Evidentiary Decoupling

Signal-Outcome Alignment in Neurodevelopmental Regulation and Capability Assessment

Flyxion
Independent Researcher

September 2026

Abstract

A 2026 management article warns that generative AI is producing a "capability mirage": polished work output that no longer reliably indicates the underlying skill of the person who produced it. A separate 2026 neurodevelopmental paper argues that spoken words such as "no" regulate behavior not through meaning but through a learned, precisely timed mapping between acoustic signal and consequence, a mapping that can fail under inconsistent caregiving or competing signals even while comprehension remains intact. Read side by side, these two literatures describe the same formal event in different substrates. Both concern a signal that keeps its surface form while losing its informational grip on the state it once reliably indexed. This essay makes that structural identity explicit, shows that the developmental account supplies a mechanism, precision-weighting collapse under misaligned predictive input, that the organizational account only describes statistically, and proposes evidentiary decoupling as a domain-general principle: a signal's regulatory or evidentiary force is a property of a currently maintained inverse mapping, not of the signal itself, and that mapping can fail silently, leaving the signal intact and functionally hollow.

1 Two Unrelated Papers

Two texts, read independently, have nothing to do with each other. The first is an organizational-psychology article warning that AI-assisted work output is starting to function as a facade: teams look capable, and the underlying skill that used to be legible through their output has become uncertain, possibly eroding, possibly intact, possibly never having existed (Swift, Andreu, & Hernandez, 2026). The second is a neurodevelopmental paper arguing that language regulates behavior through sound, timing, and predictive alignment rather than through meaning, using the acquisition of "yes" and "no" as a case study in how a spoken signal becomes, and can stop being, an effective trigger for a specific neural response (Kamecka, 2026).

Nothing about subject matter connects a discussion of AI-augmented legal memoranda to a discussion of a toddler learning that "no" means stop. But both papers, independently and without shared vocabulary, arrive at the same underlying claim: a signal can remain perfectly well formed while the relationship between that signal and the state it is supposed to indicate quietly comes apart. This essay treats that convergence as evidence of a real, substrate-independent phenomenon rather than a coincidence of two people writing about regulation in the same year, and argues that the developmental paper, read carefully, supplies a mechanism for the organizational paper's central claim that the organizational paper does not itself possess.

2 The Shared Formal Structure

Strip both arguments to their skeleton. In each case there is an observable signal, call it S, and a latent state that the signal has historically been used to infer, call it G. Historically, an observer, an organization, a caregiver, a nervous system, learns some usable inverse mapping $S \rightsquigarrow G$: work output indicates capability, a spoken "no" indicates that stopping is the expected consequence of continuing.

The forward relationship, however, is never $G \to S$ alone. It always runs through an intervening process P that shapes how G becomes S:

$$S = f(G, P).$$

In the organizational case, P is AI assistance: prompting, drafting, editing, and repair capacity that a person brings to bear on a task. In the developmental case, P is the surrounding regulatory environment: the consistency of tone, timing, and consequence across caregivers, institutions, and competing sources of instruction. When P is stable and well aligned with G, the historical inverse mapping $S \rightsquigarrow G$ continues to work. When P changes, becomes more powerful, more inconsistent, or more heterogeneous across instances, the forward mapping f changes while the observer's inverse mapping does not update to match it. The result in both cases is the same:

$$S \text{ unchanged or improved}, \qquad G \text{ unchanged}, \qquad I(G; S) \text{ declines}.$$

Mutual information between signal and latent state falls even though neither term on its own has degraded. This is the formal content of both papers, and it is worth stating as a general principle before returning to the substrate that gives it a mechanism:

> A signal's regulatory or evidentiary force is not a property of the signal itself. It is a property of a currently maintained mapping between the signal and the state it indicates. That mapping can fail silently, leaving the signal intact and functionally hollow.

