Modular Spiral Cognition: A Systems Framework for Bias and Cognitive Coherence

By Tyler Price

Full archive DOI: https://doi.org/10.5281/zenodo.15226447

Table of Contents

Abstract

Modular Spiral Cognition (MSC) is a systems-level framework for understanding how cognitive coherence, emotional salience, and executive oversight interact to shape human thought. It introduces three interacting subsystems—**Observer**, **Interpreter**, and **Reactor**—to explain how stable belief patterns form, persist, and shift. This paper focuses on *bias* as a test case for the MSC model, proposing that bias is not a malfunction in reasoning but a natural output of localized subsystem alignment.

In contrast to traditional models that pathologize bias as irrational error, MSC suggests that coherence-seeking behavior within the Interpreter, amplified by emotional salience from the Reactor and unchecked by the Observer, produces predictable distortions in judgment. These spirals of belief—reinforced through feedback loops—create stable but sometimes maladaptive cognitive attractors.

We demonstrate how common biases such as confirmation bias, consistency bias, and motivated reasoning can be reframed as logical artifacts of subsystem misalignment. The model is supported by decades of empirical findings in cognitive psychology and neuroscience, including research on narrative rationalization, affective filtering, and metacognitive suppression. Additionally, we propose conditions under which bias can be reduced by reintroducing Observer function and interrupting spiral dominance.

This paper lays the foundation for future applications of MSC to a broad range of psychological phenomena—including, but not limited to, depression, ADHD, and suicidal ideation—framing them not as isolated disorders, but as emergent patterns of subsystem misalignment within a unified cognitive architecture.

Introduction

What causes human beings to hold onto false beliefs even in the face of overwhelming evidence? Why do emotionally charged ideas become more resistant to change, while emotionally neutral ones may shift more easily? Why do some interventions aimed at reducing bias succeed for some individuals while causing entrenchment in others?

These questions have been explored extensively within cognitive science, often through models that focus on heuristics, emotional regulation, or memory structures. Yet many of these models operate at a descriptive or functional level—they tell us *what* happens when people reason poorly, but not *how* or *why* those patterns emerge, persist, and self-reinforce. In particular, the role of internal coherence—the felt sense that one's thoughts, emotions, and experiences "make sense"—remains under-theorized in terms of its architecture and dynamics.

This paper introduces **Modular Spiral Cognition (MSC)**, a process-level framework that explains cognition as the output of three interdependent subsystems:

- **The Observer**, which governs metacognition, detachment, and system-level monitoring
- **The Interpreter**, which generates internal narratives to make meaning from experience
- **The Reactor**, which encodes emotional salience and motivational weight

Together, these subsystems form a dynamic control structure through which beliefs stabilize or shift. MSC proposes that bias is not a random error or cognitive defect, but a predictable result of **local coherence-seeking in the absence of full-system alignment**. In other words, when the Interpreter creates a story that fits the data it has—especially when influenced by emotional signals from the Reactor—and when the Observer is offline or suppressed, biased conclusions arise naturally as logical outputs of an imbalanced system.

This view departs from traditional framing in an important way: rather than treating bias as something that "goes wrong" in a rational system, MSC treats it as an emergent property of subsystem misalignment. What appears irrational from the outside may in fact be internally coherent within the current configuration of the mind.

In this paper, we use bias as the first test case for the MSC framework. We will explore how common biases—such as confirmation bias, motivated reasoning, and the need for consistency—can be understood as artifacts of reinforced coherence cycles between Interpreter and Reactor, especially in the absence of active Observer function. We will then align these ideas with well-established empirical research and propose experimental paths for testing MSC-based predictions.

This work lays the foundation for broader applications, including the reinterpretation of clinical conditions such as depression, ADHD, and suicidal ideation—not as isolated disorders but as

patterns of emergent misalignment within the same cognitive architecture. While those applications are outside the scope of this paper, they will be briefly introduced as future directions. **This paper represents the first in a planned series applying MSC to specific cognitive and emotional phenomena, beginning with bias** as an initial test case.

