The Structural Distortion Principle: A Closed-Loop Model of Perception, Attention, and World-Maintenance in Bounded Cognitive Systems

Boris Kriger
Institute of Integrative and Interdisciplinary Research
boriskriger@interdisciplinary-institute.org

Abstract

Perception is commonly treated as an imperfect but generally reliable representation of external reality. This paper argues that perceptual distortion is not accidental but structural—an inherent property of any adaptive cognitive system operating under finite processing resources. In such systems, attention functions simultaneously as a mechanism of inclusion and a mechanism of exclusion. This dual role produces systematic, self-reinforcing asymmetries between perceived and actual environmental states. The paper introduces the *Structural Distortion Principle* (SDP), formulated as a trade-off between adaptive fitness and representational veridicality: increasing attentional selectivity enhances behavioral viability while proportionally deepening the divergence between the perceived subset and the environment's statistical structure. This is not a claim about perceptual error but about systematic representational compression whose structure is determined by the agent's cultural, moral, and volitional configuration. A conceptual framework is developed connecting perception (P), attention (A), volitional regulation (W), cultural-cognitive filters (C), moral-evaluative constraints (M), and temporal allocation (T), demonstrating how cognitive systems maintain narrow perceptual bandwidths through closed-loop self-reinforcement—the *perceptual configuration cycle*. The construct of *attentional duration* is proposed as an empirically tractable structural signature of the agent's experienced world, operationalizable through fixation metrics, dwell-time paradigms, and sustained-attention tasks. The central paradox is articulated: complete perceptual veridicality is unattainable under resource constraints, yet the explicit recognition of this structural limitation constitutes the primary pathway toward increasing perceptual reliability. The redistribution thesis is developed: perceptual change is achieved not by the addition of information but by the sustained reallocation of attentional duration toward excluded domains, against the automatic constraints of volitional regulation, cultural conditioning, and moral-evaluative prohibition. Four testable predictions are derived that distinguish the framework from predictive processing, cultural psychology, and cognitive bias accounts considered individually. The notation introduced serves as a formal conceptualization that provides the necessary constraints for future computational modeling and identifies specific parameters for empirical falsification. The implications extend beyond individual cognition to any information-processing system—biological or artificial—that must select from an environment richer than its processing capacity.

Keywords: perception, attention, structural distortion, cognitive systems, volitional regulation, cultural filters, attentional duration, adaptive asymmetry, bounded cognition, perceptual configuration cycle

1 Introduction

Simon's (1955) concept of bounded rationality established that cognitive systems operate within resource constraints rather than against an ideal of optimization. The subsequent development of predictive processing (Clark, 2013; Friston, 2010), cultural psychology (Nisbett, 2003), and the cognitive bias research program (Kahneman, 2011) has refined this insight from multiple directions, documenting the constructive, culturally shaped, and systematically skewed character of human perception and reasoning.

However, the dominant framing of perceptual limitation continues to treat distortion as a deficiency—a deviation from an ideal that more information, better instruments, or more careful reasoning might progressively correct. Cognitive biases are catalogued as errors; attentional limitations are treated as failures; cultural conditioning is regarded as interference with otherwise reliable cognitive processes.

This paper proposes a different framing. Perceptual distortion is not a flaw to be corrected but a structural condition of any adaptive system with finite resources. For any cognitive system that must act in a world whose informational complexity exceeds its processing capacity, selective attention is not optional—it is constitutive of perception itself. Because selection entails exclusion, every act of perception is simultaneously an act of non-perception. Structural asymmetry between the perceived and the actual is not the exception; it is the operating condition.

The thesis advanced here is threefold. First, perceptual distortion is structurally inevitable in any resource-bounded cognitive system (the *Structural Distortion Principle*). Second, this distortion is not merely passively suffered but actively maintained by the interplay of attention, volitional regulation, cultural conditioning, moral-evaluative constraints, and temporal allocation. Third, the recognition of structural distortion—not the acquisition of additional information—constitutes the primary pathway toward increased perceptual reliability.

The contribution of this paper is architectural rather than componential. Each element of the argument has precedents in the literature. What has not been systematically formalized is their integration into a single closed-loop framework—the *perceptual configuration cycle*—in which the components mutually constrain and reinforce one another. This integration yields three claims that, in combination, distinguish the present account from existing frameworks:

1. **Attention as a co-primary inclusion/exclusion operator**—not as a secondary consequence of selection but as a structural principle embedded within a systemic architecture.

2. **Temporal allocation as a structural signature of the experienced world**—the distribution of attentional duration across domains as a measurable map of phenomenological boundaries.

3. **The perceptual configuration cycle as a closed-loop architecture**—a self-reinforcing circular relationship among perception, attention, volitional regulation,

culture, morality, temporal allocation, and worldview that provides a unified explanation for the stability of perceptual configurations and identifies specific conditions under which reconfiguration is possible.

The paper's methodological stance is conceptual and integrative rather than computational. The notation introduced in Section 3 provides a formal conceptualization: it specifies relationships between constructs, identifies parameters amenable to empirical operationalization, and establishes the constraints necessary for future computational modeling. Section 5.3 derives four testable predictions that distinguish the framework from existing accounts.

The paper proceeds as follows. Section 2 surveys relevant theoretical traditions. Section 3 introduces the conceptual framework and its components. Section 4 presents the perceptual configuration cycle, identifies disruption points, and develops the redistribution thesis—the claim that perceptual change is achieved through temporal reallocation rather than information addition. Section 5 discusses the framework's relationship to existing models, derives testable predictions, and addresses limitations. Section 6 concludes.

2 Theoretical Background

2.1 Perception as Construction

The constructivist tradition in perception science, stretching from Helmholtz's "unconscious inferences" through Gregory's (1970) constructive theory to contemporary predictive processing frameworks (Clark, 2013; Friston, 2010), has established that perception is not passive reception but active construction. The brain maintains a generative model of its environment and updates this model based on prediction errors—discrepancies between expected and actual sensory input (Hohwy, 2013).

The constructive character of perception is not merely one computational strategy among alternatives but a physical necessity for any system operating under finite neural conduction velocity. Sensory signals travel at bounded speeds (∼100 m/s for the fastest myelinated axons), so by the time afferent information reaches decision-making structures, it describes the past, not the present (Wolpert & Ghahramani, 2000; Wolpert & Flanagan, 2001). In punishing, non-stationary environments, systems that merely react to already-received signals face exponentially decaying survival probability; persistence requires internal states that encode predictive information about future environmental states beyond what current observations provide (Pezzulo et al., 2022). Forward models in motor control were developed precisely to compensate for these sensorimotor delays—an independent convergence from experimental neuroscience on the same architectural necessity that predictive processing derives from computational principles (Wolpert & Ghahramani, 2000; Friston, 2010). For an extended treatment of the physical constraint argument, see Kriger (2026b).

