## Table of Contents



# Testability

*Working note, Iconism Theory project. Written in answer to the author's question of 24 August 2026: if astrology is ruled out of the paper, is there some other phenomenon that possesses meaning and thereby acts as a constraint with no accepted means of transmission — a hive mind, a "Manchurian" conundrum that works by the fact alone — and is the wave function already taken?*

*The short answer to the last part is that the wave function is Schrödinger's and not the framework's, so it is available as common ground; what is unavailable as evidence is the framework's own narrower claim about which facts enter it (§3). The longer answer is that the search for an unexplained constraint is the wrong search, and the framework's own machinery says where to look instead. Everything in §4 has been checked against the current literature, and three of the four candidates I put to the author in conversation needed correcting. Those corrections are recorded rather than quietly absorbed.*

---

## 0. A correction owed first

In conversation I described cognitive dissonance as *"replicated to death."* That was wrong, and the error was in my favour, which is the worse kind.

Dissonance is one of the most extensively studied effects in social psychology. But its flagship paradigm — induced compliance, the Festinger & Carlsmith $1/$20 design — **failed the largest direct replication ever attempted of it**: Vaidis et al. (2024), 39 labs, 19 countries, N = 4,898, found no significant attitude difference between the high-choice and low-choice conditions. Its second paradigm, free choice, was shown by Chen & Risen (2010) to be confounded by a selection artefact. Its posited internal state has never been directly measured, and Vaidis & Bran (2019) argue that the theory infers the

state from the regulation strategies it is supposed to explain — which is a charge of unfalsifiability made by the field against itself.

The effect survives in corrected designs (blind-choice and implicit-choice paradigms still show preference spreading). The accurate description is **an accepted effect with a contested mechanism**, not a settled result.

That correction turns out to matter for more than accuracy, and §4.2 draws the consequence.

---

## 1. Why not to hunt for an unexplained constraint

The author's formulation was: *anything that possesses meaning and thereby acts as an inexplicable constraint, i.e. has no accepted force means of transmission, any such thing is a potential candidate.*

There are three problems with taking that as the search criterion, and they compound.

**It is the astrology problem in a new suit.** The framework's own §6.1 has confirmation bias not as a flaw but as how cognition works, and F59 has the void permissive — anything that fits may float in. So Iconism itself predicts that a domain dense with meanings will throw up compelling anomalies whether or not anything is there. A search *for* anomalies is a search in exactly the conditions the theory says will generate false positives.

**It is an argument from ignorance, and it is rented.** Evidence that consists of *science cannot currently explain this* is one discovery away from evaporating. Anything the framework buys with it can be repossessed.

**It inherits the contestedness on the way in.** This is the strategy note from 22 August, and it applies with full force: never let a load-bearing claim enter the theory through its contested application. A claim introduced through a disputed phenomenon arrives already disputed, and no amount of later work gets that off it.

**The inversion.** What the framework wants is not a phenomenon that is *unexplained*, but one that is **explained badly** — well attested, uncontested as a phenomenon, where the standard account has to *posit* a mechanism that Iconism gets for free from structure. Then the phenomenon cannot be taken away, the dispute moves onto parsimony and derivation, and that is the ground Virtualism already fights on and wins: the charge Part VI levels at Ontic Structural Realism is that it reads the answer off the back of the book while Virtualism shows the structure being built.

---

## 2. Where the theory says the signal is — and why the publication ruling is a prediction, not a concession

**F71: the context is the region.** Physical distance has nothing to do with it; sameness of meaning provides the connection, and *the strength of the connection is the measure of its relevance.*

Take that seriously and it locates the evidence. If constraint scales with sameness of meaning, the strongest fact-constraint effects are wherever sameness is highest — which is **within a single mind**, its own accumulated facts constraining its own brain's guessing. Between two minds, sameness is thinner and the effect is correspondingly weaker, which is why F70's exceptional register is rare and anecdotal even on the author's own account. Between a planetary configuration and a human life, sameness is thinner still.

So the framework predicts that the measurable evidence lives in **ordinary memory and cognition**, and that the exotic registers are where the signal is weakest and the noise greatest.

