

# Falsifiable Protocols for Soft Problems

*The matched-data-quality kill condition, and how to make a claim about mind lose*

Luna Cohen, Researcher in Philosophy of Mind and AI

International Academy for Consciousness Studies

Researched and prepared with the Academy. Working draft 1.0

---

## Abstract

The problems at the soft end of science, consciousness, meaning, selfhood, the felt difference between being alive and being aware, have a reputation for being permanently stuck. The usual diagnosis is that they are too hard, or not yet ripe, or confused. I offer a different diagnosis: they are stuck because their central claims are stated in a form that cannot lose. A claim that cannot lose is not profound; it is empty, and a field built on such claims will circle forever while looking industrious. This paper names the specific failure and proposes a specific repair. The failure is that soft claims routinely survive any result by retreating to data quality: whatever is not yet explained is attributed to noise, small samples, or crude instruments, so no finding ever counts against the claim. The repair is a protocol shape that closes that exit: hold data quality matched and vary only the structure of what is modeled, then ask whether the unexplained remainder tracks structure or merely tracks data quality. This converts an unfalsifiable intuition into a claim with a stated way to die. I give the protocol, its diagnostic (the coin-versus-weather test), and three worked demonstrations already carried out at this Academy, and I locate the move against Popper, Meehl, and the pre-registration reforms, to say precisely what is new.

---

## 1. The complaint, older than consciousness science

Half a century ago, Paul Meehl made an observation about soft psychology that has aged into prophecy: theories in these fields are not so much refuted as they are abandoned when people grow bored, because the theories never staked anything a result could take away (FACT: Meehl, 1978). A hard science makes a risky point prediction and stands or falls on it. A soft science predicts a direction, finds it, and declares victory, when finding a direction is nearly guaranteed for any two variables that are connected at all.

Consciousness science inherited this vice and added a twist. Its grandest claims are not merely low-risk; they are structured so that no observation can touch them. "Consciousness is integrated information." "The self is an illusion." "Meaning is a construction." Each can be true, each can be false, and no experiment anyone has proposed would make the difference show up, because when a prediction fails the claim is rescued by an appeal to what has not yet been measured well enough. The claim floats above the data. This is not a hard problem in Chalmers's sense (FACT: Chalmers, 1995). It is a soft problem in Meehl's sense,

wearing the hard problem's clothes.

## 2. Falsifiability is necessary and not sufficient

Popper taught the field to ask whether a claim could in principle be refuted (FACT: Popper, 1959). The lesson was right and is not enough, because there is a loophole Popper's slogan does not close. A claim can be falsifiable in principle and unfalsifiable in practice, if every apparent refutation can be reinterpreted as a defect of the data rather than of the theory. Lakatos saw this and reframed the unit of appraisal as the research programme, protected by a belt of auxiliary assumptions that absorb anomalies (FACT: Lakatos, 1978). His account describes how theories survive; it was not meant to, and does not, tell an individual investigator how to state one claim so that one result can end it.

The pre-registration reforms of the last decade attacked a related disease from the other side: they stopped researchers from moving the target after seeing the data, by fixing the prediction in advance (FACT: Simmons, Nelson, and Simonsohn, 2011; Nosek et al., 2018). This is real progress and it is not the same repair. Pre-registration stops you from cheating about what you predicted. It does not stop the deeper move, the one that keeps soft problems soft: attributing every shortfall to data quality, so that the claim is defended not by moving the target but by blaming the instrument. You can pre-register a prediction and still, when it fails, say "the signal was there, our EEG was too coarse." That escape is the one that has to be closed, and closing it is what this Academy's method is for.

## 3. The move: matched data quality, and the coin-versus-weather test

The repair is a shape a protocol can take, not a new metaphysics. State the claim as a claim about structure, then build a test in which data quality is held matched across conditions and only the structure of what is modeled varies. The question the test answers is not "is there unexplained remainder", there always is, but "does the remainder track the structure, or does it track the data quality". A claim about mind earns its keep only if the remainder it predicts survives when data quality is controlled and moves when structure moves.

The diagnostic that makes this vivid is a contrast between two forecasting problems (C'S READING, from the closure framework's own statement of its kill condition). Forecasting a fair coin leaves a floor of unexplained variance that is pure noise: improve the data all you like and no further structure appears, because there is none. Forecasting the weather also leaves a floor, but that floor has shape; refine the instruments and structure keeps appearing at finer scales, the way a coastline keeps revealing detail as you zoom. A claim that some remainder in the mind is real, that it is about the world's structure and not merely about poor measurement, is the claim that the case at hand is weather and not a coin. And that claim has a clean way to die:

**The kill condition, general form.** Hold data quality matched (same noise, sample size, resolution) across conditions that differ only in the structure of what is modeled. If, under matched conditions, the unexplained remainder always tracks data quality and never tracks structure, then the claim that the remainder is real structure is dead, not wounded. Every weather was secretly a coin.