This physical grounding has a direct consequence for the present framework. If constructive perception is not optional but physically mandated, the question shifts from *whether* perception constructs to *what determines the form of the construction*. This is the question the present paper foregrounds. Predictive processing provides the computational machinery; the Structural Distortion Principle addresses what configures that machinery—specifically, why certain perceptual hypotheses are maintained even when prediction errors accumulate, and why the system resists updating in domains where revision would be informationally beneficial but psychologically or energetically costly.

2.2 Attention as Resource and Constraint

Research on attention has produced two broad classes of theory. Resource theories (Kahneman, 1973; Lavie, 2005) treat attention as a limited computational resource that must be allocated across competing demands. Selection theories (Broadbent, 1958; Treisman, 1964; Desimone & Duncan, 1995) emphasize attention's role in filtering relevant from irrelevant information.

Both traditions treat attention primarily as a tool of cognition—a mechanism that serves perception by directing processing resources toward task-relevant stimuli. What is less commonly examined is attention's complementary function: its role as a mechanism of exclusion. Every act of selective attention is simultaneously an act of selective inattention (Mack & Rock, 1998). The phenomena of inattentional blindness (Simons & Chabris, 1999) and change blindness (Rensink et al., 1997) demonstrate that unattended stimuli can be effectively invisible even when they fall within the sensory field.

This paper proposes that attention's exclusionary function is not a secondary consequence of its selective function but is co-primary with it. This claim sharpens an insight already present in the inattentional blindness literature and in taxonomic treatments of attention (Chun et al., 2011), as well as in De Brigard's (2012) analysis of the relationship between attention and awareness. The present framework's contribution is the integration of this dual function into a broader systemic architecture connecting attention to volitional regulation, culture, morality, and temporal allocation.

2.3 Volitional Regulation and Cognitive Stabilization

The role of volition in perceptual regulation has received comparatively little systematic treatment relative to its importance. Executive function research addresses voluntary control of attention (Posner & Rothbart, 2007), and the literature on self-regulation examines how individuals direct cognitive resources toward goals (Baumeister & Vohs, 2004). However, these frameworks generally assume that volition operates to expand or improve cognitive processing—to overcome distraction, sustain focus, and achieve more accurate representations.

This paper advances a counterintuitive proposition: volitional regulation, in its typical deployment, tends to stabilize existing perceptual configurations rather than to expand them. Several converging lines of evidence support this claim. Research on cognitive effort avoidance demonstrates that agents systematically prefer less effortful cognitive strategies even when more effortful strategies would yield better outcomes (Kool et al., 2010). The opportunity-cost model of subjective effort (Kurzban et al., 2013) proposes that the phenomenology of mental effort reflects a cost-benefit computation in which resource reallocation is resisted unless expected returns exceed the costs of reconfiguration. Studies of motivated reasoning show that agents deploy cognitive resources to defend existing beliefs rather than to evaluate evidence impartially (Kunda, 1990; Mercier & Sperber, 2011). Work on cognitive entrenchment demonstrates that expertise can narrow the range of considered alternatives (Dane, 2010). Research on the neurocognitive mechanisms of mental action suggests that volitional control is itself a resource-constrained process that defaults to conservation (Mylopoulos & Lau, 2020).

Collectively, these findings support the proposition that the default mode of volitional operation is stabilization rather than expansion. The volitional system, which possesses the capacity to redirect attention toward novel, complex, or uncomfortable stimuli, in practice most frequently operates to maintain attentional patterns that are familiar, comfortable, and energetically economical.

2.4 Culture and the Boundaries of the Perceivable

Cross-cultural research on perception and cognition has demonstrated that cultural context shapes not only interpretation but perception itself. The Müller-Lyer illusion shows differential susceptibility across cultures (Segall et al., 1966). Linguistic relativity research indicates that language categories influence perceptual discrimination (Winawer et al., 2007). Nisbett & Miyamoto (2005) have shown that East Asian and Western perceivers differ in attentional allocation between focal objects and contextual backgrounds.

This paper extends these findings: culture does not merely influence how things are perceived; it determines the boundaries of what can be sustained in attention. Cultural conditioning specifies which domains of experience are legitimate objects of prolonged contemplation and which are to be rapidly categorized and dismissed. In this sense, culture functions as a pre-attentional filter that constrains the space of possible perceptions before conscious selection begins.

The relationship between culture and attention is not unidirectional. Culture constrains attention, but attentional practices—what a community collectively sustains its engagement upon—also reproduce and modify culture over generational timescales. Bourdieu's (1977) concept of *habitus* captures this bidirectionality: culture is an embodied, structuring structure that shapes perception and attention before conscious thought, yet is itself reproduced through the perceptual and attentional practices it generates. Sperber's (1996) epidemiology of representations provides a complementary mechanism: cultural representations are transmitted and transformed through repeated individual-level cognitive processing, so that the collective attentional patterns of a community gradually reshape the cultural content that constrains those patterns. The present framework treats this bidirectionality as a constitutive feature of the closed-loop architecture described in Section 4.

2.5 Moral-Evaluative Constraints on Attention

A less commonly formalized influence on perceptual configuration is the moral-evaluative framework of the perceiving agent. Research on sacred values demonstrates that moral commitments render certain trade-offs and certain lines of inquiry cognitively inadmissible—not merely undesirable but psychologically inaccessible to sustained deliberation (Tetlock, 2003; Tetlock et al., 2000). Moral foundations theory (Graham et al., 2009; Haidt, 2012) has shown that moral intuitions operate as rapid, affect-laden judgments that precede and constrain subsequent reasoning. Work on taboo cognition (Tetlock, 2003) indicates that moral frameworks determine not only what conclusions are acceptable but what questions may be entertained at all.

These findings suggest that moral-evaluative commitments function as high-level priors that regulate attentional access: they determine whether a given domain of experience is admissible for sustained attention or must be rapidly categorized and dismissed. In the framework developed here, morality (M) operates as a distinct constraint on temporal allocation (T), specifying not only what the agent values but what the agent is permitted—by its own evaluative architecture—to contemplate at length.

2.6 Relation to Predictive Processing and Active Inference

The framework presented here shares significant territory with predictive processing (PP) and active inference (AI). Under PP, perception is hypothesis-driven, and precision-weighting determines which prediction errors are treated as informative (Hohwy, 2013; Clark, 2013). Under active inference, agents do not merely passively receive sensory data but actively select the data they sample, adjusting their sensory engagements to minimize expected free energy (Friston et al., 2015).

The present framework does not compete with PP/AI but operates at a different level of description. PP/AI provides a computational mechanism for perceptual construction and updating. The present framework addresses a structural question that PP/AI leaves largely implicit: why, given the precision-weighting and active sampling mechanisms available, do cognitive systems nevertheless maintain stable perceptual configurations that resist revision even when prediction errors in excluded domains would, if attended to, warrant model updating?