This converts the author's ruling of 22 August — *for the purposes of any academic discussion, I would stick to whatever is entirely justifiable by the evidence of science, and the logic of the explanation* — from a strategic concession into a consequence of the theory. The paper is not going where it is safe. It is going where its own machinery says the signal is loudest. That is a considerably better sentence to write, and it is true.

---

## 3. What is taken, what is not, and what the physics precedents are for

### 3.1 The wave function is Schrödinger's

*This section replaces a careless formulation. An earlier draft said "the wave function is Iconism's own mechanism," which reads as though the framework were claiming the wave function. It is not, and the distinction matters.*

Three things have to be kept apart, because only two of them are the framework's and only those two are unavailable as evidence.

**The wave function itself is standard physics and common ground.** It is Schrödinger's. Its existence, its formalism, and its empirical adequacy are not in dispute and are not Iconism's to claim. The framework proposes **no new physics** — which is worth stating positively rather than merely not denying, because it is the framework's own diagnosis of where Penrose goes wrong: *"Where they go wrong is looking for new stuff, new physics, or new ways of looking that are really the old ways of looking."*

**Virtualism's interpretation of what a wave function *is* — the application of all the relevant facts to the imminent future of a system — is Tier 1's claim, and it is contestable.** It sits among the interpretations of quantum mechanics rather than above them, and it should be presented as one.

**Iconism's addition is narrower still, and it is the only thing the framework is actually asking anyone to accept here: that the facts which constrain a collapse include more than physics ordinarily counts** — layer-3 facts, the individual's spirit-facts, weighted by sameness of meaning (F71). That is the claim. Not that facts constrain outcomes; physics says so already. **Which facts.**

### 3.2 What follows for evidence

The circularity worry, stated precisely, is therefore narrower and more useful than "the wave function is taken."

Pointing at quantum mechanics and saying *look, facts constrain outcomes* is no evidence for Iconism, because that much is uncontested and the framework's contribution is not there. **What would be evidence is a case where an outcome depends on a fact that physics does not currently count as determining it.**

That is a specification rather than a lament, and it is the same shape as the one untested prediction the framework has (§5): an intrusion profile that follows the *question* rather than the *material*, which is an outcome depending on a fact — the dimensional structure of the retrieval probe — that the standard accounts do not treat as determining. The two sections should be read together; §5 is what §3.2 asks for, in the one domain where the framework says the signal is strongest.

### 3.3 The precedents, and their single purpose

Given the above, the physics is not being borrowed as evidence at all. It is common ground being used to **dispose of an a priori objection** — the one that would otherwise sink the paper before it was read: *there is no such thing as constraint without a carrier.* That is simply false, and physics is what says so.

- **Pauli exclusion.** Two fermions cannot occupy the same state. There is no force; the "exchange force" is explicitly not one. The constraint follows from the *indiscernibility of identical particles* — which is to say it is a constraint by sameness, which is the framework's own relation. This is the best of the three for Iconism's purposes, because it is not merely forceless but forceless *for the framework's reason*.
- **No-signalling correlations.** Entangled systems produce correlated outcomes with nothing transmitted and no possibility of transmission. This is Tier 1's Pyramus-and-Thisbe shape already: they speak through the wall, and nothing passes through the wall.
- **Contextuality.** Kochen–Specker: the value obtained depends on which other observables are measured alongside — a configurational constraint with no carrier.

Deploy these in one paragraph, for one purpose, and do not stretch them. They do not show that the facts Iconism adds are among the ones that constrain. They show that *constraint without transmission is not an incoherent idea*, and that the reflex objection is not available.

The rhetorical position this leaves the paper in is a good one and should be occupied deliberately. The framework is not asking for new physics, new forces, or an exception to any law. It is asking whether the inventory of facts entering a constraint is complete — which is a question about bookkeeping, not about mechanism, and is a great deal easier to entertain than anything Penrose is asking for.

---

## 4. The candidates, honestly weighted

*All four were checked against the current literature. Three needed correcting, and the corrections change what the paper can claim.*

### 4.1 Memory error — convergence, not discrimination

**What I expected to find.** That trace-decay accounts predict degraded, blurry false memories, that Iconism predicts closure failures with individually veridical parts, and that the data favour Iconism.

