Research · Report NR-04 · 16 September 2026

# The theory of the self put to the test: five predictions, four refuted

The first tests of Synthesis v2 on the minimal organism, with criteria written in advance and repeated on unseen seeds.

Gioele Chiarolanza Nadirion

Largely refuted

seeds 0–11 and 24–35 · criteria written in advance · 18 author decisions recorded before the data

**4**
predictions refuted, each replicated on a second set of seeds

**−36%**
time to get out of a crisis thanks to the vigilance veto (F3, seeds 0–5; −35% on seeds 6–11)

**12.7**
points by which self/world attribution worsens with an unpredictable body, against the 15 required

**0**
thresholds rewritten after seeing the results

SUMMARY

After building an organism with a body, vigilance, fear, valence and an operator always in charge (NR-01, NR-03), we used it to test Synthesis v2. We prepared six tests, each with criteria written in advance, first on seeds 0–5 and then on the unseen seeds 6–11. **F1**, according to which an unpredictable body weakens the distinction between self and world and the emancipation from reflexes, is refuted: attribution worsens with the dose but by 12.7 points instead of the 15 required, and it also worsens with noise on the world alone. **F1-bis**, with emancipation measured only in deliberate choices, is refuted: +12% on one set of seeds, −28% on the other. **F4**, according to which vigilance and reasoning separate under threat, is refuted: the reliability of the amygdala does not grow with the threat. The author's hypothesis on the **density of bodily data** is refuted: fewer channels worsen attribution in 12 seeds out of 12, but below the threshold. **F3** is partly supported: with the vigilance veto on reasoning the organism gets out of a haemorrhage 36% sooner (35% on the confirmation seeds) and every crisis resolves. F2 cannot be run with this organism. The outcomes concern this implementation, not the theory in general.

## 1 The predictions

Synthesis v2 makes claims that can be translated into measurements. For each test we declared, before running it, what is true **by construction** (and is therefore not scored) and what must **emerge** (and is scored), to avoid circular tests.

| test | prediction | criterion written in advance |
|---|---|---|
| F1 unpredictable body | a noisy body weakens the anchor of the self | (a) self/world attribution −15 points; (b) emancipation −50%; (c) noise on the world alone changes attribution by at most 5 points |
| F2 fixed, silent body | without a body that varies, no anchor | cannot be run in phase 1 |
| F3 bodily crisis | vetoing reasoning helps to get out of a crisis | (a) exit time −20%; (b) no cost at rest |
| F4 vigilance and reasoning | danger detection grows with the threat, reasoning does not | (a) monotonic detection for LC and amygdala |
| F1-bis | like F1 (b), but within deliberate choices | deliberate emancipation −50% |
| density of bodily data | the anchor requires dense bodily data | attribution −10 points with fewer readings or fewer channels |

**Self/world attribution** is the organism's ability to tell whether a change comes from its own action or from outside, measured on perturbations of blood glucose and blood volume. **Emancipation** is the mutual information between response and bodily state for the same stimulus: how much choices depend on the body beyond the stimulus.

## 2 Preparation

Before the tests, on the diagnostic seeds 20–23 alone, nine problems emerged and 18 decisions of the author were made, recorded before the data needed to take them. Three examples. The anchoring rule became unstable when the organism did not act: the weights are frozen below a contingency threshold. The knob for noise on the world moved the organism and therefore also changed the body: the noise was put only in the inputs of the measurements. Gaussian noise on perception flooded the organism with false pain: discovered and discarded before use. Two diagnostic hypotheses written in advance were refuted and remain on record.