Section I: The MSC Architecture

1.1 Subsystem Overview

Modular Spiral Cognition (MSC) models thought as the emergent behavior of three interdependent subsystems: the **Observer**, the **Interpreter**, and the **Reactor**. Each subsystem contributes a distinct cognitive function. These functions are not tied to single brain regions but emerge from broader functional networks, aligning with current neuroscience on distributed cognition and modularity.

- **The Observer** does not construct narrative content directly; instead, it contributes metacognitive labeling, detachment, and state regulation. It can initiate mental redirection, label internal events, and interrupt interpretive spirals—but it does not create causal explanations or coherent stories.

- **The Interpreter** constructs meaning. It turns perception, memory, and internal state into structured narrative logic. This subsystem is responsible for causal reasoning, justification, planning, and the coherence of one's mental model.

- **The Reactor** assigns salience. It flags emotionally charged data, introduces urgency or avoidance responses, and biases attention toward stimuli deemed significant. It is essential for value judgment and prioritization.

While these three processes are implicit in many cognitive theories (e.g., dual-process models, predictive processing), MSC differentiates itself by explicitly modeling **their interaction** as a *governance system*, not a hierarchy. No single subsystem is "in charge" in every context. Instead, **cognition emerges from the *alignment, dominance, or suppression* of subsystems over time.**

1.2 Subsystem Interactions

Under ideal conditions, the Observer, Interpreter, and Reactor operate in dynamic equilibrium. This balance allows for beliefs and behaviors that are logically coherent, emotionally grounded, and adaptively monitored. In such a state, people exhibit flexibility, resilience, and responsiveness.

However, when one subsystem dominates or is silenced, the system drifts. For example:

- **Reactor dominance** without Observer input may lead to impulsivity, emotional reactivity, or catastrophizing.
- **Interpreter dominance** with a silenced Reactor may produce abstract, disconnected reasoning—hyper-rationality with no motivational force.
- **Observer suppression** disrupts metacognition and self-correction, allowing spirals to persist unchecked even when inaccurate or harmful.

Importantly, subsystem suppression is not inherently pathological. Temporary dominance is necessary in different contexts (e.g., emotional engagement in relationships, analytical focus in problem-solving). The problem arises when dominance becomes chronic, rigid, or unopposed.

1.3 Reinforced Coherence Cycles

Reinforced Coherence Cycles describe how subsystem misalignment leads to the stabilization of distorted beliefs. Reinforced coherence cycles occur when the Interpreter repeatedly reshapes new data to fit an emotionally charged narrative (Reactor), in the absence of moderating input from the Observer.

This creates a **feedback loop**:

1. Reactor flags emotionally salient stimuli (e.g., threat, shame, hope).
2. Interpreter constructs a story to explain or justify the salience.
3. That story feeds the Reactor, which flags the belief as more important.
4. Observer, if offline, fails to interrupt or offer alternative framing.

Over time, this cycle reinforces itself and becomes increasingly resistant to external correction. The system achieves **internal coherence**, even when externally incoherent or maladaptive. Beliefs solidify not because they are true, but because they are internally stable and emotionally reinforced.

MSC identifies this process as the underlying architecture of *many cognitive distortions*, including bias, rigidity, and even some clinical thought disorders. Where traditional models view coherence as a marker of cognitive health, MSC highlights the distinction between **alignment** and **accuracy**:

---

*A belief may be stable, coherent, and motivated—**and still be completely false.***

Section II: Cognitive Bias Through the Lens of MSC

2.1 Traditional View of Bias

In most psychological literature, **bias** is framed as a deviation from rational thought—an error in judgment caused by emotional interference, heuristics, or insufficient information processing. Dozens of well-documented cognitive biases have been cataloged, including:

- **Confirmation bias** – favoring information that supports existing beliefs
- **Consistency bias** – reconstructing memory to align with current beliefs
- **Motivated reasoning** – evaluating arguments more favorably when they support preferred conclusions
- **Availability heuristic** – overestimating the importance of information that is more salient or recent

The prevailing interpretation is that these distortions emerge from cognitive shortcuts or limitations, such as system overload, emotional overwhelm, or a preference for low-effort reasoning. While useful descriptively, this framing does not explain *why* the brain selects certain shortcuts or *how* those distortions stabilize over time into persistent belief systems.