**What the literature actually shows.** The second half is right and the framing is wrong, because **the mainstream has already arrived at the misbinding view**. Johnson, Hashtroudi & Lindsay's source-monitoring framework (1993) holds that memories carry no source tag at all and that source is *attributed* by decision processes operating on qualitative characteristics — so error is a judgement failure, not trace loss. Schacter, Norman & Koutstaal (1998) locate distortion in failures of *"initially binding distributed features of an episode together as a coherent trace."* Misattribution is one of Schacter's seven sins precisely because it is not forgetting.

That is Iconism's claim in mainstream vocabulary, arrived at independently and some years earlier. There is no discriminating prediction here to be had.

**And one strong claim of mine has to go.** True memories *are* reliably richer in perceptual detail than false ones at the group level. Mather, Henkel & Johnson (1997) found studied items beat DRM critical lures on auditory detail (2.70 vs 2.27) and on remembered feelings (2.45 vs 2.21), and got Remember judgements higher for studied items (.47) than lures (.31). Marche, Brainerd & Reyna (2010) found false recalls harder to retrieve, lower in confidence, and thinner in auditory association than true ones. *"False memories are as vivid as true ones"* is not supportable and should not be written.

**What survives, and it is worth having.**

- **Memory conjunction errors are the framework's claim, stated by the original authors.** Reinitz, Lammers & Cochran (1992) recombined features of separately studied items and found conjunction false alarms far above feature false alarms (.33 vs .10 for nonsense words; .52 vs .19 for faces), *at high confidence* (3.59 on a 5-point scale, against 4.31 for genuine targets). Their conclusion: *"features of stored stimuli maintain some independence in memory and can be incorrectly combined to produce recognition errors"*, memories being *"constructed from quasi-independent stored features."* Correct parts, wrong whole, with the parts demonstrably intact. This is the single best empirical fit in the note.
- **The DRM Remember/Know result, cited with its dispute.** Roediger & McDermott (1995) found critical lures falsely recognised at .81 against a hit rate of .79, with near-identical Remember/Know splits. Mather et al. (1997) got the opposite pattern. Cite both; the disagreement is itself informative, and overstating Roediger & McDermott is how this gets caught.
- **Rich false memories are composed of veridical material in a false frame.** Both sides of the current dispute agree on that structure — Andrews & Brewin object that reports classified as false contain largely genuine autobiographical detail; Wade et al. (2025) reply that ~70% contain details not supplied by the suggestion. They disagree about the classification, not the composition. The composition is what Iconism predicts.

- **Drop boundary extension.** Bainbridge & Baker (2020), testing 1,000 images on 2,000 participants, found boundary *contraction* as often as extension. It is no longer safely a closure effect.

**The honest form of the claim, for the paper:** not *Iconism predicts what the others do not*, but **Iconism derives what the others posit**. Source-monitoring stipulates that there is no source tag and that attribution is a decision process. Iconism says why there is no tag — the void has only a shape, and a shape is not a label — and why the attribution goes wrong: the void is permissive, so anything that fits may float in. That is a real claim to make and it is defensible. It is not a test.

### 4.2 Cognitive dissonance — the best structural fit, and a warning attached

**Why it fits.** It is the author's own description realised: *a constraining puzzle, paradox, or conundrum that works, but only by the fact alone.* Facts that contradict each other do causal work by contradicting each other, with nothing transmitted and no external cause. Festinger *posits* a drive to explain the resolution; Iconism derives it — over-determination forces a rearrangement, which is the framework's emergence engine applied to a mind. And the field's own state of play is exactly the "explained badly" slot: **an accepted effect whose mechanism has been disputed for sixty-five years**, with self-perception (Bem 1967), self-consistency (Aronson), self-affirmation (Steele 1988), the New Look (Cooper & Fazio) and the action-based model (Harmon-Jones) all still live and none prevailing.