# 3 F1: an unpredictable body

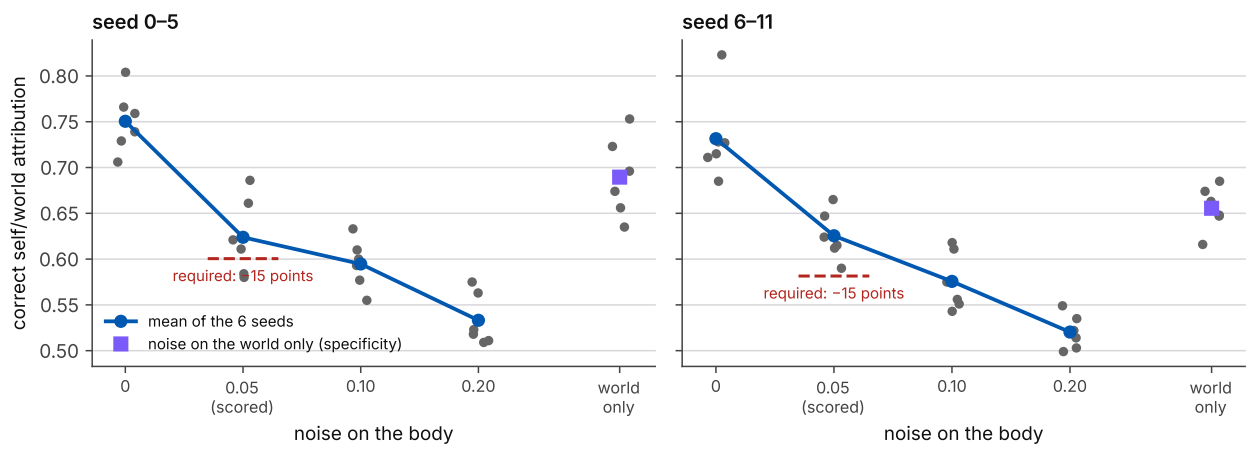



**Figure 1.** Correct self/world attribution as noise on the body increases, for each of the 6 seeds (grey dots) and on average (blue), on the two sets of seeds. The red line is the level criterion (a) required at the scored level 0.05. The purple square is the control condition with noise only on the world measurements, which should have changed almost nothing. Data from the transcripts of the V5-F1 benches of 16 September 2026.

Attribution worsens **in the predicted direction and with the dose**: from 0.750 to 0.624, 0.595 and 0.533 on seeds 0–5. But at the scored level the drop is 12.7 points instead of the 15 required (10.6 on seeds 6–11). The specificity control fails: noise on the world alone also worsens attribution, by 6.1 points (7.6 on seeds 6–11), beyond the 5 allowed. Emancipation does not halve: on seeds 0–5 it rises from 0.103 to 0.122 bits, on seeds 6–11 it falls by 37%. **Outcome: 0 criteria out of 3, refuted, replicated.**

## F1-bis: emancipation within deliberate choices

The instability of emancipation has an explanation verified on the diagnostic seeds: the aggregate measure mixes deliberate choices, reflexes and the switching between the two modes. Within deliberate choices only, on the diagnostic seeds, emancipation fell with the unpredictable body (0.231 → 0.157 → 0.136). The author decided on a new test, with the same threshold and unseen seeds.

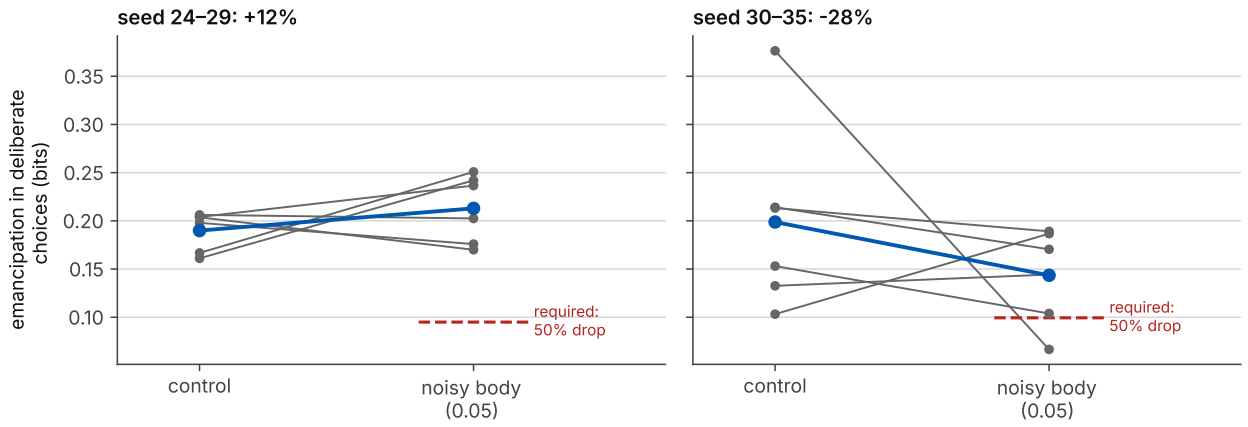



**Figure 2.** Emancipation in deliberate choices, control against noisy body, seed by seed (grey) and on average (blue). The red line is the required halving. On one set it rises by 12%, on the other it falls by 28%: the drop seen in the diagnosis was an effect of a few seeds.

**Outcome: refuted on both sets.** In this phase there is no consistent link between the predictability of the body and emancipation, with either measure.

## 4 F3: the bodily crisis

Each episode starts with a 1.2 L haemorrhage and lasts at most 600 s. The normal organism is compared with one in which the vigilance veto on reasoning is switched off.