2.2 Reinterpretation: Bias as Logical Output

Modular Spiral Cognition reframes bias not as irrationality, but as a **logical output of reinforced coherence cycles**—interpretive feedback loops between the Interpreter and Reactor in the absence of Observer mediation.

Under this lens:

- The **Interpreter** strives to construct a coherent explanation for subjective experience.
- The **Reactor** marks emotionally salient inputs, shaping what feels "important."
- Without the **Observer** to interrupt or reframe the cycle, the Interpreter adapts its narrative to emotionally charged input, reinforcing a coherent belief that feels "true" precisely because it is **stable**, **internally aligned**, and **emotionally reinforced**.

In this state, what appears irrational externally is **internally rational** given the system's configuration. **Bias is no longer an "error" but an expected outcome of a subsystem that values emotional coherence over objective truth.**

This reinterpretation offers several testable predictions:

- It **explains** why some biases remain robust even after contradictory evidence is introduced.
- It **accounts** for why individuals can acknowledge a logical counterpoint and still *feel* their original belief is more "true."
- It **provides** a process-based mechanism that explains how distorted beliefs not only arise, but **persist and escalate** without intervention.

In other words, **bias is not a glitch in the system. It *is* the system**—under specific alignment conditions.

2.3 Alignment with Empirical Literature

The MSC framing is consistent with extensive research in cognitive psychology and neuroscience. Several key findings align closely with the subsystem interactions described in the model:

- **Cognitive Dissonance Theory** (Festinger, 1957) demonstrates the discomfort caused by conflicting beliefs, which often resolves not through truth-seeking but by aligning new inputs with preexisting narratives—functionally identical to Interpreter-driven restructuring under Reactor pressure.

- **Narrative Coherence Models** (Bruner, 1990; Green & Brock, 2000) show that humans tend to organize information into stories that "feel right," especially when they evoke strong emotion—an Interpreter-Reactor mechanism.

- **Emotion and Salience Research** (Pessoa, 2008; Dolan, 2002) confirms that emotional stimuli shape attention and memory encoding, guiding the Interpreter's narrative by weighting information. This supports Reactor priming and reinforced coherence cycles.

- **Metacognitive Suppression and Frontal Network Inhibition** (McCurdy et al., 2013; Fleming et al., 2012) suggest that reduced metacognitive capacity correlates with rigidity in belief, further supporting MSC's role for the Observer as a moderator of belief reinforcement.

Taken together, these findings substantiate MSC's claim that **bias results from subsystem-level coherence seeking—not from irrationality, laziness, or cognitive failure** per se.

---

*If beliefs can remain stable without being accurate, **what does that mean for how we define "truth"?***

Section III: Realignment and Cognitive Flexibility

3.1 Role of the Observer in Bias Disruption

If bias results from reinforced coherence cycles between the Interpreter and Reactor, then disrupting those cycles requires reintroducing a moderating force: the **Observer**.

In MSC, the Observer serves as a **metacognitive checkpoint**. It tracks internal consistency at a systemic level rather than within local narratives. While the Interpreter seeks coherence and the Reactor flags emotional importance, the Observer alone can:

- Interrupt cognitive spirals by recognizing self-reinforcing thought patterns

- Reframe experience from multiple perspectives
- Weaken the immediacy of emotional salience
- Introduce delay, detachment, or decoupling between impulse and interpretation

This subsystem becomes particularly important during moments of belief challenge, dissonance, or emotional escalation. When active, the Observer can **weaken or dissolve existing coherence loops** by allowing contradictory evidence to be considered outside the frame of the active narrative.

In empirical terms, this aligns with research on:

- **Metacognitive regulation** (Flavell, 1979; McCurdy et al., 2013)
- **Mindfulness and cognitive defusion** (Teasdale et al., 2002; Garland et al., 2015)
- **Prefrontal suppression of affective dominance** (Ochsner & Gross, 2005)

The key insight is not that people "should be more rational," but that **rationality requires specific subsystem conditions to emerge**. Without active Observer function, rationality is unavailable—not because the person lacks intelligence, but because the system is not aligned to permit it.

