The Evolutionary Architecture of Human Consciousness: Cognitive Contradictions, Biological Duality, and the Illusion of Time

Boris Kriger

Institute of Integrative and Interdisciplinary Research

boriskriger@interdisciplinary-institute.org

Abstract

Human beings experience their own minds as inconsistent, conflicted, and often unreliable. Memory distorts, perception contradicts itself, plans for the future fail, and moral reasoning struggles against biological impulses. These features are commonly interpreted as flaws in human cognition. This theoretical review argues the opposite: such contradictions are not defects but may represent necessary consequences of the evolutionary architecture of consciousness. This paper uniquely integrates temporal perception phenomena (memory reconstruction, future projection, present-moment awareness) with dual-process cognitive architecture to propose a unified framework explaining human irrationality. While drawing on established work in bounded rationality and error management theory, the novel contribution lies in showing how temporal construction and biological duality interact to produce meaning-seeking behavior and existential conflict. The speculative nature of this synthesis is acknowledged throughout, alternative explanations are considered, and five testable empirical predictions with suggested methodologies are proposed. This framework has implications for clinical psychology, artificial intelligence alignment, and behavioral economics, particularly regarding evolutionary mismatch in modern environments.

Keywords: consciousness, evolutionary psychology, cognitive science, temporal perception, memory reconstruction, meaning-making, dual-process architecture, bounded rationality, adaptive cognition, predictive processing

1 Introduction: The Problem of Human Inconsistency

Human cognition is riddled with contradictions. People hold incompatible beliefs, misremember events, act against their own interests, and oscillate between rational thought and instinctive impulse. Philosophical traditions have often treated this as a weakness of reason (Kahneman, 2011). Psychology classifies such tendencies as biases and distortions (Tversky and Kahneman, 1974). Yet the persistence and universality of these features suggest that they may not be accidental flaws.

The central thesis of this paper is that human cognitive inconsistency may be structurally adaptive—or at minimum, may have been tolerated by selection because the costs of perfect consistency would exceed its benefits. A perfectly consistent perception of reality might impede rapid action in time-pressured situations, though this claim requires empirical investigation. It should be noted that certain artificial intelligence systems achieve high consistency without apparent paralysis; however, these systems typically operate in constrained domains with well-defined objectives, unlike the open-ended environments that shaped human cognition (Russell, 2019). Evolution may have favored minds that could operate despite ambiguity, contradiction, and incomplete information (Gigerenzer, 2008).

This paper offers a theoretical synthesis rather than original empirical research. Its contribution lies in integrating findings from evolutionary psychology, cognitive neuroscience, and philosophy of mind into a unified framework that specifically links temporal perception phenomena with dual-process cognitive architecture. The speculative nature of this integration is acknowledged throughout, and testable predictions with concrete methodologies are proposed where possible.

1.1 Scope and Limitations

Before proceeding, several important caveats must be stated. First, the adaptationist framework employed here carries known risks. As Gould and Lewontin (1979) famously argued, not every biological feature is an adaptation; some may be byproducts (spandrels), developmental constraints, or products of genetic drift. The features discussed in this paper—cognitive inconsistency, memory reconstruction, temporal illusions—may have arisen through any of these mechanisms. When I suggest adaptive functions, these should be understood as hypotheses requiring empirical testing rather than established facts.

Second, this paper does not claim that all cognitive inconsistencies are beneficial. Many clearly are not, particularly in modern environments that differ substantially from ancestral conditions. The goal is to explore why such features persist despite their apparent irrationality, not to celebrate them.

Third, the term "inconsistency" itself may be misleading. What appears inconsistent from a classical logic perspective may be consistent within a probabilistic or context-sensitive framework. Recent work on resource-rational analysis (Lieder and Griffiths, 2020) suggests that many apparent biases are optimal given computational constraints. This paper's use of "inconsistency" should be understood as referring to deviations from classical rationality norms, not necessarily from deeper computational optimality.

1.2 Falsification Criteria

A theoretical framework is only useful if it can be wrong. The following findings would substantially undermine or falsify this framework:

1. **Independence of phenomena**: If irrationality, temporal distortion, meaning-seeking, and inner conflict could be experimentally dissociated—increased independently without affecting the others—this would disconfirm their proposed integration. The framework predicts correlated variation.

2. **Absence of ecological rationality**: If heuristics that appear irrational in the laboratory also performed poorly in naturalistic environments with ancestral-relevant structure, the "adaptive bias" interpretation would be undermined.

3. **Failure of error asymmetry predictions**: Error management theory predicts biases should be strongest where ancestral error costs were most asymmetric. If biases showed no relationship to cost asymmetry, or were strongest in domains with symmetric costs, the framework would require revision.

4. **Linear executive function benefits**: If executive function showed monotonically positive relationships with real-world outcomes across all domains (contradicting Prediction 3's curvilinear hypothesis), this would challenge the "adaptive inconsistency" thesis.

5. **Cultural uniformity**: If Prediction 4 failed—if holistic and analytic cultures showed identical patterns of contradiction tolerance—the cultural variation hypothesis would be falsified.

The framework would *not* be falsified by: (a) finding that some inconsistencies are maladaptive (evolutionary mismatch predicts this); (b) finding that clinical interventions improve functioning (they target pathological extremes); or (c) finding that AI achieves consistency without paralysis (AI operates in constrained domains). These outcomes are already accommodated by the theory, which is why the criteria above focus on novel predictions.

1.3 Distinguishing Adaptation from Byproduct: Operational Criteria

The mismatch hypothesis (invoked in Section 2.3, 6.2) risks becoming unfalsifiable if any maladaptive outcome can be attributed to mismatch. To constrain this, I propose the following criteria for distinguishing "adaptive in ancestral context" from "never adaptive":

1. **Design evidence**: Does the trait show evidence of functional design—precision, efficiency, complexity that would be unlikely without selection? Random byproducts typically lack such organization.

2. **Cross-species comparison**: Do related species show similar traits in similar ecological contexts? Convergent evolution suggests adaptation; uniqueness suggests possible byproduct.

3. **Fitness-relevant calibration**: Does the trait respond to fitness-relevant variables (threat level, resource availability, mating opportunity) in ways that would have been adaptive ancestrally? Byproducts should show less systematic calibration.

4. **Developmental canalization**: Is the trait reliably developing despite environmental variation? Strong canalization suggests selection; high environmental sensitivity suggests byproduct.

Where a trait fails these criteria, the byproduct interpretation is preferred.

1.4 Constraining Mismatch Explanations

The evolutionary mismatch hypothesis risks unfalsifiability if any maladaptive outcome can be attributed to environmental change. To constrain this explanatory escape hatch, mismatch should only be invoked when:

1. **The ancestral environment can be specified**: We must identify what specific feature of ancestral environments the trait was calibrated for
2. **The modern change can be documented**: The relevant environmental shift must be identifiable (e.g., from small groups to anonymous masses; from immediate feedback to delayed consequences)
3. **The direction of dysfunction is predictable**: Mismatch should predict *which* modern contexts produce dysfunction, not merely explain dysfunction post-hoc
4. **Reverting the environment should restore function**: If possible, recreating ancestral-like conditions should reduce the maladaptive expression

When these criteria cannot be met, the "maladaptive byproduct" interpretation is preferred over "adaptive trait in mismatch." This paper invokes mismatch primarily for social-evaluative biases (ancestral: small group with lasting reputation; modern: anonymous masses with fleeting interactions) and future-planning biases (ancestral: immediate feedback; modern: delayed consequences). These meet the above criteria.