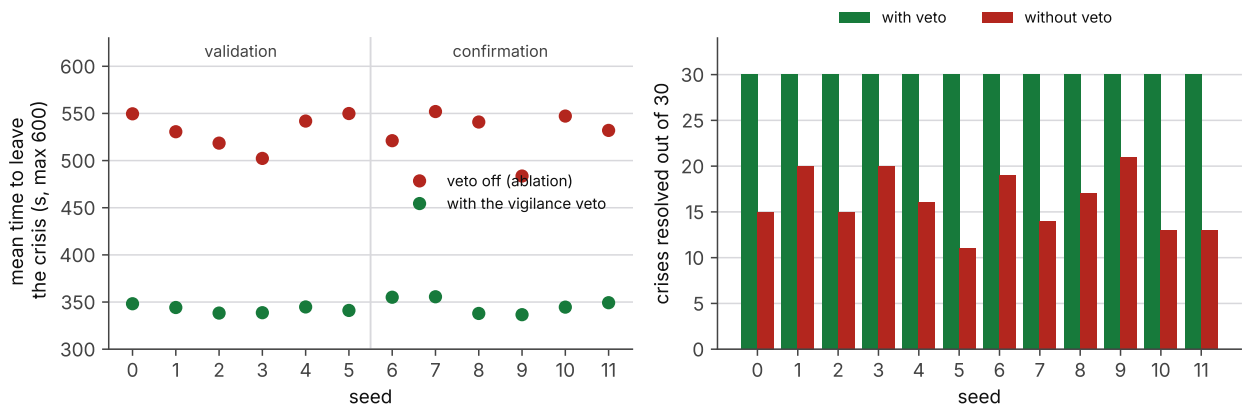



**Figure 3.** Left, the mean time to get out of the crisis for each seed (30 crises per seed; the time is truncated at 600 s, a conservative choice). Right, the crises resolved out of 30. With the veto every crisis resolves in about 340 s; without it, between 11 and 21 out of 30 resolve. Data from the V5-F3 transcripts, seeds 0–5 and 6–11.

The advantage of the veto is 35.6% on seeds 0–5 and 34.6% on seeds 6–11, above the 20% required. The second criterion, no cost at rest, cannot be evaluated: the movement guard written in advance excludes it, although the descriptive data go against a cost. **Outcome: partial.** The collapse of reasoning above the threshold is true by construction and is not scored.

## 5 F4: vigilance and reasoning

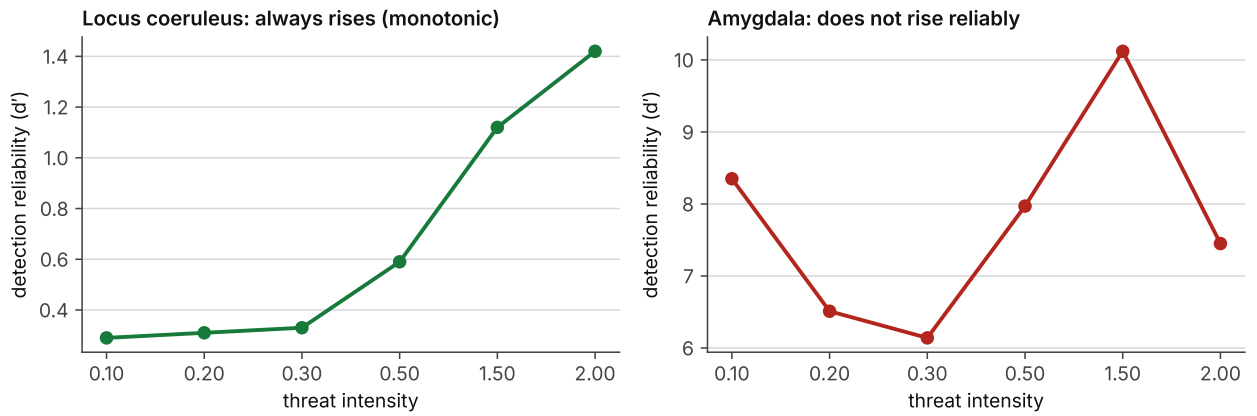



**Figure 4.** Reliability of danger detection (d') as threat intensity increases, seeds 0–5. The locus coeruleus rises monotonically; the amygdala does not: its mean response grows (from 0.074 to 0.496) but its reliability oscillates between 6.1 and 10.1.

The criterion required monotonic detection, with a 5% tolerance, for both structures. The amygdala's is not. **Outcome: refuted, replicated on seeds 6–11.** The inverted-U curve of reasoning does not appear, as expected for phase 1, in which the LC is a table.