2.1 Why the surface cannot tell you

Neither paper locates the failure in the signal's legibility. Swift, Andreu, and Hernandez are explicit that AI-assisted output can be excellent on its own terms and still be uninformative about who produced it and how resilient it is under pressure (Swift, Andreu, & Hernandez, 2026). Kamecka is equally explicit that the regulatory failure of a word like "no" is not a failure of semantic comprehension, the child still understands the word, but a disruption of the acoustic-temporal-affective alignment that had previously made the word predictive of a consequence (Kamecka, 2026). In both accounts, asking harder questions about the surface artifact, is the prose fluent, is the word pronounced clearly, will not recover the missing information, because the missing information was never encoded in the surface artifact to begin with. It was encoded in a relationship between the artifact and something else, and that relationship is exactly what changed.

3 The Mechanism the Organizational Literature Lacks

The capability-mirage literature describes evidentiary decoupling but does not explain why it happens inside a person's head or an organization's evaluative practice. It gestures at dry rot, Potemkin villages, the man behind the curtain, metaphors for concealment rather than mechanisms for how a previously reliable inference stops being reliable. Kamecka's paper, read as a mechanism rather than a topic in its own right, fills exactly this gap.

Her argument rests on predictive processing: the nervous system continuously generates expectations about incoming input and updates those expectations when prediction errors occur, with the precision, or confidence weighting, assigned to a given signal determined by how reliably that signal has predicted outcomes in the past (Kamecka, 2026; Friston, 2010; Clark, 2013). A word like "no" acquires high precision through repeated, consistent pairing of acoustic form, timing, affective tone, and consequence. That precision is not a fixed property of the phoneme. It is a running estimate, continuously revised, of how much the signal is currently worth trusting as a predictor. When the pairing becomes inconsistent, competing caregivers deliver contradictory consequences, timing becomes unpredictable, tone becomes decoupled from outcome, the precision weighting assigned to the signal falls. The word keeps its acoustic form. Its regulatory grip on behavior does not.

This is a mechanistic account of exactly the statistical claim the organizational literature makes without explaining it. Mutual information between signal and latent state is not an abstract quantity that declines for no reason; in a predictive system, it declines because the precision assigned to the inverse mapping falls as the forward process generating the signal becomes less consistent, less legible, or more heterogeneous in its causes. Applied to the organizational case, a manager evaluating AI-assisted work is, structurally, running the same kind of precision-weighted inference a nervous system runs on an auditory stop signal: is this artifact, right now, a reliable predictor of what this person can do next. As the population of possible causes behind any given artifact widens, unaided competence, AI-assisted competence, AI-assisted incompetence papered over, the precision an evaluator can rationally assign to output-as-evidence falls, whether or not the evaluator notices that it has fallen. Confidence in the inference and the actual reliability of the inference come apart, which is precisely the mirage the article is worried about, and precisely the condition Kamecka describes when a caregiver keeps using "no" as though it still carries the precision it carried before the child's environment became inconsistent.

4 Regulation Versus Control as the Two Faces of Decoupling

Kamecka draws a distinction late in her paper between regulation and control. Regulation aligns prediction, affect, and consequence, and becomes internalized as a predictive model the person carries forward. Control relies on external enforcement without building that internal model, and can produce short-term compliance that looks identical to genuine regulation from the outside (Kamecka, 2026). This distinction is the developmental version of the capability-mirage problem stated at the level of the person rather than the artifact. A child who complies with "no" because an internal predictive model has been built, and a child who complies with "no" because of proximate external pressure with no internal model behind it, can produce observably identical compliance, S, while differing entirely in G, the underlying regulatory capacity that will or will not generalize to a context where the external enforcement is absent.