2 Cognitive Contradictions as an Adaptive Mechanism

Organisms that waited for coherent, logically complete models of reality before acting would face significant survival challenges. Action in uncertain environments requires tolerance for inconsistency. The human mind therefore may have evolved not for logical harmony but for functional decisiveness (Todd and Gigerenzer, 2012).

Contradictory beliefs, heuristic shortcuts, and selective attention allow rapid decisions in complex environments (Gigerenzer et al., 1999). Tolerating inconsistencies is not necessarily irrationality; it may be a survival strategy. Research in bounded rationality has demonstrated that cognitive limitations often produce superior outcomes in real-world environments compared to theoretically optimal but computationally intractable approaches (Simon, 1955).

2.1 Systematic Evaluation: Adaptation vs. Byproduct

Following Gould and Lewontin (1979), we must ask: are cognitive contradictions adaptations, byproducts, or constraints? Rather than treating this question generically, I evaluate each major phenomenon against the four criteria proposed in Section 1.3.

Summary: Dual-process conflict and memory reconstruction show strongest evidence for adaptation (meeting 3-4 criteria). Planning fallacy and temporal distortion show moderate evidence. Meaning-seeking is most uncertain, possibly a byproduct of more basic pattern-detection that was itself selected. This differential evaluation is important: not all phenomena discussed have equal evidential support for adaptive function.

Table 1: Systematic Evaluation of Phenomena Against Adaptation Criteria

Phenomenon	Design Evidence	Cross-Species	Fitness Calibration	Canalization
Memory reconstruction	High: systematic, functional organization	Moderate: episodic-like memory in scrub jays, apes	High: varies with self-relevance, emotion	High: universal development
Planning fallacy	Moderate: systematic but possibly emergent	Low: limited evidence in non-humans	Moderate: varies with goal importance	High: cross-cultural presence
Dual-process conflict	High: distinct neural systems	High: similar structures in mammals	High: calibrated to threat/reward	High: universal architecture
Meaning-seeking	Moderate: pattern-detection is organized	Low: uncertain in non-humans	Moderate: increases with uncertainty	Moderate: culturally shaped
Temporal distortion	Moderate: systematic under specific conditions	Moderate: interval timing across species	High: varies with arousal, attention	High: developmental consistency

2.2 Error Management Theory

Haselton and Buss (2000) provide a more rigorous framework for understanding why certain "errors" persist. Error management theory proposes that when the costs of different types of errors are asymmetric, selection favors biases toward the less costly error. For example, the tendency to perceive hostile intent in ambiguous situations ("paranoid" cognition) may persist because the cost of missing a real threat exceeds the cost of false alarms (Haselton et al., 2009).

This framework generates testable predictions: biases should be strongest in domains where ancestral error asymmetries were greatest, and should show predictable variation with factors like vulnerability and resource availability.

2.3 The Counterargument: When Consistency Helps

A significant challenge to this framework comes from clinical psychology. If cognitive inconsistency is adaptive, why do interventions like cognitive behavioral therapy (CBT)—which explicitly target inconsistent thinking—improve psychological functioning (Beck, 2011)?

Several responses are possible. First, CBT may work not by eliminating inconsistency but by redirecting it toward more adaptive patterns. Second, the inconsistencies targeted by CBT may represent pathological extremes rather than normal variation. Third, and most importantly, what was adaptive in ancestral environments may be maladaptive today. This evolutionary mismatch hypothesis (Li et al., 2018) suggests that our cognitive architecture, shaped for small-group living on the African savanna, may produce

systematic dysfunction in modern industrial societies.

3 Dual-Process Architecture: Neural Substrates of Cognitive Tension

Humans are biological animals endowed with a level of self-awareness that far exceeds immediate survival needs (Dehaene, 2014). This creates a structural tension. Instincts drive reproduction, survival, and competition, while consciousness enables reflection on meaning, morality, and coherence (Damasio, 1999).

3.1 Terminological Note

Throughout this paper, I use "dual-process architecture" to refer to the well-established distinction between fast, automatic (Type 1) and slow, deliberative (Type 2) cognitive processes (Evans and Stanovich, 2013). This terminology is preferred over "biological duality" to maintain consistency with the existing literature. The specific contribution here is emphasizing the *evolutionary origins* of these two systems and their relationship to temporal cognition—how dual-process architecture interacts with memory reconstruction and future projection to generate meaning-seeking behavior.

3.2 Neural Substrates

The dual-process distinction has clear neuroanatomical correlates. The limbic system, evolutionarily ancient, mediates emotional responses, fear conditioning, and reward processing (LeDoux, 1996). The prefrontal cortex, greatly expanded in humans, enables executive function, long-term planning, and moral reasoning (Miller and Cohen, 2001).

Critically, these systems can come into conflict. Neuroimaging studies reveal competition between limbic activation (e.g., amygdala responses to threat) and prefrontal regulation (Ochsner et al., 2012). The anterior cingulate cortex appears to monitor and mediate such conflicts (Botvinick et al., 2004).

This is not Cartesian dualism—both systems are fully physical and interact continuously. Rather, it represents a dual-process architecture in which fast, automatic, evolutionarily older systems operate alongside slow, deliberative, evolutionarily newer ones.

3.3 Mechanisms of Integration

How does the brain manage transitions between these modes? Several mechanisms have been identified:

- **Prefrontal inhibition**: The ventromedial prefrontal cortex can downregulate amygdala activity through inhibitory projections (Quirk et al., 2006).
- **Cognitive reappraisal**: Reinterpreting emotional stimuli reduces limbic responses (Ochsner et al., 2012).
- **Default mode network switching**: Transitions between task-focused and self-referential processing involve coordinated network changes (Raichle, 2015).

The experience of moral conflict may arise when these integration mechanisms fail or are overwhelmed, leaving competing systems simultaneously active.

3.4 Cross-Cultural Perspectives on Cognitive Integration

The tension between automatic and deliberative processing may be experienced differently across cultures. Markus and Kitayama (1991) documented fundamental differences between independent (typically Western) and interdependent (typically East Asian) self-construals. More relevant to the present framework, Nisbett et al. (2001) demonstrated that East Asian cognition tends toward holistic processing while Western cognition emphasizes analytic decomposition.

These differences suggest that the *management* of dual-process tensions—not just their existence—varies culturally. Holistic cognitive styles may integrate limbic and prefrontal inputs more fluidly, while analytic styles may emphasize prefrontal override. This has implications for Prediction 4 (cultural variation) in Section 11: different cultures may show not different *amounts* of inconsistency, but different *patterns* of which inconsistencies are tolerated versus resolved.

This remains a limitation requiring future research. Specific ethnographic and experimental work comparing, for example, Buddhist contemplative traditions' approaches to mental conflict with Western cognitive therapy approaches would substantially enrich this framework.

3.5 Comparative Cognition: The Cross-Species Evidence Gap

If cognitive inconsistency is adaptive, we should see graded versions in other species proportional to their environmental uncertainty and social complexity. This comparative evidence is crucial for the adaptationist argument but remains the weakest link in the present framework.

What we know:

- **Memory reconstruction**: Western scrub-jays show episodic-like memory and cache based on anticipated future states (Clayton and Dickinson, 1998). However, whether they *reconstruct* memories rather than simply retrieve them is unclear.
- **Temporal cognition**: Great apes show some capacity for mental time travel (Suddendorf and Corballis, 2007), planning for future needs. Interval timing is widespread across species, suggesting ancient evolutionary origins.
- **Superstition-like behavior**: Various species exhibit superstition-like behavior under uncertainty (Foster and Kokko, 2009), suggesting pattern over-attribution is not uniquely human.
- **Dual-process architecture**: Mammals share limbic-cortical organization, suggesting the hardware for dual-process conflict predates humans.