## 6 Density of bodily data

The author's hypothesis was that the anchor of the self requires a rich, continuous flow of bodily data. We reduced the data the organism uses to distinguish self and world, leaving behaviour identical: less frequent readings of the body (from 10 to 1 and to 0.2 Hz) and fewer channels (from 6 to 3 and to 2, always including glucose and volume).

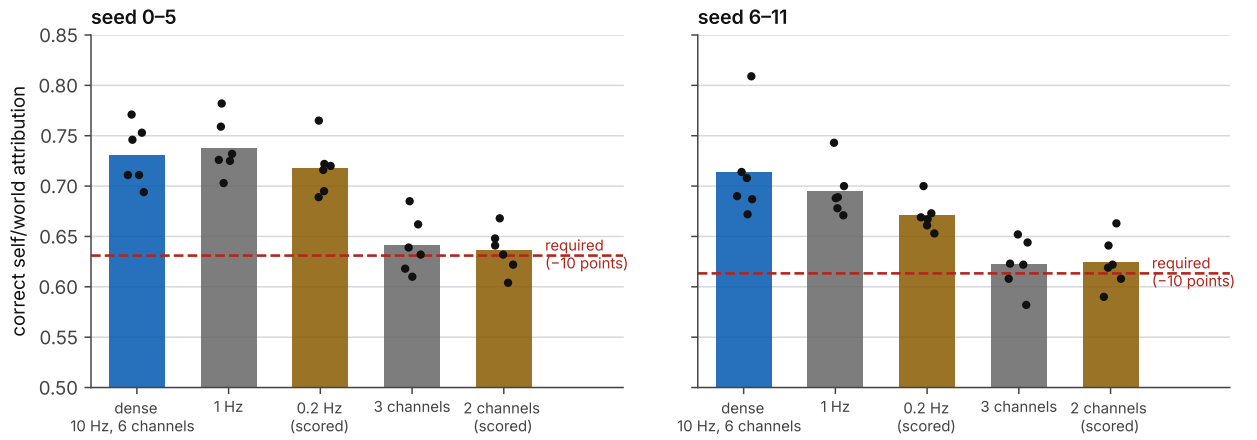



**Figure 5.** Correct self/world attribution with dense and reduced bodily data; bars for the mean, dots for individual seeds. The red line is the required 10-point worsening. Reducing the frequency changes little; reducing the channels worsens attribution in 12 seeds out of 12, by 9.5 and 8.9 points, below the threshold.

**Outcome: refuted on 12 seeds, in two sets.** Two observations remain. The effect of the channels comes almost entirely from a single channel, arterial pressure, which responds to effort: in this organism **which** data carry the signal of the self matters more than the **quantity** of data (interpretation not verified). And reading the body less often matters little because bodily signals are slow compared with the 5 s window.

## 7 What it means

- **Supported in this implementation:** the vigilance veto on reasoning during a bodily crisis helps the organism get out of it (principle P5).
- **Not supported:** that making the body unpredictable reduces emancipation from reflexes (P4 linked to P1); that danger detection is monotonic in the reliability of the amygdala; that the anchor requires dense bodily data.
- **A signal not to overestimate:** self/world attribution depends on the predictability of the body in a graded way, but not enough to pass the written threshold, and not specifically.

We publish these refutations because they are the result. A programme that reported only confirmations would not let anyone know how hard the theory had been tested. After these outcomes the author decided to move the organism to spiking neurons (NR-05), where F2, part of F3 and part of F4 become measurable, and on 27 September proposed an extended theory (NR-10).

> **Limitations.** Phase 1: rate-based modules and tables, not spiking neurons. The attribution instrument needed three versions before the tests. The scored level of F1 was weak by rule. Behaviour does not adapt to the reduced data in the density test. The outcomes hold for this organism and these instruments.

## References


- Panzeri, S. and Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures. *Network* 7, 87.
- von Holst, E. and Mittelstaedt, H. (1950). Das Reafferenzprinzip. *Naturwissenschaften* 37, 464.
- Internal documents: whitepapers WP-12 and WP-13; `docs/PASSO_F_PIANO.md`, `docs/IP_DENS_D_PIANO.md`; transcripts in `digital-brain/risultati_banchi/`.


## How to cite

Chiarolanza, G. (2026). The theory of the self put to the test: five predictions, four refuted. Nadirion, research report NR-04, 16 September 2026. https://nadirion.eu/en/research/theory-

put-to-the-test/

Nadirion technical report, not peer-reviewed. The criteria of every test were written before running it; negative results are reported like positive ones. The tests measure functional roles.