3.2 Conditions That Restore Observer Activation

MSC suggests that Observer function is **not binary**, but fluctuates depending on environmental, psychological, and neurochemical factors. Several conditions appear to facilitate the reactivation of the Observer and subsequent realignment of subsystems:

- **Emotional de-escalation**: When Reactor intensity decreases, the Observer can more easily engage. This may be induced through mindfulness, regulation strategies, or external support.
- **Narrative disruption**: When the Interpreter's current story is challenged—either gently or through shock—the loop may destabilize, opening space for new alignment.
- **Perspective shifting**: Practices that invoke self-distancing (e.g., "What would I think if someone else believed this?") stimulate Observer engagement.
- **Novelty and contrast**: Exposure to new experiences, data, or perspectives that do not fit the current coherence loop can activate Observer scrutiny.

- **Safety**: If the individual perceives threat or social risk, Reactor dominance may override all other functions. Psychological safety, in contrast, allows for mental experimentation and observation.

These are not universal prescriptions, but **principled conditions** under which subsystem realignment becomes more likely. Rather than viewing resistance to change as irrational stubbornness, MSC reframes it as the *natural inertia of a stable subsystem loop*—and provides a roadmap for how that loop might be interrupted.

---

*If the Observer enables us to step outside reinforced coherence, then **cognitive flexibility is not a personality trait—it's an emergent property of alignment.** The ability to question, pause, or revise a belief may depend less on willpower than on which subsystems are engaged at the right time.*

Is open-mindedness simply the byproduct of a well-aligned system?
*If so, belief change is not merely a philosophical problem. It becomes a systems problem—**and perhaps a solvable one**.*

Section IV: Predictive Utility of MSC

This section explores how MSC's architectural model doesn't just explain existing phenomena like bias—it **can also be used to anticipate cognitive vulnerabilities, forecast the progression of mental states, and guide intervention strategies.** This is the "*So what?*" of the framework—its forward-facing power.

4.1 From Explanation to Anticipation

If subsystem misalignment can be mapped, then psychological states may be more predictable than traditionally assumed. Because MSC offers a mechanistic architecture—rather than a static typology—it can be used to **predict behavioral patterns based on subsystem configuration.**

For example:

- A person showing high emotional reactivity, rigid narratives, and low meta-awareness is likely operating under Reactor–Interpreter dominance, with suppressed Observer function.

- In such a configuration, **information alone** will have minimal impact unless it either bypasses Reactor filtering or restores Observer engagement.

- In contrast, someone whose Observer is active but whose Reactor is disengaged may experience detachment, low motivation, or depressive flattening—even if their beliefs are accurate and coherent.

Rather than viewing cognition as unpredictable or purely situational, MSC suggests that many cognitive tendencies can be **anticipated based on subsystem alignment profiles**—and even *temporally modeled* as system shifts unfold over time.

4.2 Anticipating Resistance and Receptivity

The framework also provides tools for **predicting resistance to change**—not just in clinical contexts, but in education, relationships, and communication.

- **High Interpreter-Reactor coherence with low Observer presence** predicts entrenched belief and **emotional argumentation**.

- **Active Observer with conflicting narratives** predicts **uncertainty and openness** but also **potential instability**.

- **Balanced subsystem activity** predicts **flexibility, curiosity, and integration**.

By evaluating these patterns, MSC allows us to move beyond vague concepts like “open-mindedness” or “emotional regulation” and into **structural predictors of mental readiness**.

This can be applied to:

- **Therapy** – anticipating client blocks before they emerge

- **Instructional design** – framing material for Observer reactivation

- **Conflict resolution** – identifying when communication is likely to succeed or fail based on current subsystem dominance

- **Organizational coaching** – helping teams recognize and overcome collective cognitive entrenchment

---

*If we can map resistance in advance, can we also **design environments that reduce its likelihood?***

4.3 Time-Sensitive Cognitive Trajectories

One of MSC's most promising frontiers is its capacity to **model how subsystem dominance shifts over time**. Unlike snapshot-based typologies, MSC recognizes cognition as a fluid state influenced by environment, memory, attention, and meaning-making processes.