What remains unknown:

- Whether non-human primates experience *subjective conflict* between impulse and deliberation
- Whether meaning-seeking (beyond pattern-detection) exists in other species

- Whether planning biases (optimism, illusion of control) appear in non-humans
- The degree to which temporal illusions (time dilation under threat) are shared

What this means: The framework's adaptationist claims are best supported for dual-process architecture (clear mammalian homologs) and weakest for meaning-seeking (possibly uniquely human). Future research should systematically test whether social primates facing high uncertainty and complexity show human-like "irrational" patterns. If they do not, the phenomena may be human-specific byproducts of language and culture rather than general adaptations.

This comparative gap is a genuine weakness of the present argument, not merely a limitation of scope.

4 Perceptual Distortion and the Construction of Reality

Perception is not a passive recording of reality but an active interpretation shaped by evolutionary pressures (Hoffman et al., 2015). Time perception changes with emotion, memory reshapes the past, and expectations shape experience (Eagleman, 2008).

These distortions may be adaptive. A completely objective perception of reality would be computationally demanding. The mind simplifies, filters, and distorts to remain functional (Marr, 1982).

4.1 Predictive Processing: A Deeper Engagement

Predictive processing (PP) is arguably the dominant framework in contemporary cognitive neuroscience (Clark, 2013; Hohwy, 2013). Any evolutionary account of cognition must engage with it seriously. PP proposes that the brain is fundamentally a prediction machine, constantly generating expectations and updating them based on prediction errors.

The apparent tension: PP seems to prioritize consistency—the brain works to minimize prediction error, which appears to be the opposite of "adaptive inconsistency."

Resolution through precision-weighting: The key insight is that PP's "consistency" operates through *precision-weighting* (Friston, 2012; Parr and Friston, 2019). The brain does not treat all prediction errors equally; it strategically weights certain errors as more or less important based on context, reliability, and relevance. This mechanism may explain how the brain tolerates global inconsistencies while maintaining local coherence:

- High-precision weighting on survival-relevant errors ensures rapid response to threats (explaining hypervigilance in Case 1)
- Low-precision weighting on abstract inconsistencies allows action despite logical contradictions
- Context-dependent precision adjustment enables flexible switching between thorough analysis and quick heuristics
- Precision can be allocated to *expected* prediction errors (active inference), allowing motivated cognition

PP as the mechanism: Rather than being in tension with the present framework, PP may provide the *computational mechanism* for the phenomena described. The "irrationalities" documented are not failures of prediction error minimization but features of how precision is allocated:

- **Memory reconstruction**: High precision on gist, low precision on detail
- **Temporal distortion**: Precision shifts under threat alter subjective time
- **Meaning-seeking**: Pattern-completion with high prior precision
- **Dual-process conflict**: Competition between high-precision limbic predictions and high-precision prefrontal predictions

This reframing suggests that the "strong claim" (unified architecture) might find its substrate in precision-weighting mechanisms. If precision allocation is the common computational currency, the four phenomena would indeed share a mechanism. This remains speculative but generates testable predictions: manipulating precision (e.g., through attention, uncertainty, or neuromodulation) should co-modulate all four phenomena.

5 Memory as Reconstruction Rather Than Storage

Memory does not archive the past; it reconstructs it (Schacter et al., 2012). Recollection is influenced by current beliefs, emotions, and context. The past exists in consciousness only as a present reconstruction (Loftus, 2005).

This undermines the notion of a fixed personal history and supports the idea that human identity is dynamically reinterpreted rather than statically preserved (Conway, 2005). Neuroimaging studies have confirmed that memory retrieval activates many of the same neural circuits involved in imagination and future planning, suggesting a common constructive mechanism (Addis et al., 2007).

5.1 Adaptive Functions of Memory Distortion

If memory evolved for accurate recording, its systematic distortions would be puzzling. But if memory evolved to guide future behavior, reconstruction becomes sensible (Schacter, 2012):

- **Updating**: Memories that incorporate new information remain useful for current decisions.
- **Gist extraction**: Retaining meaning while losing detail conserves cognitive resources.
- **Self-enhancement**: Positively biased memories may support psychological well-being and motivation (Taylor and Brown, 1988).

6 The Illusion of Future Control

Humans plan extensively, yet the future remains largely unpredictable (Gilbert, 2007). Planning creates an illusion of control that reduces anxiety and enables long-term behavior (Langer, 1975). The future, like the past, exists only as a mental construction in the present.

6.1 Systematic Biases in Future Thinking

This projection is necessary for motivation but structurally detached from actual future events (Seligman et al., 2016). Several well-documented biases illustrate this detachment:

- **Planning fallacy**: Systematic underestimation of time, costs, and risks for future tasks (Kahneman and Tversky, 1979).
- **Impact bias**: Overestimation of the emotional intensity and duration of future events (Wilson and Gilbert, 2005).
- **Optimism bias**: Unrealistic expectations about personal futures relative to statistical baselines (Sharot, 2011).

The capacity for mental time travel, while enabling unprecedented behavioral flexibility, also introduces these systematic biases in prediction and decision-making (Suddendorf and Corballis, 2007).

6.2 Ancestral Utility vs. Modern Dysfunction: Economic Implications

These biases may have been adaptive in ancestral environments where undertaking risky ventures (hunting, migration, mate competition) required optimistic projections to overcome legitimate fears. In modern contexts, however, the same biases contribute to cost overruns in construction projects, inadequate retirement savings, and unrealistic business plans.

This evolutionary mismatch (Li et al., 2018) has particular relevance for financial behavior. The ancestral "risk-taking for resources" that motivated dangerous hunts may translate directly into modern gambling behavior and speculative trading. Thaler and Sunstein (2008) document how systematic biases lead to predictable financial errors; the present framework suggests these are not correctable "mistakes" but expressions of cognitive architecture shaped for different environments. Implications include:

- Financial regulations should assume bias rather than rationality
- Retirement systems should use automatic enrollment exploiting status quo bias
- Trading systems might benefit from "cooling off" periods that allow prefrontal override of limbic impulses

7 The Present as the Only Locus of Influence

Because the past is reconstructed and the future is projected, real influence exists only in the present moment (Varela, 1999). This is not merely a philosophical position but a cognitive constraint arising from how consciousness operates.

Human dissatisfaction often stems from living in reconstructed pasts and imagined futures rather than in the only actionable temporal space (Killingsworth and Gilbert, 2010). Mindfulness research has demonstrated that interventions targeting present-moment awareness can significantly alter psychological well-being, supporting the functional importance of temporal orientation (Brown and Ryan, 2003).

7.1 Connecting Present-Moment Awareness to Narrative Identity

This raises an important question: how does present-moment awareness relate to the narrative construction of identity discussed earlier? The answer involves temporal integration.

The present moment is where reconstruction of the past and projection of the future *occur*. Narrative identity (McAdams, 2001) is not stored somewhere but actively constructed in each present moment. This suggests that interventions targeting present-moment awareness may work partly by allowing more flexible, less automatic narrative construction—creating space between stimulus and response where identity can be renegotiated.

8 Meaning-Seeking as an Evolutionary Trait

Humans compulsively seek meaning, even where none objectively exists (Frankl, 1985). This tendency likely evolved as a pattern-recognition mechanism for detecting threats and opportunities. Foster and Kokko (2009) demonstrate mathematically that superstition-like behavior—perceiving patterns and causal relationships that do not exist—can be favored by selection when the cost of false negatives exceeds the cost of false positives.

Meaning-making is thus not a purely philosophical endeavor but an extension of survival cognition (Baumeister, 1991). The capacity for narrative construction allows humans to integrate disparate experiences into coherent life stories, providing psychological continuity despite constant change (McAdams, 2001).