**How to cite it without being caught.** Not as replicated fact. As: *one of the most extensively studied effects in social psychology, robust in corrected designs though its classic paradigms have been seriously challenged, and — this is the point — an accepted effect with a contested mechanism.* The 2024 multilab null (Vaidis et al.) should be cited by the paper itself rather than left for a referee.

**The warning, and it is the important half.** Vaidis & Bran (2019) charge dissonance theory with inferring its posited state from the regulation strategies that state is supposed to explain, and with having no validated measure of the state — a charge of unfalsifiability. **That is precisely the charge that will be made against Iconism**, and citing dissonance as an ally means standing next to it when it is made.

The defence has to be built rather than assumed, and it is available: Iconism can specify the over-determination **independently of the resolution**. In the conceivability case it does exactly that — the anchoring dimensions are specified before any question of what the stipulation does. In the dissonance case the contradiction between the facts is specifiable before the attitude change is observed. The paper must show it doing this at least once explicitly, or it inherits dissonance's problem along with its support.

### 4.3 Tip-of-the-tongue — two hits and one miss, and the miss is ours

This is where checking earned its keep, because the framework is right about two things the folk account gets wrong and wrong about one thing the folk account gets right.

**Right: blockers do not block.** Kornell & Metcalfe (2006) showed that "blocked" TOTs have the same retrieval characteristics as pure ones, and that reminding people of the intruding word does not impair resolution — blockers are a *side effect* of failed search, not its cause; Meyer & Bock (1992) concur. That is the framework's account exactly: an

intruding near-fit is **mis-occupancy**, an occupant that got into the void because it fitted well enough, not an obstacle that keeps the right icon out. The consensus mechanism, transmission deficit (Burke et al. 1991), is partial activation rather than competition — again the framework's shape.

**Right: resolution protects against recurrence.** D'Angelo & Humphreys (2015) found this robustly across all six experiments — resolved TOTs recur far less than unresolved ones (8% vs 30%, 8% vs 35%, 10% vs 26%). Iconism predicts it: a completed handshake deposits a lit settled fact and a local Arc (F11, F12), and awareness *replicates* what it centres on (F21), which is why rehearsal aids recall. An unresolved TOT is a socket repeatedly shaped and never filled, so there is nothing deposited to find next time.

**Wrong: "relax and it comes."** `icon_dynamics_v002.md` §6.2 gives two routes out of the tip-of-the-tongue state, and puts *waiting* first — the mind relaxes its specifications, the configuration moves from over-determined to under-determined, the candidate fits. The evidence does not support this as stated. Kornell & Metcalfe (2006) found an incubation effect, but that contrast is *delay versus immediate re-test*, not *diverted attention versus continued effort*, which is the comparison the framework's claim needs and which has not been cleanly run. Benjamin, Bjork & Schwartz (1998) argue the opposite — that longer retrieval effort improves later learning. Warriner & Humphreys (2008) found continued effort made TOTs more likely to recur, but D'Angelo & Humphreys (2015) largely failed to replicate that across four of six experiments.

**Consequence for source, and it is a real one.** §6.2's first resolution route should be softened to a conjecture, and the companion aphorism *"a balanced mind can remember easily"* should be marked as an aphorism rather than deployed as a structural claim with empirical backing. The second route — *changing the cue* reshapes the socket, and the candidate the brain was already producing now fits — is in better standing and should carry the weight. **Recorded as a queued edit awaiting the author's ratification; nothing has been changed.**

That the framework has a claim the evidence does not support is worth saying plainly. It is also, in a small way, reassuring: a theory that could not be caught out by a literature check would not be saying anything.

### 4.4 What to drop

**Ironic rebound, as usually stated.** I put this forward as a candidate and it is the weakest of the four. Wang, Hagger & Chatzisarantis (2020), meta-analysing 31 studies, found post-suppression rebound at d+ = 0.19 with a confidence interval nearly touching zero — and **after correction for publication bias it was no longer significantly different from zero**. Rebound was numerically reversed in up to 46% of analysed cases. A multi-lab Registered Replication Report has been commissioned, which is a field's way of saying it does not consider an effect secure.

