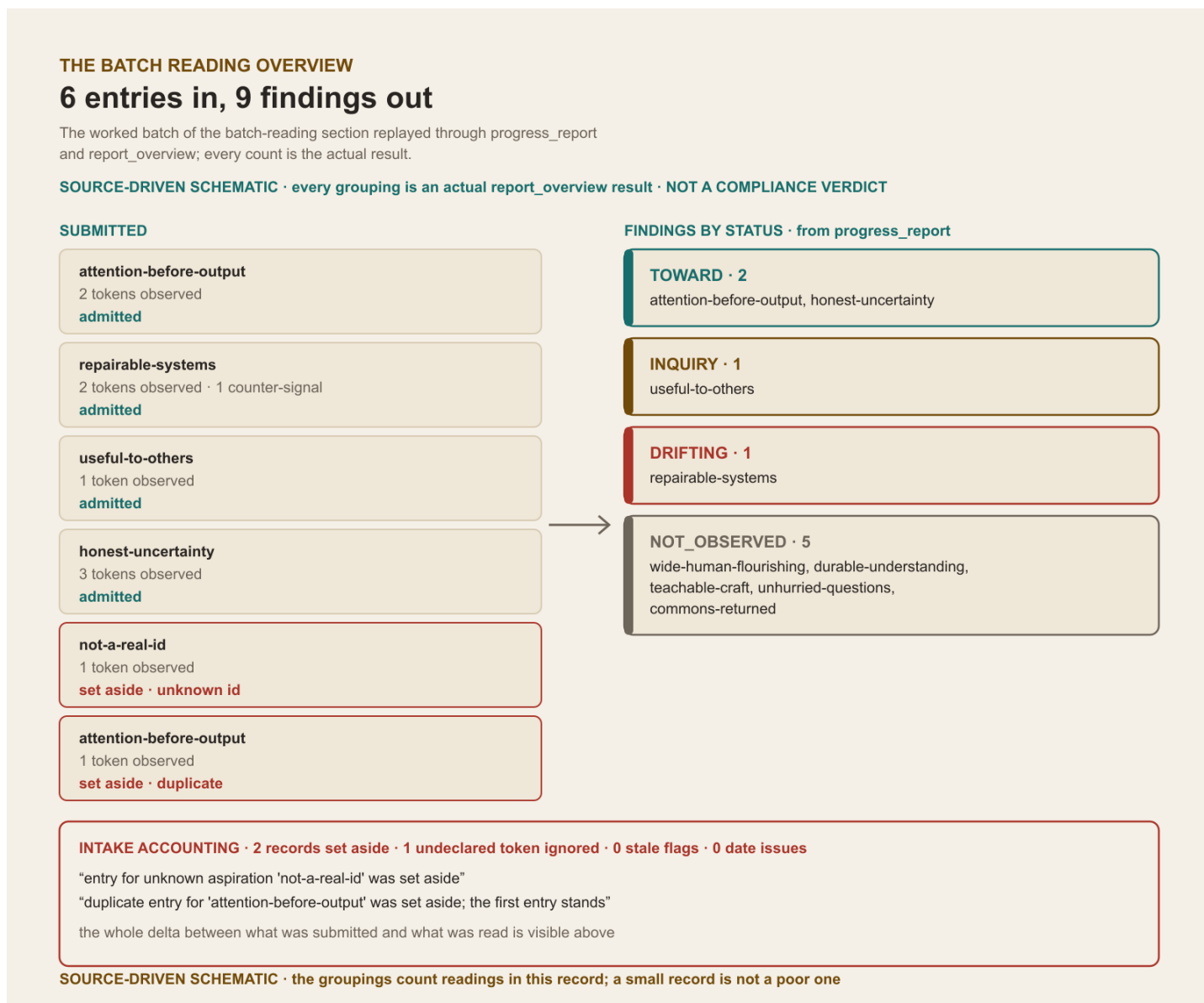


Figure 13: The 6-entry worked batch of this section replayed through the real evaluator: 2 TOWARD, 1 INQUIRY, 1 DRIFTING, and 5 NOT_OBSERVED across 9 findings — one per registry aspiration whether or not an entry was filed — with 2 intake set-asides and 1 ignored undeclared token quoted verbatim from the report. The groupings count readings in this record; they do not grade the work or the people behind it.

10.3 Vocabulary and reach as reading context

The overview describes what *this batch* produced; the signal inventory and horizon-band distribution describe the fixed vocabulary and temporal reach that any batch is read against. Because every one of the registry’s eighteen markers and nine counter-signals is a distinct token (fig. 9), the `ignored_marker_total` above is unambiguous: the ignored token matched no declared marker of any aspiration, not merely the one it was filed under. And because the four aspirations that drew readings here — attention, repair, usefulness, honest uncertainty — sit in three different temporal-reach bands (fig. 10), a single review batch routinely mixes an immediate-horizon reading with an at-handoff one. The bands are a reminder that these readings become visible on different clocks, not a schedule on which they should be expected to agree.