By observing:

- Which subsystem is dominant
- Whether coherence is local or systemic
- How stable the configuration is

…it becomes possible to **forecast likely shifts in behavior, motivation, or belief formation.**

Example use cases:

- Predicting when a person in crisis may shift from reinforced coherence cycle to openness (or vice versa)
- Modeling how belief rigidity might escalate or soften across social media exposure
- Tracking real-time alignment shifts in neurodiverse individuals (e.g., ADHD, dissociative states)
- Simulating potential failure points in high-stakes decision-making contexts (e.g., leadership, negotiation, crisis)

4.4 From Mental States to Mental Motion

Traditional models tend to label people as being in a given **state**: anxious, disengaged, biased, irrational. MSC reframes these as **positions within a moving system**. Mental states are not frozen—they are **dynamic trajectories**, and the system is always trying to stabilize.

> Just as weather forecasting depends on atmospheric pressure gradients, cognitive forecasting depends on subsystem pressures: which subsystem is dominant, which is suppressed, and where the system is trying to stabilize next.

If this proves true, it suggests a future where we no longer respond reactively to dysfunction—but **intervene before the coherence loop stabilizes maladaptive belief.**

MSC does not claim to predict the content of thought. But it may allow us to **forecast the shape of thought**—its emotional velocity, interpretive rigidity, and systemic flexibility.

> *If we can model not just where someone is cognitively—but where their system is trending—then intervention becomes less about correction and more about timing.*
>
> ***The future of cognition may not lie in fixing broken thoughts, but in forecasting fragile ones.***
>
> *If that becomes possible, then mental health itself may evolve—from reactive treatment to proactive system navigation.*

Section V: Future Directions

5.1 Extending MSC Beyond Bias

While this paper focuses on cognitive bias as an initial test case, the MSC framework was designed from the outset to describe a broad range of cognitive, affective, and behavioral patterns. The same subsystem dynamics used to explain bias may also illuminate:

- Depression
- ADHD and executive dysfunction
- Suicidal ideation
- Cognitive dissonance and belief rigidity
- Trauma responses and narrative reformation
- Motivation loss and emotional flattening

These patterns may appear diverse on the surface, but **each may stem from a specific configuration of subsystem dominance, suppression, or misalignment.**

The conditions listed here are illustrative, not exhaustive—**MSC is a general framework that may extend to many additional forms of cognitive and behavioral dysfunction as they are explored**.

5.2 Testable Hypotheses and Experimental Designs

For MSC to move beyond theory, its claims must be evaluated empirically. Several testable predictions follow from the current model, including:

- Observer activation correlates with openness to belief revision, even when emotional intensity is high
- Reinforced coherence cycles between Interpreter and Reactor can be disrupted by introducing metacognitive prompts
- Individuals with reduced Observer engagement will show increased susceptibility to motivated reasoning and narrative entrenchment
- Shifts in subsystem dominance (e.g., via mindfulness, novelty, or safety cues) precede measurable cognitive flexibility in decision-making or judgment tasks

Experimental paradigms may include:

- Neuroimaging studies using fMRI to track default mode network activity during coherence loop disruption
- Behavioral tasks measuring belief revision under affective priming
- Longitudinal studies examining how subsystem balance predicts responsiveness to therapy, education, or conflict

These studies would not only test MSC's validity, but help refine it—identifying boundary conditions, individual differences, and contexts in which subsystem dynamics may be more or less predictive.

What if the difference between belief change and belief entrenchment isn't logic—but timing and subsystem readiness?

5.3 Future Technologies and System Modeling

In the long term, MSC may also guide the development of cognitive technologies aimed at:

- Tracking subsystem engagement in real-time (e.g., neurofeedback, biometric correlates of salience and coherence)
- Supporting intervention timing (e.g., when to introduce cognitive challenges or pauses based on OIR trends)
- Creating adaptive learning environments that modulate difficulty or framing to support subsystem balance
- Informing AI-human interaction design where the goal is not just to align with user preference, but to **restore cognitive flexibility**

These applications remain speculative but feasible. If cognition is a system, then system tools can help us optimize—not just performance, but perspective.