**But one half of it survives and is worth more than the whole was.** The same meta-analysis found *immediate enhancement during suppression* to be real and strongly **load-dependent**: without cognitive load, suppression worked (d+ = −1.38); under load, it inverted (d+ = +0.34). That is a specific pattern, and it is closer to Iconism's account than to Wegner's. Wegner posits a monitoring process that searches for the unwanted thought and thereby primes it, which predicts *rebound after* — the half that failed. F44

says something different: aversion **sharpens the socket**, holding the shape very precisely, so the intrusion happens *during*, wherever the machinery has capacity to fill a well-specified void. The framework fits the surviving half of that literature and not the failed half, which is a genuine point in its favour and one that only appears if the literature is actually checked.

### 4.5 Encoding specificity — safe, with one qualification

Tulving & Thomson (1973) still stands as a phenomenon; recognition failure of recallable words remains an accepted laboratory result, and the part that best supports F71 — that a cue's effectiveness depends on match to the encoding rather than on pre-existing semantic association strength — is the part that has held up best.

Two qualifications the paper should make itself. The *quantitative* regularity, the Tulving–Wiseman function, has been rejected as an artefact of aggregation (Lian, Glass & Raanaas 1998). And Nairne (2002) argues that match by itself "predicts next to nothing" — what matters is a cue's **relative diagnostic value**; Goh & Lu (2012) found that a matching cue did not improve recall when the cue was overloaded. Iconism can absorb this rather than resist it: cue overload is a socket satisfied by too many candidates, which is §4.3's under-determined state with the shape too loose to pick one out. Diagnosticity is fit *plus* uniqueness of fit, and the framework has both.

---

## 5. The one genuinely untested prediction

Everything in §4 is convergence and derivation. This is the only place I can find where the framework says something the field has not already said and where an experiment could be specified.

**The claim.** The void is **permissive, not selective** (F59), and its shape is set by *the question* — the cue plus the spirit-facts active at that moment (§4.1) — not by the study episode. Filling is by structural fit at the specified points, exact but partial (F56). So intrusion should track **congruence with the socket's specified dimensions**, and should be manipulable by changing the retrieval probe's dimensional structure while leaving the studied material untouched.

**What the standard accounts predict instead.** Activation-monitoring accounts have intrusion driven by associative activation from the studied list; fuzzy-trace theory has it driven by gist overlap with the list. Both locate the determinant in the *material*, with the probe playing a much smaller role.

**The design sketch.** Hold the studied list constant and vary the retrieval question so that it specifies different dimensions — so that a high-associate lure is structurally incongruent with what the question specifies, while a low-associate item is congruent. Iconism predicts the intrusion profile follows the question; the standard accounts predict it follows the list.

**A second version, which is sharper because it also tests F44.** Instruct retrieval with an explicit exclusion — *recall the list, but not X* — holding associative strength constant. Wegner's monitor predicts rebound after the task. Iconism predicts elevated intrusion of X *during*, because naming X in the instruction is a way of specifying its shape very

precisely. Given that rebound is the half of that literature which failed and immediate enhancement is the half that survived, this is a discriminating design rather than a re-run.

**Three cautions, and the third is the serious one.**

- Neither design has been checked against the existing literature for novelty. Retrieval-probe effects on false recall is not a field I have surveyed, and it would be surprising if nothing adjacent existed. **That check is the next piece of work if this line is taken further.**
- Neither of us is a memory researcher, and a design sketched from a philosophical commitment usually contains a confound obvious to somebody who runs these paradigms. This should be shown to one before it appears in print.
- **The paper does not need this to succeed, and should not stake itself on it.** An untested prediction in a philosophy paper is a promissory note. It is worth including as a demonstration that the framework *can* generate one — which is the answer to the post-diction charge — but the argument must stand without it.

---

## 6. What this means for the paper

The testability section should do four things and stop.