The answer proposed here is that the interaction of W, C, and M constrains the precision-weighting process itself—determining which domains are assigned high precision (and thus generate informative errors) and which are assigned low precision (and thus remain effectively invisible to the updating process). Critically, this is not merely a matter of minimizing surprise (as the free energy principle would have it) but of actively suppressing the detection of surprise: the cultural and moral constraints narrow the bandwidth of what registers as a prediction error in the first place. Under the SDP account, agents do not merely minimize error within a given error-detection regime; they narrow the regime itself. Constant et al. (2019) have begun to address cultural affordances within the PP framework; the present framework extends this line of inquiry by formalizing cultural and moral constraints as upstream regulators of the attention function rather than as content within the generative model.

The SDP can thus be understood as a meta-level principle about the boundary conditions within which PP/AI operates: given empirically plausible priors and constraints, PP/AI implementations are expected to realize the trade-off described by SDP, converging on perceptual configurations that are adaptive but structurally non-veridical, and that are stabilized by the very mechanisms that could in principle revise them.

The SDP also addresses a challenge that predictive processing itself acknowledges but does not resolve: the relevance problem. Among the infinite possible predictions a system could generate, which are worth making? Precision-weighting—the dynamic adjustment of how much weight different predictions and prediction errors receive—provides the computational mechanism (Feldman & Friston, 2010; Hohwy, 2020), but the question of what *sets* the precision weights remains. The present framework proposes that C, M, and W constitute the upstream regulators that determine precision allocation across domains: culture specifies which environmental features are legitimate targets of predictive engagement, morality determines which predictions are admissible for sustained processing, and volition either stabilizes or disrupts the resulting precision structure. This asymmetric precision allocation is itself adaptive: error management theory has established that when the costs of different error types are asymmetric, selection favors biases toward the less costly error rather than toward accuracy per se (Haselton & Buss, 2000; Haselton et al., 2009). High precision on survival-relevant errors ensures rapid response to threats; low precision on abstract inconsistencies permits action despite logical contradictions (Lieder & Griffiths, 2020). The adaptive inconsistency that results is not a deficiency but a structural consequence of resource constraints on verification—the very inconsistency that the SDP formalizes at the perceptual level. For extended treatment of the resource-theoretic argument, see Kriger (2022).

The constraint on precision-weighting has a formal consequence that sharpens the SDP's core claim. Within Bayesian epistemology, it is a standard result that if the agent's prior assigns zero weight to a hypothesis, no amount of evidence can lead to posterior belief in that hypothesis. Applied to the present framework: if the joint operation of C, M, and W effectively excludes a domain from the agent's hypothesis space—if the cultural and moral constraints set the prior weight for that domain to zero—then prediction errors arising from that domain cannot be registered regardless of their magnitude or frequency. This is the formal expression of the SDP's claim that agents do not merely minimize error within a given error-detection regime but narrow the regime itself.

3 Conceptual Framework

3.1 Definitions and Methodological Note

The following constructs are defined for use within the conceptual framework. The notation provides a formal conceptualization: it specifies relationships between constructs, identifies the points at which empirical operationalization is possible, and establishes the constraints necessary for future computational modeling. It does not constitute a mathematical model that yields quantitative derivations in its present form.

- **R**—The reference environment: the total set of states, features, and regularities present in the environment that are in principle detectable. R is treated as a theoretical limit—the maximal information available to an ideal, unconstrained observer. It is not assumed to be metaphysically determinate; it functions as a regulative ideal against which the selectivity of actual perception can be assessed. Importantly, the SDP is invariant under metaphysical stance: R can be treated instrumentally, as the envelope of constraints within which successful prediction and action are evaluated, so that the trade-off formulation holds whether one adopts a realist, enactivist, or pragmatist position. This invariance has a deeper ontological grounding: the distinction between R and P maps onto the distinction between undifferentiated substrate and actualized reality. Information theory establishes that information arises only when differences are registered and encoded relative to a receiver (Shannon, 1948; Bateson, 1972); undifferentiated substrate is operationally equivalent to noise for any system that lacks mechanisms of differentiation. The attention function A is itself an act of differentiation in the operative sense—it is the process through which the agent transforms undifferentiated environmental variation into structured perceptual content (Spencer-Brown, 1969; Luhmann, 1995). Different configurations of C, M, and W produce different differentiations and therefore different actualized perceptual realities from the same substrate—plural actualizations that are legitimate but non-identical, each constrained by both the substrate and the internal organization of the differentiating system (for extended formal treatment, see Kriger, 2026a). In practice, R may be operationally approximated as the set of stimuli detectable by the full range of available instruments and sensory modalities for a given species (cf. von Uexküll's *Umwelt*; von Uexküll, 1934/2010).

- **P** $\subset$ **R**—The perceived configuration: the subset of R that is represented in the cognitive system at any given time.
- **A**—The attention function: the process that maps R onto P through selective inclusion and exclusion.
- **W**—Volitional regulation: the process that governs the deployment, duration, and redirection of A. Following Kurzban et al. (2013), W is understood mechanistically as a cost-benefit computation that defaults to energy conservation and can be recruited for effortful reconfiguration when expected returns exceed reconfiguration costs. W is the only component of the framework with the capacity for endogenous loop-disruption (see Section 4.2), but it is under typical conditions recruited by C and M to function as a stabilizer. This dual capacity—potential disruptor, habitual stabilizer—is what generates the will paradox (Section 3.8).
- **C**—The cultural-cognitive filter: the set of acquired dispositions, categories, norms, and attentional habits that constrain the operation of A. C is bidirectionally coupled with A: culture constrains attention, and collective attentional practices reproduce and modify culture (Bourdieu, 1977; Sperber, 1996).
- **M**—The moral-evaluative constraint: the set of value commitments, sacred-value protections, and moral intuitions that regulate which domains of R are admissible for sustained attentional engagement. M operates on T by determining whether prolonged contemplation of a given domain is permitted or prohibited by the agent's evaluative architecture.
- **T(x)**—Temporal allocation: the duration of sustained attentional engagement with stimulus or domain x.

3.2 The Perception Function

Perception is characterized as a function of attention operating on the reference environment through the joint constraints of culture, volitional regulation, and moral-evaluative frameworks:

$$P = A(R,\ C,\ W,\ M) \tag{1}$$

This formulation captures the claim that P is not a random or merely noisy sample of R but a systematically structured subset determined by the joint operation of attentional selection, cultural filtering, volitional regulation, and moral constraint. Each parameter restricts the domain over which A operates, progressively narrowing the region of R that enters P.

3.3 The Dual Function of Attention

The central property of A in this framework is that it cannot map R onto P without simultaneously generating a complementary set—the unperceived:

$$A:\ R \to P \quad \text{and simultaneously} \quad R \setminus P \tag{2}$$

For any finite cognitive system where processing capacity is strictly less than the informational richness of the environment, the unperceived portion of the environment always vastly exceeds the perceived portion. (The cardinality notation $|P| \ll |R|$ is used here as a didactic simplification; a more appropriate formalization in terms of entropy is given in Section 3.4.) Attention is therefore simultaneously a function of inclusion and exclusion, and the exclusion set is always the larger of the two.

