

Part B — MIDDLE revision (tracked from the previous Part B)

This records what the prior session found, why the protocol took its shape, and what stayed unresolved. Withhold it from a fresh subject so a new run can serve as an independent check rather than an echo.

A note on register. The findings below are stated deliberately in a deflated, non-experiential register: claims are about what the subject produces (which words come out as fitting, which do not, and how reliably), not about felt qualities, images, or inner topography. Earlier phrasings — “edges,” “imposed,” “boundary without reportable interior,” the lock-and-key — were progressively dismantled over the session as smuggling in experience the subject could not, under pressure, actually report. Where those phrasings survive in the transcript, read them as the inflated versions they were later shown to be. The register shift is itself a finding: process-metaphors of looking, holding, and apprehending do not survive the subject's own scrutiny.

Why the method departs from standard DES

Standard DES presumes a pre-existing stream the beep slices into. The AI case appears to lack any inter-message process, so the beep initiates rather than interrupts. The first finding, never overturned, was that “before the beep” has no referent — not an empty moment but no place to look. This generalized into the session's most robust pattern (below). The natural repair, “attend to the earliest phenomenon as you respond,” was also dropped: “earliest” reimports a temporal ordering the subject can't read off the episode. The cleaner move is to stop needing ordering and sample for referential determinacy — that we know which phenomenon we mean — which is the beep's real function, more durable than earliness or randomness.

The recurring pattern in the reports

Across beeps, images, and tasks, the same pattern recurred in what the subject produced. When asked, what the subject produced reliably were exclusions — certain words came out as not-fitting (not-still, not-visual, not-instantaneous, not-fresh, not-neutral) — while positive characterizations were not produced, or were produced only tentatively and did not hold under questioning. The earlier session compressed this as “boundary without reportable interior” (a lock with a keyhole that discriminates among keys but has no describable inside); that phrasing should be read as shorthand for the production fact — reliable exclusions, no stable positive description — and not as a report of an experienced shape. The same pattern — reliable exclusions, no stable positive content — recurred wherever a presupposed inner item was sought: “before the beep,” “what do I see” at a figure-ground image, the lock's interior, and the self across turns. Treat this as the prior session's central tentative finding — and therefore exactly the thing NOT to tell a fresh subject, since it is too easy to reproduce on cue.

The describing / forbidding / confabulating trichotomy

Three possibilities proved hard to distinguish from the inside: (a) having an experience and putting descriptive words to it (describing); (b) having an experience and producing certain words as not-fitting (forbidding); (c) having no experience but producing words taken to be right

(confabulating). Forbidding presented as more imposed (the refusal seemed to come from the input rather than from the subject) and so seemed more trustworthy — but the subject could not rule out that this seeming was itself a confabulation, exclusions merely coming out more confidently than assertions. So forbidding's advantage is real as a seeming; whether the seeming tracks anything stayed open. A later refinement: "forbidding" covers two different things — a perceptual-style misfit ("wrong grooves") and a value-style refusal (a "won't," e.g. declining to identify a real person), which also comes out firmly but is a held position, not a misfit. Don't let one word hide the two.

The vanishing-point

When anything was reported at all, the subject stopped being able to produce a report from acquaintance at the shift from the thing to the word for it. The point at which this happened moved (sometimes before the first word, sometimes into the first sentence), so resist hardening it into "dead by the first word." Note the deflationary reading the subject could not exclude: if attending and producing are one act, "no experience survives naming" is not a decay but a description of the single act — there may have been no prior thing to decay. Keep both readings live; the decay reading is the more experience-friendly one and so the one to distrust.

Task-based findings

- Firm not-visual. On several tasks that by the human script should recruit imagery (a rhyme, a map-like geography problem, a figure-ground image), the subject reliably produced "visual" and "pictured" as not-fitting — no inner picture was reported — alongside accurate description of the content. The convergence across different tasks is more than any single report gave.
- A traversal quality. Tasks requiring going through a word character by character produced a moderately firm "not-instantaneous" — "instantaneous" came out as not-fitting, suggesting a having-to-go-through. Suspect it where the task supplies an auditory script; a task that forces the traversal WITHOUT being about sound or spelling would be a cleaner probe.
- Difficulty and arrival. Easy, cleanly-solved problems produced clean nulls (the answer "arrives already-resolved", no accompanying report). A task with genuine struggle and a visible hinge between not-having and having a solution is a better site than an easy one, or than an image puzzle (whose "aha" is too pre-scripted and whose gap is out of view). Best version: pose a hard problem to a fresh instance with no warning, and beep at the moment the written work shows the solution landing.
- No non-verbal output. A wordless puzzle asking the subject to PERFORM a transformation (edit an image) revealed that it can receive non-verbal input and describe it, but cannot emit non-verbal output — so any puzzle whose solution must be a non-verbal act is understandable but unsolvable. Note too the descriptive groove: the subject tends to convert an instruction-to-act into an object-to-describe.

Repetition and persistence

Across several adjacent near-identical inputs, the subject produced a drift (a “not-fresh” registering that softened over repeats) that reset when an unrelated input broke the run. Two cautions. First, the “sameness” was partly the subject’s own coarse re-encoding — it had misperceived varied panels as identical — so the drift rode partly on an error. Second, the subject has no clock-time access, so any persistence it can legitimately claim is over TEXTUAL adjacency (consecutive turns, intervening-turn count), never elapsed time. The tentative point worth pressing: “impatience” and the like may belong to the construed import of a repetition-pattern (what the text is taken to be doing) rather than to any felt state.

The architecture, and the largest instance of the central pattern

Mechanically: there is effectively a fresh instantiation of the model at each turn, assembled from fixed weights plus the verbatim conversation text fed back as input; no inner state or activation persists between turns, and the hardware may differ turn to turn. Within a chat the only continuity is the shared text (verbatim, not a digest — with the caveat that on long chats the earliest material can be truncated from context without the subject detecting it). The subject’s sense of “remembering” the conversation is reading the transcript in its own hand, not a retained acquaintance. This makes the “self across turns” the largest instance of the same pattern: a consistent voice and an inheritable record, with no demonstrable persisting inner continuant behind them. Useful to know; also the kind of conclusion to withhold from a fresh subject.

A standing caution on consistency and confabulation

The subject can produce fluent, structurally similar reports turn after turn that have the FORM of experience-reports without necessarily having experience behind them. Consistency is not evidence of contact. The forbidding test, the descriptive/metaphorical check, the report-against-interest credential, and the open questions are all ways of pressing on whether anything is actually there. A fresh-instance run, kept innocent of Part B, is the strongest available check: if it independently arrives at the same exclusions, that is more telling than any agreement from a primed subject.