None of these summaries is a grade. `report_overview` adds no claim the findings did not already carry; it only counts them. A batch that reads two TOWARD and five NOT_OBSERVED is not “worse” than one that reads nine TOWARD. It is a smaller record.

11 Limits and safeguards

Aspiration language can become sentimental, coercive, or falsely universal, and a directional instrument can be misread as a verdict. Golden Line therefore carries six hard limits. Three of them have a mechanical core the code enforces, and the proposition that pins each one is named inline; the other three are commitments the code cannot hold for us, and saying so is part of keeping them.

- *Commitment*. A registry entry is a design choice, not a discovery of humanity’s single highest good. The nine aspirations are a starting set, versioned and revisable, not a closed canon.
- *Enforced* (*Proposition 3*). A TOWARD result says only that declared markers were observed, no counter-signal was recorded, and (when temporal review is enabled) currentness was auditable from a present, parseable, in-window date. Those conditions are necessary and sufficient in the code. That the result does not prove the work good, safe, lawful, or beneficial is a reading rule, not a check.
- *Enforced* (*Proposition 5*). A NOT_OBSERVED result arises solely from the absence of an admitted entry. It is not evidence that the aspiration is absent; submitted records may have been malformed, unknown, or duplicate, and the intake notes say which.
- *Commitment*. Counter-signals are local warnings about a record, not psychological diagnoses or public labels for a person or project. Nothing in the evaluator can stop a reader from using one that way.
- *Enforced* (*Definition 7, Proposition 4*). A malformed entry is visible as an intake set-aside, and a missing or malformed date in an enabled temporal review prevents a positive record from being read as current. These are data-quality safeguards, not proof that the underlying work failed.
- *Commitment*. The instrument cannot resolve conflicts among affected people; it can only make the chosen direction and its tradeoffs easier to discuss. This is an absence rather than a check, and no test can confirm it.

The sharpest risk is the one the scholarship section names: that a published aspiration hardens into a target and is gamed rather than served [Strathern, 1997], or that stating the direction as a scorecard evokes the very behavior it was meant to describe [Merton, 1948]. The design pushes against this at every stage: no numeric score, no aggregate grade, counter-signals surfaced first, absence read as inquiry, and no positive currentness claim from malformed temporal data. What it does not touch is the adversarial case [Manheim and Garrabrant, 2018]: the evaluator matches the tokens an observer filed against the registry, so an observer who wants a TOWARD can file the tokens that produce one, and no clause ordering prevents it. The honest safeguard is to keep the readings directional and the responsibility human, in the spirit of Jonas’s long-horizon duty rather than a compliance ritual [Jonas, 1984].

11.1 Limits of the descriptive analysis layer

The descriptive helpers — `signal_inventory`, `horizon_distribution`, `temporal_currentness_sweep`, and `report_overview` — introduce their own, smaller risks of over-reading, and each is deliberately constrained so it cannot outrun the evaluator it sits beside.

- The inventory counts declared *vocabulary*, never fulfilment. That the registry declares eighteen markers and nine counter-signals says nothing about how many have ever been observed; a larger marker count is not a higher standard, only a longer list of things one could look for. Reading the inventory as a scoreboard would invert its purpose.
- The horizon bands are an interpretive reading aid declared in the analysis layer, not part of the registry contract. Their order widens temporal reach; it does not rank merit, and an “open-ended” horizon is not superior to an “immediate” one. The band map is intentionally brittle: an aspiration whose horizon it does not classify raises an error rather than being dropped, so a future registry change forces the map to be revisited by a human instead of silently mis-grouping an entry.
- The currentness sweep replays **one** fully-marked, counter-signal-free synthetic entry to expose where a reading loses currency. It is a probe of the evaluator’s temporal rule under a chosen `stale_after_days`, not a claim about any real observation, and the staleness boundary it draws is a data-quality threshold — a `stale` cell is an invitation to look again, never evidence that the underlying work failed. The lattice widens the probe to every registry entry, which rules out the boundary being an artifact of the one entry chosen; it does not make the entries any less synthetic.
- `report_overview` regroups and counts findings that already exist; it adds no claim the report did not carry and derives nothing from outside the report’s own structured fields. Its per-status tallies are a summary, and a batch that reads mostly NOT_OBSERVED is a small record, not a poor one.