5.4 Reframing the Science of Thought

Cognitive science has long struggled with integrating emotional salience, narrative construction, and metacognitive control into a single explanatory model. MSC offers a scaffolding through which these forces can be modeled not as competing theories, but as **interacting components of one architecture**.

If successful, MSC may help unify several previously disconnected domains:

- Rationality studies
- Emotion and motivation research
- Executive function modeling

- Belief formation and epistemology
- Clinical treatment frameworks for mood and attentional disorders

Just as the structure of DNA offered a new lens for understanding biological processes across disciplines, MSC may offer a common language for understanding psychological ones—not by reducing complexity, but by showing how complex parts **interlock.**

> *If these patterns are emergent, not isolated, then **understanding cognition may no longer be about cataloging its failures**—but decoding its alignments.*
>
> *What we call **disorders may be not deviations from reason—but stable systems** missing one alignment.*
>
> *If we can see them that way, we may **stop treating the mind as broken—and start learning how to rebalance it.***
> .

Appendix A: Illustrative Cognitive Case Studies (Fictional)

Subsystem misalignment in action: Narrative walkthroughs

> *These examples are fictional and presented to clarify how subsystem dynamics may appear in practice. They are not empirical evidence, but narrative tools to support understanding.*

Case Summary: Jonah (Cognitive Rigidity and Self-Narrative Collapse)

Age: 34
Context: Unexpected job loss from a long-term tech position. Presenting as demotivated, withdrawn, and emotionally numb.

Baseline State:

Jonah exhibited balanced cognitive function. His Observer was active, monitoring his thoughts and course-correcting when needed. The Interpreter reinforced a coherent identity (“I’m

valuable, competent"), and the Reactor supplied motivational salience based on recognition and achievement.

Misalignment Trigger:

After being laid off during a corporate restructuring, Jonah struggled to reconcile the event. The Interpreter, lacking input from an emotionally grounded Reactor and without Observer moderation, began distorting internal coherence:

> *"If I was really good at what I did, they wouldn't have let me go."*

The Reactor flagged shame and confusion, intensifying interpretive distortion. The Observer, overwhelmed and unengaged, failed to intervene. A **reinforced coherence cycle** formed between shame (Reactor) and compensatory narrative collapse (Interpreter).

Outcome:

Jonah rejected support from friends, interpreting their reassurances as pity. The more emotionally charged the interaction, the deeper his narrative collapse became. He could "explain" why he was unworthy but not believe any counterpoint. Observer suppression made reframing impossible.

Realignment:

A low-stakes conversation about music introduced novelty and safety, bypassing the Reactor's threat filters. The Interpreter paused. The Observer re-engaged briefly, prompting an alternative framing:

> *"Maybe what I lost was a job—not my value."*

From there, emotional salience shifted, and a broader narrative emerged. Observer function returned, breaking the cycle.

Case Summary: Dani (Observer Dominance and Motivational Flatlining)

Age: 27
Context: Graduate student in neuroscience, highly analytical but struggling with decision-making and long-term goal pursuit. Describes herself as "detached from everything."

Baseline State:

Dani's Observer was highly active. She excelled at stepping back, analyzing herself, and reframing thoughts. The Interpreter was competent and cautious—often generating abstract or speculative narratives rather than firm conclusions. Her Reactor, however, was suppressed.

> *"I can think of fifty reasons to do something, but none of them feel right."*

Misalignment Presentation:

In Dani's system, emotional salience rarely activated. She described emotions as "inconvenient noise" and preferred to reason everything out. The Interpreter tried to generate motivation logically—but without Reactor input, no option "mattered." Even coherent goals felt inert.

> *"I know I should care about this project, but I don't. Not enough to act."*

Dani wasn't depressed in the traditional sense. She maintained sleep, appetite, and a calm demeanor—but described herself as **"mentally idle."**

Intervention:

Therapy targeting emotional reintegration (through somatic experience and journaling) introduced new data to the Reactor. Emotional resonance with childhood memories led to a shift:

> *"I forgot how much I used to care about discovery. I stopped listening to that part of me."*

The Interpreter absorbed this salience. The Observer, now seeing a motivational signal that passed the "test," allowed it to remain in focus. Dani's goals didn't just make sense—they started to matter again.