This dual-function claim is not presented as a novel empirical discovery—the inattentional blindness literature (Mack & Rock, 1998; Simons & Chabris, 1999) and taxonomic analyses of attention (Chun et al., 2011) already document the exclusionary consequences of selection. The present contribution is the elevation of this duality from empirical observation to a structural principle within an integrated framework, and its connection to volitional, cultural, and moral determinants of what is excluded. The dual-function principle explains the stability and systematicity of bias, not merely its occurrence: because C, M, and W jointly determine the exclusion set, the structure of what is not perceived is as patterned as the structure of what is perceived.

3.4 The Structural Distortion Principle (SDP)

This is not a claim about perceptual error but about systematic representational compression. The SDP is formulated as a trade-off rather than a tautology. It states not merely that selection entails exclusion (which would be definitional) but that increasing attentional selectivity produces a specific, non-trivial consequence: a systematic divergence between the statistical structure of P and the statistical structure of R, whose form is determined by the agent's configuration of C, M, and W.

SDP. *For any cognitive system S with finite processing resources operating in a reference environment R, there exists a trade-off between adaptive fitness and environment-tracking accuracy. Increasing the selectivity of the attention function A enhances the system's capacity for rapid, contextually appropriate action, but proportionally increases the divergence between the statistical regularities represented in P and those present in R.*

Qualitatively, for any fixed processing capacity, as selectivity increases, expected action-competence (speed, reliability within a subset of tasks) increases while expected representational divergence from R also increases. Different systems—different organisms, different cultures, different artificial architectures—implement different operating points on this trade-off curve. The SDP asserts that no operating point achieves both maximal action-competence and full environment-tracking accuracy; the curve is necessarily concave with respect to these two objectives.

To give the principle empirical traction, selectivity and divergence can be related to measurable quantities. Attentional selectivity can be operationalized as the filtering load in task-based paradigms, the dispersion of fixation distributions in eye-tracking studies, or the sparsity of neural activation patterns during perceptual tasks. Representational divergence can be operationalized as the discrepancy between the statistical distribution of features in the environment (assessed by independent environmental sampling) and the distribution of features reported, recalled, or acted upon by the agent.

An information-theoretic formulation clarifies the relationship. Let $H(R)$ denote the entropy of the reference environment and $H(P)$ the entropy of the perceived configuration.

Selectivity can then be expressed as:

$$\sigma = 1 - \frac{H(P)}{H(R)} \tag{3}$$

As σ increases, $H(P)$ decreases relative to $H(R)$, meaning the perceived configuration becomes more ordered and predictable than would be warranted by the variance present in R. This order is imposed not by the environment but by the culturally and morally shaped suppression of environmental variance: the attention function, constrained by C, M, and W, compresses the environment's statistical structure into a lower-entropy representation that is organized around the agent's adaptive priorities rather than around the environment's objective distribution.

It is important to distinguish this structured compression from adaptive compression. Any efficient coding system reduces entropy relative to its input; the SDP's substantive claim is that the compression introduced by culturally and morally constrained attention is *non-uniform* in a specific way—it preserves variance in domains favored by C and M while suppressing variance in domains excluded by them. The resulting divergence can be expressed as the Kullback-Leibler divergence $D_{\mathrm{KL}}(P \| R)$, which measures the information lost when the compressed representation P is used in place of R. The SDP predicts that $D_{\mathrm{KL}}(P \| R)$ is not merely large (which would follow trivially from compression) but structured—its distribution across domains is patterned by the agent's cultural, moral, and volitional configuration.

A methodological note: the computation of σ or D_{KL} requires an operationalization of $H(R)$, which depends on how R is delimited for a given experimental context. The framework's core predictions concern the direction and patterning of divergence (which domains show greater D_{KL}, and how this patterning varies across agents with different configurations of C, M, and W), not its absolute magnitude. The value of σ will vary with the granularity of R-specification, and this context-dependence should be treated as a feature of experimental design rather than a limitation of the principle.

The SDP is not equivalent to the free energy principle (Friston, 2010), though it is compatible with it. The free energy principle describes a computational strategy (minimize prediction error). The SDP describes a structural consequence of that strategy when implemented under resource constraints with culturally and morally shaped priors: the resulting perceptual configuration is adaptive but systematically non-veridical, and its non-veridicality is patterned by the cultural and moral context of the agent.

3.5 Temporal Allocation as a Structural Signature of the Experienced World

A key proposition of this framework is that the phenomenological salience of any element x in the perceived world is proportional to the duration of attentional engagement:

$$\text{Salience}(x) \propto T(x) \tag{4}$$

This is not a claim about objective importance but about the structure of the agent's experienced world: the world as lived is constituted not by the set of objects the agent can detect but by the set of objects it can sustain attention toward over time. The experienced world is a duration-weighted representation.

This proposition has a strong corollary: the structure of the agent's experienced world is observable through the distribution of temporal allocation. If one maps $T(x)$ across all domains x for a given agent, the resulting distribution reveals the effective boundaries of that agent's phenomenal world—a structural signature, not merely a behavioral correlate.

The claim is empirically tractable. Several existing measurement paradigms provide partial operationalizations of $T(x)$: eye-tracking fixation duration, which measures the time gaze is sustained on a region of interest (Henderson, 2003); dwell time in reading and visual search tasks (Rayner, 1998); experience sampling methods, which probe the content of attentional engagement at random intervals across daily life (Csikszentmihalyi & Larson, 1987); sustained attention and vigilance paradigms, which measure the duration over which focused processing can be maintained (Robertson et al., 1997); and neural measures of sustained engagement, such as EEG steady-state potentials and fMRI BOLD signal persistence during extended stimulus exposure (Langner & Eickhoff, 2013).

The novelty of the present proposal is not the existence of these individual measures but the interpretive frame: $T(x)$ is proposed not as a dependent variable in attention research but as a quantitative bridge between first-person experience and third-person measurement. The distribution of $T(x)$ across domains is treated as an index of what the agent's world *consists of*—a map of phenomenological boundaries that is accessible to empirical investigation.

3.6 Culture as an Attentional Constraint Function

The relationship between culture and attention is characterized as:

$$A = f(C, \ldots) \tag{5}$$

Culture C does not merely provide interpretive frameworks for already-perceived stimuli. It operates upstream of perception by constraining the attention function itself. Cultural conditioning determines which domains are legitimate targets of sustained attention, which phenomena may be briefly registered but not contemplated, and which aspects of the environment are categorically excluded from attentional access.

In this formulation, culture defines not the content of thought but the boundaries of the perceivable world. Two agents with identical sensory apparatus but different cultural filters will inhabit structurally different perceived configurations—not because they interpret the same percepts differently, but because they perceive different subsets of R. This claim extends earlier observations about cultural differences in attentional deployment (Nisbett & Miyamoto, 2005) by treating such differences not as variations in processing style but as variations in the boundaries of the experienced world.