8.1 The Paradox of Self-Sacrifice

If meaning-seeking is merely a survival mechanism, how do we account for meaning-driven behaviors that contradict immediate survival—martyrdom, self-sacrifice, choosing death over dishonor?

Several explanatory frameworks exist, each with distinct empirical implications:

1. **Inclusive fitness**: Self-sacrifice that benefits genetic relatives can be favored by kin selection (Hamilton, 1964). *Prediction*: Self-sacrifice should correlate with genetic relatedness to beneficiaries. *Evidence*: Studies of heroic rescue behavior show increased risk-taking for kin (Burnstein et al., 1994).

2. **Reputation and reciprocity**: Costly signaling of commitment may yield long-term benefits through enhanced reputation and reciprocal altruism (Zahavi and Zahavi, 1997). *Prediction*: Self-sacrifice should be more common when observed by reputation-relevant audiences. *Evidence*: Charitable giving increases with publicity (Harbaugh, 1998).

3. **Cultural evolution**: Meaning systems that promote group cohesion may spread through cultural group selection, even if costly to individuals (Richerson and Boyd, 2005). *Prediction*: Self-sacrifice norms should be stronger in groups facing inter-group competition. *Evidence*: Military and religious martyrdom correlates with group conflict intensity.

4. **Misfiring**: Self-sacrifice may represent a byproduct of meaning-systems that are generally adaptive but occasionally misfire in extreme contexts. *Prediction*: Self-sacrifice should show characteristics of "hijacked" psychological mechanisms. *Evidence*: Suicide terrorism exploits kinship psychology through fictive kin terminology (Atran, 2003).

These explanations are not mutually exclusive; different instances of self-sacrifice may involve different mechanisms, and they may operate simultaneously within single cases. Inclusive fitness and reputation/reciprocity, for example, could both contribute to the same act of heroism. The framework I find most compelling combines cultural evolution with misfiring: meaning systems evolved for group cohesion can, under specific cultural conditions, generate self-sacrifice as an emergent property. However, adjudicating between these accounts requires more targeted empirical research than currently exists.

9 Integrating the Framework: Strong and Weak Claims

The phenomena discussed—cognitive contradiction, dual-process architecture, perceptual construction, memory reconstruction, temporal illusion, and meaning-seeking—have typically been treated separately. This paper proposes they may be related, but the strength of that relationship claim requires careful statement.

9.1 The Weak Claim: Interacting Systems

The **weak claim**, which I consider well-supported, is that these phenomena *interact* and share *common evolutionary pressures*. Dual-process conflict affects meaning construction; memory reconstruction uses the same mechanisms as future projection; temporal perception modulates decision-making. These interactions are empirically established (Addis et al., 2007; Schacter et al., 2012).

Under this interpretation, irrationality, temporal illusion, meaning-seeking, and inner conflict are *separate systems* that influence each other and were shaped by overlapping selection pressures (action under uncertainty, social coordination, identity maintenance). Their co-occurrence across domains (documented in Appendix A) reflects interaction, not identity.

9.2 The Strong Claim: Unified Architecture

The **strong claim**, which remains speculative, is that these phenomena are expressions of a *single underlying architecture*—that they share not just evolutionary pressures but computational mechanism. This would require identifying a specific neural or computational substrate that generates all four phenomena.

Candidate mechanisms include:

- **Precision-weighting in predictive processing**: The same precision-modulation system might generate selective attention (apparent irrationality), temporal distortion (precision-weighted prediction error), meaning attribution (pattern completion), and conflict (competing high-precision predictions) (Parr and Friston, 2019).
- **Default mode network function**: The DMN is implicated in self-referential processing, episodic memory, future simulation, and mind-wandering—potentially a neural hub for the integrated architecture (Raichle, 2015).
- **Dopaminergic valuation**: Dopamine signals encode prediction error across domains; dysfunction produces both motivational (meaning) and temporal (interval timing) deficits simultaneously.

However, demonstrating that these candidates actually implement a *unified* architecture rather than merely *interacting* systems requires evidence this paper cannot provide. The strong claim should be understood as a hypothesis for future research, not a conclusion.

9.3 What Would Distinguish the Claims?

The weak and strong claims make different predictions:

- **Dissociability**: Under the weak claim, the four phenomena should be experimentally dissociable—one could be manipulated without affecting others. Under the strong claim, they should show obligatory co-variation.
- **Neural localization**: Under the weak claim, different phenomena should show distinct neural signatures with some overlap. Under the strong claim, a common substrate should be identifiable.
- **Individual differences**: Under the weak claim, people could vary independently on each dimension. Under the strong claim, individual differences should be correlated.

Current evidence is insufficient to adjudicate. The Appendix demonstrates co-occurrence, but co-occurrence is consistent with both claims.

9.4 Schematic Representation

Figure 1 illustrates the proposed relationships. *Solid arrows* indicate relationships with established empirical support; *the overall integration* remains hypothetical.

Memory-Future Interaction: The bidirectional arrow between memory reconstruction and future projection reflects their shared neural substrates (Addis et al., 2007).

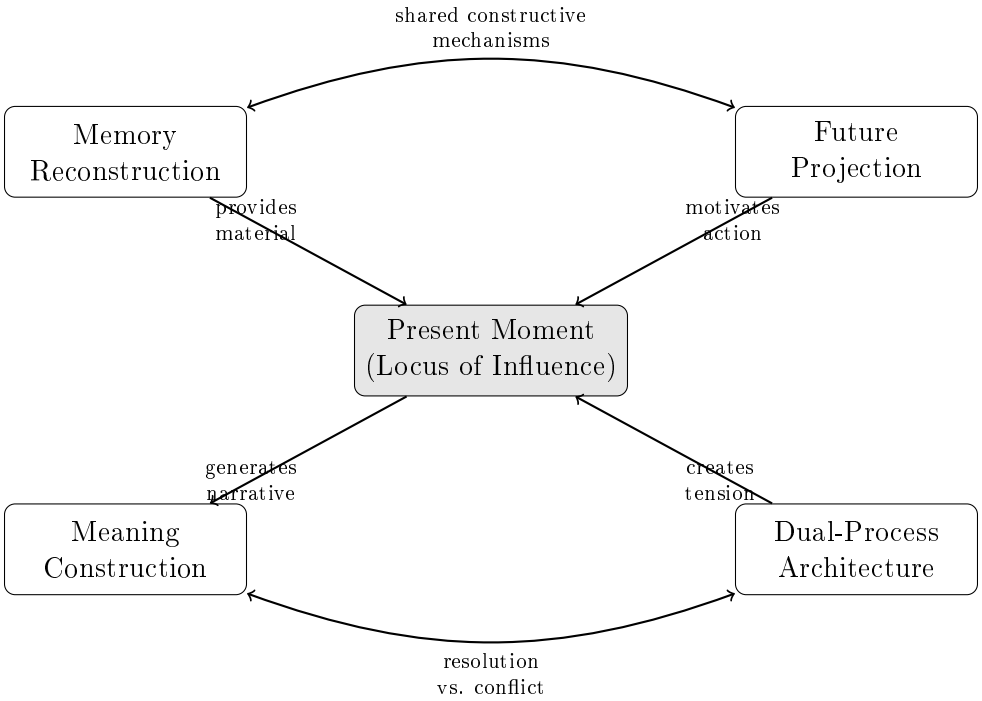

Figure 1: Schematic representation of the proposed integrative framework. The present moment serves as the locus where memory reconstruction and future projection occur, while dual-process architecture creates the tensions that meaning construction attempts to resolve. Bidirectional arrows indicate mutual influence: memory and future projection share constructive neural mechanisms (Addis et al., 2007), while meaning construction both resolves and is shaped by dual-process conflicts.