Because all four helpers are pure, deterministic, and read-only, they cannot change what a reading means — but a reader can still misuse a number. The safeguard is the same one the whole instrument relies on: keep the summaries descriptive and the judgement human.

For security boundaries, use Red Line. For how the work was performed, use Black Line. For what is missing, withheld, unobserved, or ethically left unclaimed, use White Line. A horizon report must never be used to bypass those scopes, and a TOWARD on any Golden Line aspiration can never authorize what Red Line refuses.

12 Conclusion

Golden Line keeps a question alive: what is this work for, and what would movement toward that purpose look like at the next honest horizon? Its answer is not a universal doctrine. It is a versioned set of nine threads, each with a horizon, observable markers, and counter-signals, read by a staged evaluator that returns a direction and never a grade.

The formalism shows how modest the machine is: four directional readings ([Definition 4](#)), a decision rule in which a counter-signal outranks any marker ([Proposition 2](#)) and staleness or a malformed reviewed date reopens rather than condemns ([Proposition 4](#)), and seven structural invariants each kept honest by a planted-bad proof-of-detection test ([Proposition 12](#)). Six of the thirteen figures replay those rules rather than draw them: precedence ([Proposition 2](#)), marker completeness ([Proposition 3](#)), the currentness boundary ([Proposition 8](#)), and the shape of the returned record are all read off executed output ([Proposition 9](#)). source, figures, and manuscript citations remain joined.

The scholarship shows both the age and the stakes of the commitments (see [the scholarship section](#)): ends and practical wisdom, internal goods, capability and flourishing, reflective craft, repair, commons governance, the political character of technical systems, and the long-horizon duty that technological power now demands. It also shows why the instrument must stay anti-metric: a direction hardened into a target stops measuring what it named, and an open commons without governance can simply hide who carries the cost.

The instrument is strongest when paired with the restraint of its companions, each of which does a job it cannot. Golden Line adds a direction without turning aspiration into authority, a thread worth reaching toward, held in the open, and revised whenever the evidence or the values change.

The work is built for DAF's public research index at [docxology](#), machine-readable, cross-linked, and verification-logged, with its eventual public home at [docxology/golden_line](#). The aspirational thread is meant to be kept in plain sight.

References

- Aristotle. *Nicomachean Ethics*. Various critical editions, 350 BCE. Traditionally dated to the fourth century BCE; on telos, eudaimonia, the mean, and practical wisdom (phronesis). Translated editions vary. Accessed 2026-07-18.
- Donald T. Campbell. Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1):67–90, 1979. doi: 10.1016/0149-7189(79)90048-X. Source of Campbell’s law: the more a quantitative social indicator is used for social decision-making, the more it is subject to corruption pressures and the more apt it is to distort the processes it was meant to monitor. Accessed 2026-07-28.
- Confucius. *The Analects*. Various critical editions, 500 BCE. Compiled and transmitted in early Chinese traditions; date is approximate. On learning as the cultivation of conduct within relationships. Accessed 2026-07-18.
- Wendy Nelson Espeland and Michael Sauder. Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1):1–40, 2007. doi: 10.1086/517897. Empirical study of law-school rankings identifying two mechanisms of reactivity — self-fulfilling prophecy and commensuration — by which a public measure changes the conduct it claims to describe. Accessed 2026-07-28.
- Viktor E. Frankl. *Man’s Search for Meaning*. Beacon Press, 1959. English translation; German original *Ein Psychologe erlebt das Konzentrationslager* (1946). On meaning as an orienting direction. Accessed 2026-07-18.
- Charles A. E. Goodhart. Problems of monetary management: The u.k. experience. In *Papers in Monetary Economics*, volume 1, pages 1–20. 1975. URL <https://www.econbiz.de/Record/problems-of-monetary-management-the-u-k-experience-goodhart-charles/10002525062>. Original monetary-policy context for the control-target warning later associated with Goodhart’s law. Accessed 2026-07-18.
- Steven J. Jackson. Rethinking repair. In *Media Technologies: Essays on Communication, Materiality, and Society*. MIT Press, 2014. doi: 10.7551/mitpress/9780262525374.003.0011. URL <https://academic.oup.com/mit-press-scholarship-online/book/14976/chapter-abstract/169335302>. On broken-world thinking and repair as a site of creativity, knowledge, power, and care. Accessed 2026-07-18.
- Hans Jonas. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. University of Chicago Press, 1984. English edition; German original *Das Prinzip Verantwortung* (1979). On long-horizon duty toward future life. Accessed 2026-07-18.
- C. G. Jung. *Psychology and Alchemy*. Collected Works of C. G. Jung, Volume 12; Bollingen Series XX. Princeton University Press, 1953. German original *Psychologie und Alchemie* (1944). Reads nigredo, albedo, citrinitas, and rubedo as figures for psychological transformation rather than as laboratory chemistry. Cited here only for that symbolic register, which is the register this project’s name is used in. Accessed 2026-07-27.
- Ibn Khaldun. *The Muqaddimah: An Introduction to History*. Princeton University Press (Rosenthal translation, 1958), 1377. Prolegomena to the history of the world; on how purposes and cohesion are carried by institutions. Accessed 2026-07-18.
- Alasdair MacIntyre. *After Virtue: A Study in Moral Theory*. University of Notre Dame Press, 1981. On practices, goods internal to a practice, and the narrative unity of a life toward a telos. Accessed 2026-07-18.
- David Manheim and Scott Garrabrant. Categorizing variants of goodhart’s law. arXiv preprint arXiv:1803.04585, 2018. URL <https://arxiv.org/abs/1803.04585>. Separates four mechanisms by which optimizing a proxy collapses its relationship to the goal — regressional, extremal, causal, and adversarial — and notes that Campbell’s law arguably has scholarly precedence over the later Goodhart formulations. Accessed 2026-07-28.
- Robert K. Merton. The self-fulfilling prophecy. *The Antioch Review*, 8(2):193–210, 1948. doi: 10.2307/4609267. On how a definition of a situation can evoke behavior that makes the definition come true. Accessed 2026-07-18.
- Martha C. Nussbaum. *Creating Capabilities: The Human Development Approach*. The Belknap Press of Harvard University Press, 2011. On the central human capabilities and a threshold conception of flourishing. Accessed 2026-07-18.
- Elinor Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, 1990. doi: 10.1017/CBO9780511807763. URL <https://www.cambridge.org/core/books/governing-the-commons/7AB7AE11BADA84409C34815CC288CD79>. On self-governance and institutional diversity in common-pool resource management. Accessed 2026-07-18.
- Michael Polanyi. *The Tacit Dimension*. Doubleday; reprinted University of Chicago Press, 1966. On tacit knowing: *we can know more than we can tell*. Accessed 2026-07-18.
- Ingrid Robeyns. *Wellbeing, Freedom and Social Justice: The Capability Approach Re-Examined*. Open Book Publishers, Cambridge, UK, 2017. doi: 10.11647/OBP.0130. URL <https://www.openbookpublishers.com/books/10.11647/obp.0130>. Open-access re-