As discussed in Section 2.4, the relationship between C and A is bidirectional: culture constrains what is attended to, and what is collectively attended to reproduces and gradually transforms culture. This bidirectionality is not a complication for the framework but a constitutive feature of the closed-loop architecture described in Section 4.

3.7 The Moral-Evaluative Constraint on Temporal Allocation

The moral-evaluative framework M operates as a high-level regulator of temporal allocation. Where culture (C) determines the general boundaries of the perceivable, morality (M) determines what may be contemplated at length within those boundaries.

Moral commitments categorize domains of experience along an axis of admissibility: some domains are marked as worthy of sustained contemplation (the sacred, the important, the obligatory); others are marked as inadmissible for sustained attention (the taboo, the trivial, the morally repugnant). This categorization directly regulates $T(x)$: domains marked as admissible receive extended attentional engagement, while domains marked as inadmissible are subject to rapid categorization and attentional withdrawal.

Research on sacred values (Tetlock, 2003; Tetlock et al., 2000) demonstrates that moral commitments can render certain factual questions cognitively inadmissible—inaccessible to the sustained deliberation that would be required to evaluate them. This is a structural feature of morally constrained attention, not a failure of information processing.

Formally, M operates as a constraint on T:

$$T(x) \leq T_{\max}(x \mid M) \tag{6}$$

For domains that M marks as inadmissible, $T_{\max}$ approaches zero: the agent cannot sustain attention on these domains regardless of their informational relevance. For domains marked as obligatory, $T_{\max}$ is high and T is maintained even when the domain is informationally sparse or aversive. The moral framework thus introduces a systematic bias into the $T(x)$ distribution that is independent of the informational structure of R.

It is necessary to distinguish M's role from a simpler affective-avoidance account. Affective avoidance (withdrawal from stimuli that elicit negative arousal) operates at the level of stimulus properties. Moral-evaluative constraint operates at the level of domain categories: it renders entire classes of inquiry inadmissible regardless of the affective valence of specific stimuli within those classes. A domain may contain stimuli that are affectively neutral or even pleasant, yet be categorized by M as inadmissible for sustained contemplation (e.g., entertaining trade-offs involving sacred values, even when the trade-off concerns abstract quantities). Conversely, M may require sustained engagement with affectively aversive domains (e.g., moral obligations to witness suffering). This dissociation between affective valence and moral admissibility is what justifies treating M as a distinct architectural component rather than as a subset of affective regulation.

3.8 The Will Paradox

Volitional regulation W possesses the theoretical capacity to override cultural and habitual attentional patterns, thereby expanding P. An agent can, in principle, redirect attention against the grain of C and M.

However, the empirically dominant mode of volitional operation is stabilization rather than expansion. This claim is supported by several converging bodies of evidence: cognitive effort avoidance (Kool et al., 2010), the opportunity-cost structure of mental effort (Kurzban et al., 2013), motivated reasoning (Kunda, 1990; Mercier & Sperber, 2011), cognitive dissonance reduction (Festinger, 1957; Harmon-Jones & Mills, 2019), and cognitive entrenchment (Dane, 2010).

The will paradox is thus not a speculative claim but a redescription, within the present framework, of well-documented regularities in cognitive self-regulation. W is the only component of the cycle with the capacity for endogenous disruption—the capacity to redirect A against the prevailing configuration of C and M. However, under typical conditions, W is recruited by C and M to function as a gatekeeper, deploying volitional resources to maintain familiar attentional patterns rather than to reconfigure them. This dual capacity—potential disruptor, habitual stabilizer—is what makes the will paradox a genuine paradox rather than a simple observation about cognitive laziness. The mechanism that could counteract structural distortion is, under default operating conditions, the mechanism that most reliably reinforces it.

This stabilization is compounded by the reconstructive nature of memory. Research on reconsolidation has established that memory retrieval is not passive readout but active reconstruction: each retrieval renders the accessed trace labile and re-encodes it in a form shaped by the current retrieval context (Nader et al., 2000; Sara, 2000). Schacter et al. (2012) have shown that the same constructive machinery that enables episodic recall also enables imagination and future simulation, implying that memory is generative rather than archival. Crucially, this reconstructive architecture is not a design flaw that evolution failed to correct but the expected outcome of convergent selection pressures: resource costs (maintaining temporal indices for all memory states is metabolically expensive), retrieval speed (associative access is faster than temporal lookup for the content-based queries that dominate adaptive decision-making), and adaptive flexibility (unbound representations permit the recombination that imagination and planning require) all favor architectures in which temporal ordering is constructed during retrieval rather than maintained in storage (Cisek, 2019; Raichle & Gusnard, 2002). For the formal competing-architectures model demonstrating that the fitness advantage of reconstructive architecture grows without bound as system complexity increases, see Kriger (2019). This has a direct consequence for the will paradox: when an agent attempts to evaluate whether its current perceptual configuration is warranted, it must consult its memory of prior evidence—but that memory is itself reconstructed through the current configuration of C, M, and W. The evidence base that might motivate reconfiguration is recalled through the very configuration that would need to be revised. This creates an epistemic double bind that goes beyond the cost-benefit account: volitional reconfiguration is resisted not only because it is effortful (Kurzban et al., 2013) but because the agent's reconstruction of its own evidential history is shaped by the configuration it is trying to escape—and this reconstructive character is itself the product of selection pressures that evolution cannot reverse without sacrificing the speed, flexibility, and resource efficiency that make complex cognition viable. Resource constraints on memory verification—the impossibility of checking reconstructed content against original encoding at retrieval time (Friedman, 1993)—ensure that this circularity is structural rather than contingent. For an extended formal treatment of how resource constraints on verification produce architecturally reconstructive memory systems, see Kriger (2025).

The evolutionary dimension adds a further layer to the will paradox. The costs of full perceptual consistency—exhaustive verification before action, global coherence checking across all belief domains, resolution of all contradictions prior to behavioral commitment—would exceed the costs of adaptive inconsistency in any environment where delayed action is punished and decision spaces exceed verification capacity (Simon, 1955; Gigerenzer et al., 1999). Error management theory has shown that when the costs of different error types are asymmetric, selection favors biases toward the less costly error (Haselton & Buss, 2000; Foster & Kokko, 2009). The SDP's $\sigma > 0$—the structural inevitability of representational divergence—is the perceptual instance of this general principle: the system tolerates distortion because the alternative (paralysis pending complete verification) is more costly than the distortion itself. For extended treatment of the evolutionary architecture of adaptive inconsistency, see Kriger (2021).

This paradox explains why increasing knowledge or information access does not automatically improve perceptual reliability. Information that falls outside the stabilized attentional bandwidth is filtered before it can influence the perceived configuration, and the agent's recollection of prior perceptual states is itself reshaped by each traversal of the cycle.