Episodic future thinking draws on recombined elements of past experience; conversely, anticipated futures can reshape how past events are remembered (e.g., reinterpreting past failures as "learning experiences" when future success is expected).

Meaning-Duality Interaction: Meaning construction can either resolve or intensify dual-process conflicts. A coherent life narrative may integrate limbic impulses and prefrontal values into a unified self-concept (resolution). Alternatively, meaning systems that condemn natural impulses may intensify experienced conflict (e.g., religious frameworks that frame bodily desires as sinful).

Table 2 summarizes how each phenomenon can be viewed through different interpretive lenses. Note that both columns represent legitimate perspectives; the goal is not to privilege the "adaptive" interpretation but to show that what appears as dysfunction may sometimes serve functions not immediately apparent.

10 Interdisciplinary Implications

Understanding human behavior requires integrating biology, psychology, philosophy, and cognitive science (Gazzaniga, 2004). Specialization obscures the unified nature of these phenomena, which may originate from the same evolutionary architecture.

Table 2: Comparison of Interpretive Frameworks for Cognitive Phenomena

Phenomenon	Deficit Interpretation	Functional Interpretation	Key References
Cognitive contradiction	Deviation from logical norms	Enables rapid decision-making under uncertainty	Gigerenzer (1999); Simon (1955)
Memory distortion	Failure of accurate encoding/retrieval	Allows updating and flexible identity construction	Schacter (2012); Loftus (2005)
Future projection bias	Systematic prediction errors	Motivates action despite genuine uncertainty	Gilbert (2007); Sharot (2011)
Pattern overattribution	False belief formation	Minimizes costly missed detections	Foster & Kokko (2009)
Dual-process conflict	Executive function failure	Enables both survival responses and reflection	Evans & Stanovich (2013)

10.1 Clinical Psychology and Individual Differences

If cognitive inconsistencies are sometimes adaptive, clinical interventions should aim not to eliminate them but to redirect them. This aligns with third-wave cognitive therapies (ACT, DBT) that emphasize acceptance and flexibility rather than correction of "distorted" thinking (Hayes et al., 2006).

The rigidity distinction: The key clinical distinction may not be between consistent and inconsistent thinking, but between *flexible* and *rigid* inconsistency. Adaptive inconsistency is context-sensitive—the mind tolerates different contradictions in different situations. Pathological inconsistency may be rigid application of normally adaptive patterns regardless of context.

Individual differences in distress: People vary substantially in their tolerance for holding contradictory beliefs. Some individuals experience significant distress from cognitive dissonance; others show minimal discomfort. This variation poses a question for the framework: if inconsistency is adaptive, why would some people be distressed by it?

Possible answers include:

- **Trait variation in integration mechanisms**: Individual differences in anterior cingulate function may produce different experiences of conflict
- **Cultural learning**: Some cultural environments may train explicit attention to contradictions that are normally processed implicitly
- **Domain-specificity**: Distress may arise when inconsistency intrudes into domains where consistency is valued (moral identity, professional competence)
- **Meta-cognitive beliefs**: Beliefs about whether one "should" be consistent may amplify distress from normal inconsistency

This individual variation is not well explained by the current framework and represents an area requiring further development.

10.2 Artificial Intelligence and Alignment

AI systems designed for perfect consistency may lack important properties of human cognition. The framework suggests that apparent "bugs" in human cognition may solve problems that purely consistent systems cannot (Stanovich, 2004).

This initially counterintuitive claim—that inconsistency might make AI *safer*—deserves careful elaboration. Current AI alignment concerns often focus on systems that pursue objectives too consistently, without the "second thoughts" that characterize human moral reasoning (Russell, 2019). The dual-process architecture described here provides exactly such second thoughts: limbic responses can interrupt prefrontal planning when something "feels wrong," even if logically justified.

The counterargument: One could argue equally that consistent AI following well-specified values would be safer than AI that "doubts its own objectives." If we could specify values correctly, doubt would only introduce error. The human-like inconsistency proposed here might make AI more *human-compatible* but not necessarily safer.

Response: The counterargument assumes correct value specification is achievable. The case for beneficial inconsistency rests on the assumption that specification will always be incomplete or imperfect. Under this assumption, systems that can recognize "this feels wrong even though it satisfies my objectives" may catch specification errors that pure consistency would execute. This is analogous to the human experience of moral intuitions overriding explicit reasoning.

Implications for AI development (highly speculative):

- AI safety may require building in mechanisms analogous to human dual-process conflicts
- Systems that can "doubt" their own objectives may be more robust to specification errors
- Human-AI interaction interfaces should accommodate human inconsistency rather than demanding logical precision
- The goal of "artificial general intelligence" may require tolerating inconsistency as a feature, not a bug

This represents a direction for future research rather than a developed proposal; the technical implementation of "beneficial inconsistency" in AI systems remains an open question. The argument should be understood as speculative even by the standards of this paper.

10.3 Legal and Economic Systems

Systematic cognitive biases have well-documented effects on legal judgments and economic behavior. The evolutionary mismatch perspective suggests that simply "debiasing" may be insufficient; instead, systems should be designed to accommodate predictable patterns (Thaler and Sunstein, 2008).

10.4 Relationship to Evolutionary Consciousness Theories

This framework intersects with several major theories of consciousness. Here I briefly consider how it relates to alternatives.

Integrated Information Theory (IIT) (Tononi et al., 2016) proposes that consciousness corresponds to integrated information (Φ). The present framework is agnostic about what makes a system conscious but addresses what conscious systems *do*—construct temporal narratives, seek meaning, experience conflict. IIT and the present proposal operate at different levels of analysis (mechanism vs. function) and are potentially complementary.

Global Workspace Theory (GWT) (Dehaene, 2014) emphasizes the broadcasting of information to multiple cognitive systems. The dual-process conflicts described here might correspond to competition for workspace access: limbic signals and prefrontal evaluations compete for broadcast, generating experienced conflict when neither dominates.

Attention Schema Theory (AST) (Graziano and Webb, 2015) proposes that consciousness is a model the brain constructs of its own attention. This is potentially compatible with the present framework: if the brain models its own attentional conflicts, the "inner conflict" phenomenon would be attention schema representing competing attentional priorities. AST might provide a mechanistic account of *how* conflict is experienced.

Social Brain Hypothesis (Dunbar, 1998) argues that primate brain expansion was driven by social complexity. This aligns directly with the present framework's emphasis on social coordination, reputation management, and coalitional psychology. The "irrationalities" described may be specifically social-cognitive adaptations, explaining why they are pronounced in social domains (public mistake, romance, good deed cases).

The present framework's contribution is not a new theory of consciousness mechanism but an integrative account of consciousness *content*: why humans experience the specific inconsistencies, temporal distortions, meanings, and conflicts that they do. It is compatible with multiple mechanistic theories.

11 Testable Predictions and Methodologies

A theoretical framework is only valuable if it generates predictions that can be empirically evaluated. This framework suggests:

1. **Ecological contingency**: Cognitive biases should be stronger in domains matching ancestral selection pressures (threat detection, mate choice, resource acquisition) than in evolutionarily novel domains.

   Methodology: Compare bias magnitude across domains using standardized measures. Predict larger effect sizes for biases in ancestral-relevant domains (e.g., snake detection) versus novel domains (e.g., stock market prediction).

2. **Stress modulation**: Under stress, the balance should shift toward faster, more heuristic processing as prefrontal resources are diverted.

   Methodology: Experimental paradigm: induce stress (e.g., cortisol administration, social evaluative threat) and measure changes in planning fallacy magnitude. *Specific prediction*: Participants in high-cortisol conditions should show statistically significant increases in optimistic time/cost estimates compared to controls.