- examination distinguishing the capability approach as an open framework from the specific theories built inside it; the framework does not itself supply a metric. Accessed 2026-07-28.
- Donald A. Schön. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, 1983. URL <https://www.routledge.com/The-Reflective-Practitioner-How-Professionals-Think-in-Action/Schn/p/book/9781138458284>. On reflection-in-action and the limits of reducing professional judgment to technical formulas. Accessed 2026-07-18.
- Amartya Sen. Equality of what? In Sterling M. McMurrin, editor, *The Tanner Lectures on Human Values, Volume 1*. Cambridge University Press, 1980. URL <https://tannerlectures.org/lectures/equality-of-what/>. Tanner Lecture on Human Values delivered at Stanford University, 22 May 1979. Argues that utilitarian, total-utility, and Rawlsian equality each answer the wrong question, and introduces basic capability equality. Accessed 2026-07-28.
- Amartya Sen. *Development as Freedom*. Alfred A. Knopf; Oxford University Press, 1999. On the capability approach: value measured in real freedoms to be and to do. Accessed 2026-07-18.
- Amartya Sen. Capabilities, lists, and public reason: Continuing the conversation. *Feminist Economics*, 10(3):77–80, 2004. doi: 10.1080/1354570042000315163. Sen’s reply on why he declines to fix a canonical list of capabilities: the relevant capabilities depend on the purpose of the assessment and must be settled by public reasoning. Accessed 2026-07-28.
- Richard Sennett. *The Craftsman*. Yale University Press, 2008. On craft as the desire to do a job well for its own sake, and on repair as a form of skilled attention. Accessed 2026-07-18.
- Herbert A. Simon. Rational choice and the structure of the environment. *Psychological Review*, 63(2):129–138, 1956. doi: 10.1037/h0042769. On satisficing and bounded rationality: agents seek *good enough* within a structured environment. Accessed 2026-07-18.
- Marilyn Strathern. ‘improving ratings’: audit in the british university system. *European Review*, 5(3):305–321, 1997. doi: 10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4. Source of the widely quoted formulation that when a measure becomes a target it ceases to be a good measure. Accessed 2026-07-18.
- Langdon Winner. Do artifacts have politics? *Daedalus*, 109(1):121–136, 1980. URL <https://www.jstor.org/stable/20024652>. On technical artifacts and systems embodying forms of power and authority. Accessed 2026-07-18.