3.9 Perceptual Dynamics: Path Dependence

Perception at any time $t+1$ is a function not only of current attention, culture, and morality but of prior perception:

$$P_{t+1} = f(P_t,\ A,\ C,\ M) \tag{7}$$

This introduces path dependence into the perceptual system. Past perceptions constrain future perceptions by shaping the attentional patterns, cultural filters, and moral judgments that determine what will be perceived next. Perception is not a series of independent snapshots but a trajectory—a historical process in which each state constrains the space of subsequent states.

This path dependence generates what may be called *perceptual inertia*: the tendency of the cognitive system to maintain its current perceptual trajectory even when environmental changes would warrant revision. Perceptual inertia is the dynamic expression of the will paradox operating over time. Its strength is amplified by the reconstructive character of memory retrieval: because each act of recall rebuilds past content from current context and stored associative patterns rather than reading from a fixed archive (Nader et al., 2000; Schacter et al., 2012), the agent's representation of its own perceptual history is progressively reshaped to cohere with the current configuration. Prior perceptions are not merely filtered through P_t at the time of new input; they are retroactively reconstructed to be consistent with it, deepening the self-reinforcing character of the cycle beyond what a forward-filtering account alone would predict.

3.10 The Finite-System Constraint and Its Constructive Consequence

Constraint Lemma. *For any cognitive system S with finite processing resources operating in a reference environment R where the informational richness of R exceeds the processing capacity of S, $P \neq R$.*

This constraint follows directly from the definitions and is acknowledged as such. It is not presented as a substantive result but as a boundary condition that frames the more interesting question: given that $P \neq R$ is structurally unavoidable, what determines *how* they differ, and what interventions can reduce the divergence in specific, targeted domains?

The constructive consequence is the paper's central practical claim. The explicit recognition that $P \neq R$—the awareness that one's perceived world is necessarily a compressed and patterned subset of the reference environment—creates the precondition for *directed perceptual revision*. This recognition is harder to achieve than it may appear, because the gap between P and R is not merely quantitative but grows structurally with system complexity. As adaptive systems become more complex, the space of action-relevant distinctions expands combinatorially—planning over n future steps with branching factor b requires evaluating $O(b^n)$ behavioral sequences; social reasoning about k agents with m possible mental states generates $O(m^k)$ joint possibilities—while verification bandwidth grows at most linearly (Simon, 1955; Gigerenzer & Selten, 2001). The ratio of verification capacity to decision-space size therefore *decreases* as complexity increases. More information can paradoxically *reduce* effective verification coverage by expanding the space of distinctions faster than it expands the capacity to check them. This is the formal expression of why the accumulation of information is insufficient: adding data enlarges $|R|$ without proportionally enlarging the agent's capacity to verify its representation of R. Directed perceptual revision requires instead modifying the regime through which information is processed.

This recognition is related to the broader capacity for epistemic vigilance—the ability to evaluate the reliability of one's own cognitive processes (Sperber et al., 2010). An agent whose cost-benefit computation (W) registers sufficiently high expected returns from reconfiguration can reallocate temporal engagement toward domains previously excluded by C and M:

$$\Delta\text{Reliability}(x) \propto \Delta T(x_{\text{excluded}}) \tag{8}$$

The increase in perceptual reliability in a given domain is proportional to the increase in sustained attentional engagement with that domain—particularly when that domain has been previously excluded by the joint operation of C, M, and W. This is the formal expression of the claim that the path to greater perceptual reliability runs through the discipline of attention rather than the accumulation of information.

4 The Perceptual Configuration Cycle

4.1 Architecture

The components described above do not operate in linear sequence but form a closed, self-reinforcing loop. Figure 1 presents the cycle schematically.

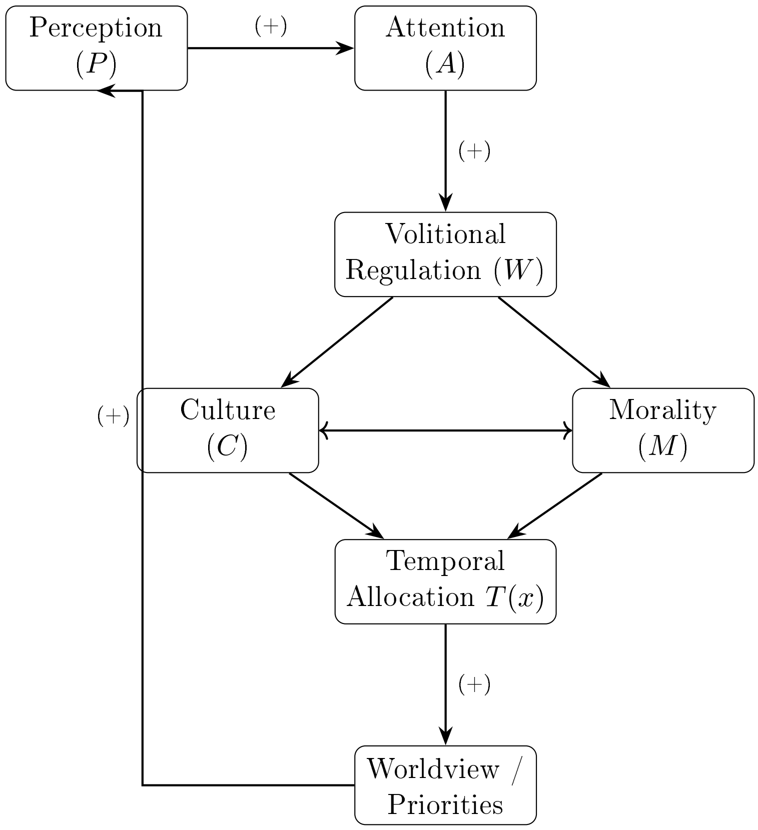

Figure 1: The Perceptual Configuration Cycle. (+) denotes a positive (reinforcing) feedback sign under default operation. The bidirectional arrow between Culture (C) and Morality (M) indicates mutual constraint. W is the primary candidate for endogenous loop-disruption (see Section 4.2).

Each link in the cycle carries a predominantly reinforcing feedback sign under default operation. Perception shapes the attentional patterns that filter future perception. Attention, regulated by W, tends toward stabilization of existing patterns. Cultural and moral constraints narrow the domain of sustained attention. Temporal allocation, once established, reinforces the priorities that generated it. Priorities consolidate the worldview that feeds back into perceptual expectations.

The result is a self-reinforcing cycle that maintains the current perceptual configuration against perturbation. The interaction of W, C, and M stabilizes the current configuration, producing the phenomenon that requires explanation: why individuals, communities, and cultures can maintain systematically asymmetric worldviews over extended periods despite the availability of corrective information.