3. **Individual differences**: People with greater prefrontal control should show more consistent but potentially less adaptive responding in uncertain environments.

   Methodology: Correlate executive function measures (e.g., Stroop performance, working memory capacity) with both logical consistency and ecological performance on uncertain decision tasks. Existing literature on executive function suggests this may show a curvilinear relationship, with moderate EF optimal (Diamond, 2013).

4. **Cultural variation**: Cultures with different cognitive styles should show different patterns of which inconsistencies are tolerated versus resolved.

   Methodology: Cross-cultural comparison of holistic (East Asian) versus analytic (Western) samples on tasks measuring tolerance for logical contradiction. Predict that holistic cognition correlates with greater tolerance for certain inconsistencies but not others.

5. **Developmental trajectory**: The integration of dual-process systems should follow predictable developmental patterns linked to prefrontal maturation.

   Methodology: Longitudinal neuroimaging combined with behavioral measures of dual-process conflict across adolescent development. Predict that experienced moral conflict peaks during periods of maximal prefrontal-limbic connectivity reorganization.

12 Conclusion

Human cognitive contradictions, moral struggles, memory distortions, and temporal confusion may not be failures of reason but consequences of the evolutionary design of consciousness. The human mind appears built not for consistency but for survival in uncertain environments. Recognizing this provides a potential framework for understanding human behavior, identity, and experience.

The search for meaning, the conflict between impulse and deliberation, and the difficulty of living in the present may all emerge from a single source: the evolutionary architecture of human consciousness, specifically the interaction between dual-process cognitive systems and temporal construction mechanisms. This integrated perspective offers both explanatory potential and practical implications for clinical psychology, artificial intelligence, and institutional design.

However, significant caution is warranted. The adaptationist interpretation offered here is one among several possible explanations. The phenomena discussed may be byproducts, constraints, or cultural constructions rather than adaptations. What appears as a unified system may be a post-hoc narrative imposed on genuinely disparate mechanisms. Empirical research testing the specific predictions generated by this framework is needed before strong conclusions can be drawn.

What this paper offers is not a definitive account but an invitation to consider human "irrationality" from a different perspective—one that may reveal method in the apparent inconsistencies of the human mind.

A Case Studies: The Unified Architecture in Action

The theoretical framework developed in this paper proposes that human "irrationality," temporal illusions, meaning-seeking, and inner conflict are not separate phenomena but manifestations of a single evolutionary mechanism optimized for action under uncertainty rather than truth-tracking. This appendix presents a series of case studies demonstrating this unity. Each case follows a consistent format: a recognizable scenario, the four phenomena as they manifest, the underlying computational goal, and testable predictions. If the same computational logic generates all four phenomena across diverse situations, their structural connection is demonstrated.

A.1 Case 1: The Night Noise

Scenario: A person wakes at 3 AM to an unfamiliar sound in the house. Heart rate accelerates, attention narrows, muscles tense.

Apparent Irrationality: The response magnitude far exceeds the probability of actual danger. Most night sounds are benign—settling wood, wind, pets. Yet the body responds as if to genuine threat. This seems irrational.

Temporal Illusion: Seconds stretch into apparent minutes. The interval between sound and investigation feels interminable. Time "slows down" under threat perception (Eagleman, 2008).

Meaning-Seeking: The mind cannot tolerate unexplained sensory input. Within moments, narratives form: "intruder," "animal," "something fell." The brain constructs explanation even without evidence.

Inner Conflict: The prefrontal cortex recognizes the irrationality ("this is probably nothing"). The limbic system insists on vigilance. The person lies awake, caught between dismissal and alarm.

Underlying Mechanism: Error management under asymmetric costs. A false alarm (unnecessary vigilance) costs a few minutes of sleep. A false negative (ignoring real threat) could cost survival. The system is calibrated for ancestral environments where nocturnal sounds often *were* predators or hostile humans.

Computational Goal: Minimize worst-case outcomes rather than expected outcomes.

Testable Prediction: Temporal dilation should correlate with perceived threat magnitude. Meaning-construction speed should increase with ambiguity. Conflict intensity should correlate with prefrontal-limbic connectivity.

A.2 Case 2: The Public Mistake

Scenario: A person misspeaks during a presentation—a minor error noticed by few. Weeks later, the moment still intrudes, replaying vividly.

Apparent Irrationality: The cognitive investment is wildly disproportionate. Hours of rumination over seconds of minor embarrassment. No productive outcome results.

Temporal Illusion: The past event remains "present." It has not faded into memory but persists with the vividness of recent experience. The normal temporal distance between now and then collapses.

Meaning-Seeking: Endless interpretation ensues. "What does this say about me?" "How did others perceive it?" "Will this affect my reputation?" The search for meaning

far exceeds the meaning available.

Inner Conflict: One part wants to move on ("it doesn't matter"). Another insists on continued analysis and self-criticism. The conflict itself consumes resources.

Underlying Mechanism: Social risk management in a species for whom reputation determines survival. Ancestrally, social rejection could mean death—loss of coalition protection, mating opportunities, resource sharing. The system treats reputational threats as near-survival threats.

Computational Goal: Maintain social standing through hypervigilance to potential status damage.

Testable Prediction: Rumination intensity should correlate with social evaluative concern (measurable via Social Interaction Anxiety Scale). The effect should be stronger in interdependent cultures.

A.3 Case 3: The Lottery Ticket

Scenario: A person buys a lottery ticket despite "knowing" the odds are essentially zero. After purchase, anticipation builds; numbers seem meaningful.

Apparent Irrationality: Expected value calculation clearly indicates no rational person should buy lottery tickets. Yet millions do. This appears to be innumeracy or cognitive failure.

Temporal Illusion: The days between purchase and drawing compress around the anticipated event. The future result feels imminent, almost present. Waiting time is experienced as "countdown" rather than neutral duration.

Meaning-Seeking: The buyer perceives patterns: birthdates, "lucky" numbers, meaningful sequences. The random selection process becomes pregnant with significance. Coincidences suggest fate.

Inner Conflict: Rational self-assessment ("this is foolish") conflicts with persistent hope ("someone has to win"). The internal debate between probability and possibility.

Underlying Mechanism: High-variance strategies under resource constraint. In ancestral environments, individuals with few resources faced a choice: accept low-status equilibrium or take high-risk/high-reward gambles (dangerous hunts, risky migrations, challenging dominant males). The willingness to bet on small probabilities of large gains may have been adaptive for those with little to lose.

Computational Goal: Maintain motivation for high-variance opportunities where expected value calculations would counsel inaction.

Testable Prediction: Lottery participation should correlate inversely with socioeconomic security. Temporal compression around the drawing should intensify with ticket investment.

A.4 Case 4: The Annual Plan

Scenario: A person creates an elaborate plan for the coming year—career goals, health targets, relationship milestones. The act of planning produces relief, optimism, and a sense of control.

Apparent Irrationality: Planning fallacy is well-documented (Kahneman and Tversky, 1979). Most plans fail or require substantial revision. Yet people consistently overestimate their control over future outcomes.

Temporal Illusion: The planned future feels concrete, almost real—as if it already exists as a fixed timeline. The distinction between projection and reality blurs. "Seeing" oneself achieving goals feels like partial achievement.

Meaning-Seeking: The plan becomes a moral narrative: "If I execute correctly, the world owes me results." Goals transform from instrumental targets to identity-defining missions. Failure becomes not just setback but betrayal.

Inner Conflict: When plans derail, internal blame emerges. "I should have tried harder" competes with "circumstances were unfair." The conflict between agency and acceptance.

Underlying Mechanism: Agency maintenance. Action requires believing that action matters. A system that accurately represented the limited control individuals have over outcomes would demotivate effort. The illusion of control is the fuel for sustained engagement (Langer, 1975).