This is structurally the same distinction the capability-mirage article draws among four possibilities behind an excellent AI-assisted work product: genuine understanding used AI as an accelerator, sufficient understanding to detect and repair errors, minimal understanding that happened to land on a good result, and minimal understanding producing a result that merely looks good (Swift, Andreu, & Hernandez, 2026). Control, in Kamecka's sense, is the developmental analog of the last two cases. Regulation is the analog of the first two. In neither domain does the surface event, compliance, or a finished artifact, distinguish which case obtains. The distinguishing evidence, in both domains, requires a perturbation: remove the enforcing adult and see whether the behavior persists, remove the AI tool, change an assumption, or introduce contradictory evidence and see whether the underlying model repairs itself or collapses.

5 Implications

5.1 Perturbation as the only recovery procedure

If evidentiary decoupling is real in both domains, then the recovery procedure is the same in both domains, and it is not more careful inspection of the artifact. It is controlled perturbation of the situation that produced the artifact. An organization that wants to know whether a piece of AI-assisted work reflects genuine capability has to change an assumption, remove the tool, or ask the person to repair an introduced failure, precisely the operations that reveal a child's internalized regulatory model by testing whether "no" still functions when the adult who usually enforces it steps out of the room. Both domains reach for the same diagnostic move because both are diagnosing the same underlying object: a currently maintained predictive mapping, not a static competence, and mappings can only be tested by perturbing the conditions that are currently sustaining them.

5.2 Decoupling does not require deception

Neither domain requires bad faith. A polished legal memorandum produced with heavy AI assistance is not a lie; it may be entirely correct. A child's compliant response to "no" under external enforcement is not a performance; the child is genuinely stopping. The mirage metaphor in the organizational literature slightly misleads for this reason, because a facade implies that the surface itself conceals a deficiency. What decoupling describes instead is an inference that has stopped being warranted, not a surface that has stopped being real. The artifact and the compliance can both be perfectly sound. What has failed is a rule that used to reliably connect them to something else, and that rule fails silently, without announcing itself, precisely because nothing about the surface event changes when it does.

5.3 Environments that widen P widen decoupling

Both papers converge on a further point. Decoupling does not require the intervening process P to worsen; it only requires P to become more variable across instances of a signal that used to have a narrow, well-understood range of causes. Kamecka links weakened regulatory clarity in adolescence and in contemporary childhood to environments with more competing signals, more inconsistent consequence, and more sources of instruction operating in parallel (Kamecka, 2026). The organizational literature links weakened evidentiary clarity to a widening range of possible AI-mediated production processes behind any given artifact (Swift, Andreu, & Hernandez, 2026). In both cases the culprit is not decline. It is heterogeneity: a signal that could previously have arisen from only a narrow band of underlying causes can now arise from a much wider one, and the inverse mapping an observer relies on was calibrated to the narrow band.

6 Conclusion

Two 2026 papers, one on organizational trust in the age of AI-assisted work, one on the neurodevelopmental mechanics of "yes" and "no," describe the same event using no shared terminology. A signal keeps its form. A latent state it used to index remains unchanged or even improves. The inference connecting the two, built up through repetition and consistency, quietly stops working, and nothing about inspecting the signal more closely will recover it. The developmental paper is not merely an analogy for the organizational one. Read as mechanism rather than topic, it supplies the missing account of how mutual information between a signal and a latent state actually falls: through precision-weighting collapse in a predictive system whose calibration no longer matches the

process currently generating the signal. Performance, whether a child's compliance or a finished work product, can survive the destruction of its evidentiary value, and the surest sign that this has happened is not visible in the performance itself. It is visible only in what happens when the performance is perturbed.

A word that still sounds like a command is not evidence that it still functions as one.

References

1. Kamecka, M. (2026). In the Beginning Was the Signal: Language as a Neurodevelopmental Regulatory System.
2. Swift, M., Andreu, T., & Hernandez, D. (2026). How AI Creates a Capability Mirage. MIT Sloan Management Review, September 14.
3. Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews Neuroscience, 11(2), 127–138.
4. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behavioral and Brain Sciences, 36(3), 181–204.