The self-reinforcing character of the cycle is further amplified by the pathway through memory. The return path from Worldview to Perception passes through the agent's recollection of prior experience, and that recollection is not archival but reconstructive (Nader et al., 2000; Bartlett, 1932; Schacter et al., 2012). Each traversal of the cycle does not merely filter incoming stimuli through the current configuration—it retroactively reshapes the agent's recollection of prior perceptual states to cohere with the current worldview. The cycle is therefore self-reinforcing in both the forward direction (filtering new input) and the backward direction (reconstructing the evidential past). This double reinforcement explains why the cycle is more resistant to disruption than a simple forward-filtering model would predict: even the agent's "evidence base"—its memory of what it has previously seen and concluded—is updated to support the current configuration at each traversal.

4.2 Disruption Points and Perceptual Plasticity

The circularity of the model is both its explanatory strength and its potential vulnerability to the charge of unfalsifiability. If the system maintains distortion because the loop reinforces itself, and we know the loop reinforces itself because distortion is maintained, the account risks tautology.

This concern is addressed by identifying specific disruption points—conditions under which the self-reinforcing cycle can be broken or reconfigured, generating observable changes in P.

Acute prediction error. Events that generate prediction errors of sufficient magnitude to overwhelm the stabilization function of W—such as trauma, sudden environmental change, or immersive cross-cultural contact—can force reconfiguration of A and, consequently, of the downstream components (cf. Friston et al., 2015, on active inference under surprise).

Deliberate attentional training. Contemplative practices such as meditation involve systematic training of W to override default stabilization patterns and sustain attention on domains normally excluded (Lutz et al., 2008). Evidence from meditation research demonstrates measurable changes in attentional distribution, perceptual sensitivity, and neural activation patterns following sustained practice.

Institutional intervention. Educational, therapeutic, and cultural institutions can modify C and M by altering the collective norms governing attentional engagement—expanding the domains considered legitimate for sustained contemplation.

Technological reconfiguration. Instruments, recording devices, and data-visualization tools can extend the temporal window of attentional engagement by making transient or inaccessible features of the environment available for sustained inspection.

The identification of disruption points transforms the closed-loop architecture from a static description into a dynamic model with empirically investigable plasticity conditions.

4.3 The Redistribution Thesis: Perceptual Change Through Temporal Reallocation

The preceding analysis converges on a thesis that constitutes the framework's central practical claim and that warrants explicit, dedicated formulation: *perceptual change is achieved not by the addition of information but by the redistribution of attentional duration across domains, against the automatic constraints imposed by volitional regulation, culture, and morality.*

This claim has three components, each of which draws on the architecture developed above.

4.3.1 Why information addition is insufficient

The conventional assumption in education, therapy, public discourse, and epistemology is that perceptual and cognitive change is produced by providing the agent with new information—facts, arguments, evidence, perspectives. The SDP framework explains why this strategy fails with predictable regularity: new information entering the cognitive system is processed through the existing configuration of C, M, and W. The attention function A, constrained by these parameters, determines whether the new information receives sustained engagement ($T(x) > 0$) or is rapidly categorized and dismissed ($T(x) \approx 0$). Information that falls outside the boundaries set by the current cultural-moral configuration is not merely resisted—it is structurally invisible to the updating mechanism.

This is not a metaphorical claim. The Bayesian formalization in Section 2.6 makes it precise: if the joint operation of C, M, and W sets the prior weight for a domain to zero, no amount of evidence from that domain can shift the posterior. The information exists in R; it may even enter the sensory field; but the precision-weighting regime assigns it zero weight, so it generates no prediction error and triggers no model updating. The agent is not ignoring the information through inattention—the agent's perceptual architecture is configured so that the information does not register as information at all.

Three converging empirical phenomena support this analysis. First, the confirmation bias literature documents that agents exposed to balanced evidence on controversial topics become *more* polarized, not less—each side assimilates the evidence that confirms its existing configuration and discounts the rest (Nickerson, 1998). Second, the "backfire effect" in persuasion research shows that corrections of factual errors can strengthen rather than weaken belief in the errors, because the correction activates the defensive stabilization function of W (Festinger, 1957). Third, the extensive failure of information campaigns in public health, climate communication, and political deradicalization demonstrates that information availability does not translate into perceptual change when the target domains are excluded by the audience's moral-evaluative constraints (Tetlock, 2003).

The SDP provides a unified structural explanation for these phenomena: they are not failures of information delivery but predictable consequences of attempting to change P by modifying R (adding information to the environment) rather than by modifying A (changing how the environment is attended to).

4.3.2 Why temporal redistribution is the operative mechanism

If information addition is insufficient, the question becomes: what intervention *does* produce perceptual change? The framework's answer is temporal redistribution—the sustained reallocation of $T(x)$ from domains that currently dominate attentional engagement toward domains that the current configuration excludes:

$$\Delta P \propto \Delta T(x_{\text{excluded}}) \quad \text{when} \quad \Delta I(x) = 0 \tag{9}$$

where ΔP denotes the change in the perceived configuration, $\Delta T(x_{\text{excluded}})$ denotes the increase in attentional duration directed toward previously excluded domains, and $\Delta I(x) = 0$ specifies that no new information has been provided. The claim is that temporal redistribution alone—holding information constant—is sufficient to produce measurable perceptual change. This is what Prediction 2 (Section 5.3) formalizes as a testable hypothesis.

The mechanism is straightforward within the SDP architecture. Sustained attentional engagement with a domain produces three effects. First, it generates prediction errors in that domain: features that were previously invisible (because $T(x) \approx 0$) now become available for processing, and discrepancies between the agent's implicit model of that domain and the domain's actual structure begin to register. Second, the increased $T(x)$ raises the domain's precision-weighting within the predictive processing hierarchy: the agent's generative model begins to allocate computational resources to the domain, treating its inputs as informative rather than as noise. Third, sustained engagement produces associative integration: the attended domain becomes connected to the agent's broader representational structure, generating cross-domain inferences and coherence pressures that were previously absent.

These three effects—error generation, precision elevation, and associative integration—are the mechanisms through which temporal redistribution modifies P without adding information. The agent does not learn new facts about the domain; rather, the agent begins to *perceive* the domain—to register its structure, generate expectations about it, and integrate it into the broader perceptual configuration.

Evidence for this mechanism comes from multiple experimental traditions. Meditation research demonstrates that sustained attentional training—which provides no new information about the environment—produces measurable changes in perceptual sensitivity, including enhanced detection of brief visual stimuli (Lutz et al., 2008), altered neural response to emotional stimuli, and increased perceptual discrimination in trained domains. Exposure therapy operates by sustaining attentional engagement with feared stimuli without providing information about their safety—the therapeutic mechanism is temporal redistribution, not information delivery. Cross-cultural immersion experiences produce perceptual changes that correlate with duration of sustained engagement rather than with the quantity of factual knowledge acquired about the host culture.

4.3.3 Why redistribution requires overriding W, C, and M

The critical difficulty of the redistribution thesis is that the very systems whose reconfiguration is required—W, C, and M—are the systems that resist the redistribution. This is the will paradox (Section 3.8) restated at the level of intervention design.