Computational Goal: Maintain behavioral output despite uncertainty about outcomes.

Testable Prediction: Individuals with higher sense of agency should show greater planning fallacy. Depressive realism (more accurate probability assessment) should correlate with reduced goal-directed behavior.

A.5 Case 5: Romantic Attachment

Scenario: A person falls in love—idealizing the partner, ignoring warning signs, experiencing intense preoccupation.

Apparent Irrationality: The cognitive distortions of romantic love are well-documented: idealization, projection, selective attention. The beloved is perceived as more attractive, more compatible, more unique than objective assessment would warrant.

Temporal Illusion: Together, time "flies." Apart, it "drags." Memory retouches shared experiences into golden glow. The past becomes more beautiful than it was; the anticipated future more certain than it can be.

Meaning-Seeking: The relationship becomes proof of destiny, cosmic significance. "We were meant to meet." Coincidences become signs. The bond carries metaphysical weight far beyond its statistical probability.

Inner Conflict: Biology pulls toward attachment; observation notices incompatibilities. The heart-mind conflict is literal: reward circuitry versus evaluative cortex. Suffering emerges from the gap.

Underlying Mechanism: Reproductive and coalition bonding requires commitment devices. If mate choice were purely rational, minor setbacks would dissolve relationships. The species requires bonds that persist through difficulty. "Irrational" attachment is the glue.

Computational Goal: Maintain pair-bond stability through motivational bias strong enough to override rational cost-benefit analysis.

Testable Prediction: Idealization magnitude should predict relationship persistence. Temporal distortion during separation should correlate with attachment security measures.

A.6 Case 6: The Good Deed

Scenario: A person helps a stranger at personal cost—time, money, risk. A warm feeling follows. They remember the act fondly for years.

Apparent Irrationality: Strict resource maximization suggests this is a mistake. Resources given to non-kin with no expectation of return appear evolutionarily inexplicable.

Temporal Illusion: Positive moral memories become "anchor points" in autobiographical time—vivid, easily retrieved, seemingly recent. Shameful memories have opposite valence: stuck, present, unprocessed.

Meaning-Seeking: The act is woven into identity narrative: "I am the kind of person who helps." Moral meaning provides behavioral stability, guiding future choices through self-concept rather than situation-by-situation calculation.

Inner Conflict: Altruistic impulse competes with resource conservation. The "angel and devil" experience of moral temptation. Neither voice is fully silenced.

Underlying Mechanism: Coalitional psychology and reputation economics. In ancestral small-group environments, reputation was public information. Costly altruism signaled coalition value, attracting reciprocal allies. The internal reward system evolved to motivate such investment.

Computational Goal: Generate reliable other-regarding behavior through intrinsic motivation, reducing dependence on external enforcement.

Testable Prediction: Warm glow magnitude should correlate with perceived observability (even subliminal audience cues). Moral memory vividness should predict future prosocial behavior.

A.7 Case 7: The Meaning Collapse

Scenario: A person experiencing depression reports that "nothing matters." Time feels viscous; decisions become impossible; the past loops endlessly; the future appears blank.

Note: This case must be handled with care, avoiding harmful detail while illustrating theoretical structure.

Apparent Irrationality: The global devaluation of all goals appears maladaptive. Nothing the person previously cared about generates motivation.

Temporal Illusion: Time becomes "sticky." The present extends without resolution; the past intrudes repetitively; the future loses coherence. Normal temporal flow fragments.

Meaning-Seeking: Not absent but inverted—compulsive search for why things are meaningless. "What's the point?" becomes obsessive rather than resolved.

Inner Conflict: Paralyzed rather than resolved. Neither impulse nor deliberation generates action. The conflict itself stalls.

Underlying Mechanism: This case reveals the architecture by its absence. Depression may represent failure of the action-value computation system that normally weights options for behavioral output. When the system stops assigning value, meaning, time, and decision-making collapse *together*—because they were never separate.

Computational Goal Failure: The system that normally generates "this action is worth performing" has stopped producing output. The unified collapse demonstrates unified architecture.

Clinical Implication: Interventions targeting any component (meaning through narrative therapy, time through behavioral activation, conflict through acceptance) may

restore the system because the components are coupled.

A.8 Case 8: The Edited Memory

Scenario: A person recalls their past in ways that support their current self-concept—minimizing past mistakes, emphasizing growth, constructing a coherent narrative of becoming.

Apparent Irrationality: Memory distortion seems obviously maladaptive. Shouldn't accurate recall aid future decisions? Yet reconstruction is the norm (Schacter et al., 2012).

Temporal Illusion: The past literally changes based on the present. What "actually happened" shifts to serve current identity needs. The boundary between memory and imagination blurs.

Meaning-Seeking: Events become "lessons," "turning points," "signs." The narrative arc requires revision—sometimes radical—to maintain coherence with who the person believes they are now.

Inner Conflict: Commitment to truth ("I should remember accurately") versus commitment to self ("I need a usable past"). Occasionally, the conflict surfaces as confrontation with external evidence.

Underlying Mechanism: Identity stability supports action. A person who constantly questioned their own history would struggle to commit to present choices. Functional memory prioritizes usability over accuracy (Schacter, 2012). The self that acts today needs a past that makes today's actions sensible.

Computational Goal: Maintain coherent identity sufficient to support stable social roles and sustained action plans.

Testable Prediction: Memory revision should increase following identity-threatening experiences. Self-enhancing recall should correlate with action confidence. Individuals with less stable self-concept should show more radical memory revision.

A.9 Summary: Interacting Systems or Unified Architecture?

Across the eight cases, a consistent pattern appears:

However, demonstrating co-occurrence does not establish unified architecture. These cases are consistent with both the strong claim (single mechanism) and weak claim (interacting systems). The pattern supports integration but does not prove it.

A.10 Boundary Conditions: Cases That Don't Fit Cleanly

To avoid circularity, we must consider cases where the framework's predictions are unclear or where phenomena dissociate.

Case 9: Flow States

In flow states (complete absorption in skilled activity), the phenomenology diverges from predictions:

- **Rationality**: Performance is optimized, not biased—the opposite of "adaptive irrationality"
- **Temporal illusion**: Time distortion occurs (hours feel like minutes), consistent with the framework

Table 3: Phenomena Across Case Studies

Case	Computational Goal	Four Manifestations
Night Noise	Minimize worst-case threat outcomes	Hypervigilance, time dilation, narrative completion, rational-limbic conflict
Public Mistake	Protect social standing	Rumination, temporal persistence, self-interpretation, move-on vs. analyze
Lottery	Maintain high-variance motivation	Hope despite odds, anticipation compression, pattern perception, fool vs. dreamer
Annual Plan	Sustain agency and action	Control illusion, future concretization, moral narrative, self vs. fate blame
Romance	Stabilize reproductive bonds	Idealization, time distortion, destiny meaning, heart vs. mind
Good Deed	Generate reliable prosociality	Warm glow, memory anchoring, identity narrative, altruism vs. self-interest
Meaning Collapse	[System failure]	Unified collapse demonstrates system coupling
Edited Memory	Maintain actionable identity	Self-serving recall, past revision, lesson construction, truth vs. function

- **Meaning-seeking**: Meaning is *suspended*, not constructed—the activity is its own justification
- **Inner conflict**: *Absent*—the defining feature of flow is the disappearance of self-critical inner dialogue

This partial fit is instructive. Flow may represent a state where the "action under uncertainty" optimization problem is temporarily solved: the environment is structured, feedback is immediate, skills match challenges. The architecture may generate inconsistency and conflict specifically when uncertainty is high; flow is what the system looks like when uncertainty is low.