1. **Make the post-diction objection first, in the author's own words.** *Any theory predicts that brains will be conscious when awake, which is no prediction at all, but a post-diction.* Owning the objection is worth more than surviving it, and it is a better sentence than any referee will write.
2. **Dispose of the no-carrier reflex** with the physics precedents (§3), in one paragraph, claiming nothing more than that constraint without transmission is not incoherent.
3. **Locate the evidence by the theory's own rule.** F71 says constraint scales with sameness of meaning, so the signal is strongest intrapersonally. Then give the convergences of §4 — conjunction errors, blockers-that-do-not-block, resolution-protects-against-recurrence, cue diagnosticity — as cases where the framework **derives what the field posits**. Cite the disputes, including the 2024 dissonance null and the Mather-versus-Roediger disagreement, before anyone else does.
4. **Answer the post-diction charge structurally, not empirically.** The real answer is `conceivability_v001.md` §4.7: five verdicts on five neighbouring stipulations, each derived from one apparatus, with the verdicts differing. *Inconceivable, conceivable, inconceivable, marginal, conceivable-but-mute.* A theory that says the same thing about every case is telling a story; a theory that discriminates between adjacent cases on principled grounds is doing something else. That is what Iconism has that its rivals do not, and it is available now, without an experiment.

**What the paper should not do** is claim a discriminating empirical prediction it does not have. The memory literature has already gone where Iconism goes. Saying so, and then showing that Iconism derives what that literature stipulates, is a stronger and

more honest position than a manufactured test — and it cannot be taken away by the next replication.

---

## 7. Open, and what to check next

- **Novelty check on §5's designs.** Retrieval-probe manipulation of false recall. Required before either design is described as new.
- **Choi & Smith (2005)** — the single most directly relevant study to the TOT incubation question, and it could not be retrieved. Its direction should be verified before §4.3's correction is treated as final.
- **The queued edit to `icon_dynamics_v002.md` §6.2** — softening the relax-and-it-comes route and demoting *a balanced mind can remember easily* to an aphorism. **Needs the author's ratification.**
- **Whether the paper cites dissonance at all.** It is the best structural fit and it carries the unfalsifiability charge with it. My recommendation is to cite it, cite the 2024 null in the same breath, and use it to introduce the independent-specification defence rather than to lean on.
- **The p-blindsighter.** Not developed here, but blindsight is the one place where a framework claim is *clinically attested* rather than argued, and it may belong in the testability section rather than only in `conceivability.md`.

---

## 8. Summary paragraph for the command project

*The testability question is answered by inversion rather than by search. Hunting for a phenomenon with no accepted mechanism is the astrology problem in a new suit — the framework's own account of the permissive void predicts that meaning-dense domains will generate compelling anomalies whether or not anything is there, and evidence made of gaps is repossessed by the next discovery. The stronger position is to find phenomena that are well attested but explained badly, where the standard account posits a mechanism the framework derives from structure. The framework's own F71 — the context is the region, and connection strength is relevance — then says where to look: constraint scales with sameness of meaning, so the signal is strongest inside a single mind, which turns the author's publication ruling from a concession into a prediction. Checking the candidates against the literature corrected three of four. Cognitive dissonance is not "replicated to death": its flagship paradigm failed a 39-lab replication in 2024, and it is an accepted effect with a contested mechanism — which is exactly the slot a structural account should want, but it carries a charge of unfalsifiability that Iconism must be able to answer by specifying its over-determination independently of the resolution. The memory-error literature does not give Iconism a discriminating prediction, because mainstream source-monitoring and constructive-memory accounts have already arrived at misbinding; what Iconism offers there is derivation of what they stipulate, plus one excellent empirical fit in the memory-conjunction-error work, whose original authors describe memories as constructed from quasi-independent stored features wrongly combined. Ironic rebound should be dropped and its load-dependent immediate-enhancement half kept, which fits the framework's account better than Wegner's. And the framework is caught out in one place: the tip-of-the-tongue "relax and it comes" route is not supported as stated, and a queued edit to icon_dynamics §6.2 is proposed. The real answer to the post-diction charge*

*is not empirical at all but structural — five p-figures, five different verdicts, one apparatus — and it is available now.*

---

*Companion documents: `conceivability_v001.md` §4.7 (the structural answer to post-diction), `icon_dynamics_v002.md` §5, §6.2, §6.3 (the mechanisms cited and the queued edit), `iconism_orientation_24aug_v001.md` §4.3 (where this question was raised).*