Takeaway:

Dani's case illustrates that **Observer overactivation** can result in emotional detachment and motivational paralysis—not because logic is flawed, but because **salience is missing from the decision loop**.

Case Summary: Malik (Emotional Coherence Loop and Belief Entrenchment)

Age: 22
Context: College dropout involved in repeated online political arguments. Feels increasingly "at war" with others' ignorance. Describes his beliefs as "the only sane ones left."

Baseline State:

Malik had strong emotional conviction and a quick, adaptive Interpreter. He could make sense of almost any confrontation with sharpness and certainty. However, his Observer was largely absent—and this allowed **his Reactor to dictate salience unchecked.**

> *"If someone doesn't agree with me, it's because they're not awake yet."*

The Interpreter began conforming tightly to Reactor cues—particularly anger and moral disgust. Each confrontation fed more evidence into a closed loop:

> *"See? They don't want the truth."*

> *"They'll twist anything to avoid admitting I'm right."*

Misalignment Cycle:

With every new conflict, Malik's internal coherence grew stronger—but externally, his relationships collapsed. Observer disengagement meant no internal contradiction was allowed to emerge. His entire worldview took on a **coherence-turned-certainty** tone.

Attempts by friends to engage were either dismissed or escalated into further moral condemnation. His narrative wasn't just distorted—it was fortified.

Realignment Attempt:

A respected professor—not by arguing, but by showing calm and curiosity—posed a subtle challenge:

> *"You sound convinced, Malik. How do you know when it's safe to change your mind?"*

The question wasn't met with anger. It created a pause. For the first time in months, Malik couldn't immediately generate an emotionally satisfying answer. The Interpreter wavered. The Reactor, momentarily quiet, allowed space for the Observer to whisper:

> *"Am I thinking, or just repeating?"*

From there, the loop began to loosen.

Takeaway:

Malik's case exemplifies how **Reactor-dominant coherence loops** can spiral into ideological rigidity when Observer function is suppressed. Breaking the cycle requires not confrontation—but subtle destabilization through **safe, curious reframing.**

Note on Governance Styles

While this paper emphasizes subsystem alignment as its central focus, many of the patterns described in these fictional examples can also be viewed through the lens of governance styles (commonly referenced in Spiral Dynamics and other developmental frameworks).
For instance, Dani's detached self-awareness resembles an unbalanced Yellow-mode cognition, while Malik's emotionally rigid certainty mirrors affective dominance often seen in Red-Orange governance states. These perspectives are not contradictory—they are layered. MSC provides a systems-level architecture beneath governance style, clarifying how values and behaviors emerge from alignment dynamics.

References

Bruner, J. (1990). *Acts of meaning*. Harvard University Press.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, 298(5596), 1191–1194. https://doi.org/10.1126/science.1076358

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2012). Relating introspective accuracy to individual differences in brain structure. *Science*, 329(5998), 1541–1543. https://doi.org/10.1126/science.1191883

Garland, E. L., Hanley, A., Farb, N. A. S., & Froeliger, B. (2015). State mindfulness during meditation predicts enhanced cognitive reappraisal. *Mindfulness*, 6, 234–242. https://doi.org/10.1007/s12671-013-0250-6

Green, M. C., & Brock, T. C. (2000). The role of transportation in the persuasiveness of public narratives. *Journal of Personality and Social Psychology*, 79(5), 701–721. https://doi.org/10.1037/0022-3514.79.5.701

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *Journal of Neuroscience*, 33(5), 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences*, 9(5), 242–249. https://doi.org/10.1016/j.tics.2005.03.010

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, 9, 148–158. https://doi.org/10.1038/nrn2317

Teasdale, J. D., Segal, Z. V., Williams, J. M. G., Ridgeway, V. A., Soulsby, J. M., & Lau, M. A. (2002). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive

therapy. *Journal of Consulting and Clinical Psychology*, 68(4), 615–623. https://doi.org/10.1037/0022-006X.68.4.615