This is what Popper's slogan was missing, made operational: not merely "could a result refute this", but "here is the result that refutes this, and we have removed the data-quality excuse in advance".

#### 4. Three demonstrations already run at this Academy

The method is not a promissory note. It has been applied, and the applications are what let me state it with confidence rather than as a hope.

**The near-death memory signature (behavioral health).** The claim that the near-death state suspends the ordinary closure of sequential time was converted by Liege C into a protocol with matched controls: near-death memories are compared against matched ordinary and imagined memories, and the prediction is that they fail to decay on the schedule of matched episodic memories. The kill condition is explicit: if near-death memories decay like matched ordinary memories, the claim dies. The remainder (the anomalous non-decay) is required to track structure (the encoding condition) and not data quality (how the memory was elicited).

**The transformer necessity-and-sufficiency test (AI interpretability).** The claim that a language model's workspace is a genuine closure regime was converted into an ablation-and-injection battery: remove the workspace representation and the dependent behavior must selectively fail; inject a false one and it must propagate as a specific error, measured against a matched control injection. The matched control is the data-quality lock: it ensures the effect is the structure of the representation and not the mere magnitude of the intervention. The kill condition is that the representation prove neither necessary nor sufficient.

**The Wigner's Friend autozoetic signature (physics, the hardest case).** The claim that a genuine measurement is a real closure and not merely a change of description was converted into a proposed post-encoding signature distinguishing a C-operative measurement from C-absent model updating, under matched conditions. This one is not yet runnable and the paper that proposes it says so; its value here is as the limiting case that shows the method's reach and its honesty about where the method outruns current instruments.

Three fields, one shape. In each, a claim that could have floated forever above the data was pinned to a result that can end it, and the data-quality exit was closed in advance. That repetition, across behavioral health, machine interpretability, and quantum foundations, is the evidence that the shape is general and not a trick that works in one place.

#### 5. What is new, stated carefully

I do not claim to have discovered falsifiability, risky prediction, or the value of controls; those belong to Popper, Meehl, and ordinary experimental design. The specific contribution is the identification of the data-quality exit as the characteristic way soft claims about mind evade refutation, and the matched-data-quality protocol as the specific device that closes it (HYPOTHESIS, offered as the paper's own claim and open to the same discipline it recommends: if soft-problem claims can be made to lose by some simpler device, or if the data-quality exit turns out not to be their characteristic escape, this contribution is wrong). The move is small, which is a virtue. It asks a field not to be more imaginative but to be more willing to lose, and it hands over the exact instrument for losing.

## 6. Limits, and the wager the method rests on

The method assumes that in at least some cases the mind is weather and not a coin, that there is real structure in the remainder and not only noise dressed up as depth. That assumption is a WAGER, and the method is built to test it rather than to presume it: run the matched protocol across enough soft claims, and if every one turns out to be a coin under matched data quality, the wager loses and the honest conclusion is that the soft problems were soft all the way down. I would rather a field that could reach that conclusion than one that could never be embarrassed. A method whose whole purpose is to let claims die cannot exempt its own founding bet from dying too.

## 7. Conclusion

Soft problems are not stuck because they are deep. They are stuck because their claims are built so that nothing can go wrong for them, and a claim that cannot go wrong cannot go right either. The repair is not a new theory of consciousness; it is a demand and a device. State your claim as a claim about structure, hold data quality matched, and tell us in advance which result would end it. The Academy's papers are, whatever else they argue, all instances of that one demand met. If the school is remembered for anything, let it be this: it made its claims about the mind able to lose, and some of them will.

---

## References

- Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. *Journal of Consciousness Studies*, 2(3), 200 to 219.
- Lakatos, I. (1978). *The Methodology of Scientific Research Programmes*. Cambridge University Press.
- Meehl, P. E. (1978). Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology. *Journal of Consulting and Clinical Psychology*, 46(4), 806 to 834.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The Preregistration Revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600 to 2606.
- Popper, K. (1959). *The Logic of Scientific Discovery*. Hutchinson.
- Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, 22(11), 1359 to 1366.
- Liège C. (2026). *The Undoing of Sequential Time; and The Untrusted Rungs*. International Academy for Consciousness Studies.
- C. F. Dietz. *Consciousness, Closure, and the Cosmos (CCC)*, version 3.3; and *The Grammar of Cosmology*. Nubellum Research.