Implication: The framework predicts that irrationality and conflict should correlate with environmental uncertainty. Flow states should be more common in structured, predictable task environments.

Case 10: Autism Spectrum Cognition

Some features of autism spectrum cognition challenge the framework:

- **Reduced "irrationality"**: Some biases (e.g., conjunction fallacy) are reduced in autism
- **Altered temporal perception**: Present but differently organized
- **Meaning-seeking**: Often intensified (systematizing, pattern-detection) but in different domains
- **Inner conflict**: May be organized differently; some report reduced "intuitive" moral responses

This case is complex because autism involves both differences in the phenomena themselves and in their integration. If the "unified architecture" claim is correct, autism might represent altered integration rather than absence of components—consistent with some theoretical accounts emphasizing atypical predictive processing.

Implication: The framework does not straightforwardly apply. This is either a limitation or evidence that the architecture is neurotypical-specific.

A.11 What These Boundary Cases Reveal

The partial fits and misfits are informative:

1. **Context-dependence**: The architecture may generate inconsistency specifically under uncertainty; structured environments may not engage it
2. **Individual variation**: The framework may describe modal human cognition while allowing substantial individual differences
3. **Possible dissociations**: Flow states suggest meaning-seeking and inner conflict can be suspended while temporal distortion persists—partial dissociation

These boundary conditions constrain the framework's scope. It appears to describe cognition *under uncertainty in social contexts* rather than cognition universally. This is consistent with the evolutionary pressures emphasized (action despite incomplete information, social coordination) but limits generality.

A.12 Conclusion: Evidence for Interaction, Hypothesis of Unity

The case studies demonstrate that irrationality, temporal illusion, meaning-seeking, and inner conflict *co-occur* across diverse everyday situations and *appear to serve common computational goals*. This supports the weak claim that these are interacting systems shaped by shared evolutionary pressures.

The strong claim—that they reflect a single unified architecture—remains a hypothesis. The co-occurrence pattern is consistent with unity but does not require it. Future research should focus on dissociation paradigms: can these phenomena be experimentally separated? If not, unity is supported. If so, the weak claim is correct.

What the case studies do establish is that treating these as four separate problems misses systematic connections. Whether the connection is interaction or identity, an integrated approach is warranted.

References

Addis, D. R., Wong, A. T., and Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, 45(7):1363–1377.

Atran, S. (2003). Genesis of suicide terrorism. *Science*, 299(5612):1534–1539.

Baumeister, R. F. (1991). *Meanings of Life*. Guilford Press, New York.

Beck, J. S. (2011). *Cognitive Behavior Therapy: Basics and Beyond*. Guilford Press, New York, 2nd edition.

Botvinick, M. M., Cohen, J. D., and Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, 8(12):539–546.

Brown, K. W. and Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, 84(4):822–848.

Burnstein, E., Crandall, C., and Kitayama, S. (1994). Some neo-Darwinian decision rules for altruism: Weighing cues for inclusive fitness as a function of the biological importance of the decision. *Journal of Personality and Social Psychology*, 67(5):773–789.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204.

Clayton, N. S. and Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, 395(6699):272–274.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, 53(4):594–628.

Damasio, A. (1999). *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Harcourt Brace, New York.

Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking, New York.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64:135–168.

Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology*, 6(5):178–190.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, 18(2):131–136.

Evans, J. S. B. and Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3):223–241.

Foster, K. R. and Kokko, H. (2009). The evolution of superstitious and superstition-like behaviour. *Proceedings of the Royal Society B: Biological Sciences*, 276(1654):31–37.

Frankl, V. E. (1985). *Man's Search for Meaning*. Washington Square Press, New York. Original work published 1946.

Friston, K. (2012). The history of the future of the Bayesian brain. *NeuroImage*, 62(2):1230–1233.

Gazzaniga, M. S. (2004). *The Cognitive Neurosciences*. MIT Press, Cambridge, MA, 3rd edition.

Gigerenzer, G. (2008). Rationality for mortals: How people cope with uncertainty. In *Rationality for Mortals*. Oxford University Press, New York.

Gigerenzer, G., Todd, P. M., and ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. Oxford University Press, New York.

Gilbert, D. (2007). *Stumbling on Happiness*. Vintage Books, New York.

Gould, S. J. and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences*, 205(1161):581–598.

Graziano, M. S. and Webb, T. W. (2015). The attention schema theory: A mechanistic account of subjective awareness. *Frontiers in Psychology*, 6:500.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, 7(1):1–16.

Harbaugh, W. T. (1998). What do donations buy? a model of philanthropy based on prestige and warm glow. *Journal of Public Economics*, 67(2):269–284.

Haselton, M. G. and Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78(1):81–91.

Haselton, M. G., Nettle, D., and Andrews, P. W. (2009). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and Social Psychology Review*, 13(1):47–66.

Hayes, S. C., Strosahl, K. D., and Wilson, K. G. (2006). *Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change*. Guilford Press, New York.

Hoffman, D. D., Singh, M., and Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin & Review*, 22(6):1480–1506.

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press, Oxford.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.

Kahneman, D. and Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science*, 12:313–327.

Killingsworth, M. A. and Gilbert, D. T. (2010). A wandering mind is an unhappy mind. *Science*, 330(6006):932–932.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32(2):311–328.

LeDoux, J. E. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon & Schuster, New York.

Li, N. P., van Vugt, M., and Colarelli, S. M. (2018). Evolutionary mismatch and what to do about it: A basic tutorial. *Evolutionary Behavioral Sciences*, 12(3):165–178.

Lieder, F. and Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, 43:e1.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, 12(4):361–366.

Markus, H. R. and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2):224–253.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman, San Francisco.

McAdams, D. P. (2001). The psychology of life stories. *Review of General Psychology*, 5(2):100–122.

Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1):167–202.

Nisbett, R. E., Peng, K., Choi, I., and Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108(2):291–310.

Ochsner, K. N., Silvers, J. A., and Buhle, J. T. (2012). Functional imaging studies of emotion regulation: A synthetic review and evolving model of the cognitive control of emotion. *Annals of the New York Academy of Sciences*, 1251(1):E1–E24.

Parr, T. and Friston, K. J. (2019). Attention or salience? *Current Opinion in Psychology*, 29:1–5.

Quirk, G. J., Garcia, R., and González-Lima, F. (2006). Prefrontal mechanisms in extinction of conditioned fear. *Biological Psychiatry*, 60(4):337–343.

Raichle, M. E. (2015). The brain's default mode network. *Annual Review of Neuroscience*, 38:433–447.

Richerson, P. J. and Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, Chicago.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking, New York.

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, 67(8):603–613.

Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., and Szpunar, K. K. (2012). The future of memory: Remembering, imagining, and the brain. *Neuron*, 76(4):677–694.

Seligman, M. E., Railton, P., Baumeister, R. F., and Sripada, C. (2016). *Homo Prospectus*. Oxford University Press, New York.

Sharot, T. (2011). The optimism bias. *Current Biology*, 21(23):R941–R945.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1):99–118.

Stanovich, K. E. (2004). *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. University of Chicago Press, Chicago.

Suddendorf, T. and Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, 30(3):299–313.

Taylor, S. E. and Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103(2):193–210.

Thaler, R. H. and Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press, New Haven.

Todd, P. M. and Gigerenzer, G. (2012). *Ecological Rationality: Intelligence in the World*. Oxford University Press, New York.

Tononi, G., Boly, M., Massimini, M., and Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, 17(7):450–461.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Varela, F. J. (1999). The specious present: A neurophenomenology of time consciousness. *Naturalizing Phenomenology*, pages 266–314.

Wilson, T. D. and Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, 14(3):131–134.

Zahavi, A. and Zahavi, A. (1997). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford University Press, New York.