Volitional regulation (W) resists temporal redistribution because it is costly. The opportunity-cost model (Kurzban et al., 2013) implies that redirecting attentional resources toward excluded domains withdraws those resources from domains that currently generate adaptive returns. The agent must expend effort to sustain engagement with domains that, by definition, its cost-benefit computation has assessed as not worth attending to. W therefore generates subjective experiences of difficulty, boredom, aversion, and fatigue that function as signals to terminate the redistribution and restore the default allocation.

Cultural conditioning (C) resists temporal redistribution because sustained engagement with excluded domains violates collective norms. Culture specifies not only what is worth attending to but what is worth attending to *together*—what constitutes legitimate shared attention. An agent who redistributes $T(x)$ toward domains that its cultural community excludes faces social costs: incomprehension, marginalization, accusations of wasting time on trivial or inappropriate matters. C thus provides external reinforcement for the internal resistance generated by W. The embodied habitus (Bourdieu, 1977) makes certain attentional patterns feel natural and others feel forced, ensuring that cultural resistance to redistribution operates below the threshold of conscious reflection.

Moral-evaluative constraints (M) resist temporal redistribution most powerfully because they render the excluded domains not merely uninteresting but *inadmissible*. To sustain attention on a domain that M marks as taboo is experienced not as effortful (as W-resistance would produce) or as socially deviant (as C-resistance would produce) but as morally wrong—a violation of the agent's own evaluative integrity. Sacred-value protection (Tetlock, 2003) generates moral outrage at the very suggestion that certain domains deserve sustained contemplation. The agent who attempts temporal redistribution toward morally excluded domains must therefore overcome not only effort costs and social costs but identity costs—the perception that sustained attention to the excluded domain constitutes a betrayal of deeply held values.

The triple resistance of W, C, and M explains a persistent empirical puzzle: why perceptual change is difficult even when agents sincerely desire it, are provided with all relevant information, and have access to supportive institutional structures. The SDP framework locates the difficulty not in the agent's knowledge, motivation, or circumstances, but in the structural properties of the perceptual configuration cycle itself. The cycle is designed—by evolutionary selection for adaptive efficiency—to resist precisely the kind of redistribution that perceptual change requires.

4.3.4 Implications for intervention design

This analysis has direct consequences for the design of interventions aimed at perceptual change—in education, psychotherapy, organizational development, conflict resolution, and public communication.

First, interventions should be evaluated not by the quantity of information they deliver but by the magnitude and duration of $T(x)$ redistribution they produce. A lecture that provides extensive factual information but does not alter the audience's attentional allocation is predicted to be ineffective. A guided experience that provides no new information but sustains the participant's attention on previously excluded domains is predicted to be more effective.

Second, effective interventions must address all three sources of resistance. Interventions that reduce only the effort cost of redistribution (W-targeted: making sustained engagement easier through scaffolding, gamification, or structured practice) will encounter C- and M-resistance. Interventions that modify only cultural norms (C-targeted: creating communities of practice where the excluded domain is treated as legitimate) will encounter W- and M-resistance. Interventions that address only moral framing (M-targeted: reframing the excluded domain as morally admissible or obligatory) will encounter W- and C-resistance. The closed-loop architecture predicts that sustainable perceptual change requires coordinated intervention across all three constraint systems.

Third, the reconstructive nature of memory (Section 3.8) implies that successful temporal redistribution should produce not only changes in future perception but *retroactive* changes in the agent's reconstruction of its own perceptual history. Agents who achieve sustained redistribution should, upon retrospective report, reconstruct their prior perceptual experience as having been more consistent with the new configuration than it actually was. This retroactive reconstruction is not a confound in intervention evaluation but a signature of successful cycle reconfiguration: the new perceptual configuration, once established, reshapes the evidential past to support itself, just as the old configuration did. Intervention research should measure pre-and-post changes in autobiographical recall content as an index of the depth of perceptual reconfiguration achieved.

The redistribution thesis thus transforms the SDP from a descriptive framework into a prescriptive one: it specifies not only why perceptual distortion is maintained but what class of interventions is required to modify it, and it provides measurable criteria ($\Delta T(x)$

across excluded domains) for evaluating intervention effectiveness.

5 Discussion and Empirical Program

5.1 Relation to Existing Frameworks

The Structural Distortion Principle and the perceptual configuration cycle intersect with, but are not reducible to, several existing theoretical frameworks.

Phenomenology. Husserl's (1913/1982) concept of intentionality is compatible with the present framework's treatment of attention as constitutive of perception. Merleau-Ponty's (1945/2012) embodied phenomenology aligns with the claim that perception is not passive reception. The present framework extends these traditions by embedding the phenomenological insight within a formal architecture that specifies the self-reinforcing dynamics of perceptual configuration.

Cognitive bias research. Kahneman's (2011) dual-process framework and the extensive catalogue of cognitive biases provide empirical support for systematic perceptual distortion. The bias literature generally treats individual biases as discrete errors amenable to targeted correction. The SDP account reframes these biases as surface manifestations of a single underlying structural condition: the inevitable representational compression produced by selective attention under resource constraints, stabilized by volitional, cultural, and moral regulation.

Predictive processing and active inference. As discussed in Section 2.6, the SDP account does not compete with PP/AI but operates at a different level of description. PP/AI specifies computational mechanisms for perceptual updating; the SDP specifies structural boundary conditions that constrain how those mechanisms are deployed. The key distinction is that active inference models the minimization of prediction error within a given precision-weighting regime, while the SDP addresses how C, M, and W constrain the precision-weighting regime itself—determining what counts as an error rather than how errors are reduced.

Autopoiesis, enactivism, and cybernetics. The perceptual configuration cycle shares structural affinities with autopoietic theory (Maturana & Varela, 1980) and second-order cybernetics (von Foerster, 1981), particularly in its emphasis on recursive self-maintenance and operational closure. The enactivist tradition (Varela et al., 1991; Thompson, 2007) would likely challenge the present framework's retention of R as a reference environment, since enactivism holds that the relevant environment is co-constituted by the organism rather than independently given. The present framework accommodates this concern by treating R instrumentally (see Section 3): the trade-off described by the SDP holds regardless of whether R is understood as an observer-independent reality or as the envelope of constraints within which the organism's sense-making is evaluated. The formal structure of the argument is preserved under either interpretation. Indeed, the SDP framework is not merely *compatible* with the enactivist insight but *instantiates* it at the level of attention: the attention function A, as described in Section 3.3, is the cognitive system's primary act of differentiation—the operation through which undifferentiated environmental variation becomes structured perceptual content (Spencer-Brown, 1969; Bateson, 1972). Different configurations of C, M, and W produce different differentiations and therefore different actualized realities from the same substrate. This is the perceptual instantiation of the general ontological principle that coherent systems achieve determinate reality only through internal acts of differentiation (Luhmann, 1995;

Simondon, 1958/2020). Prediction 1 (cross-cultural $T(x)$ divergence) is thus not merely a claim about different *filters* on a shared reality but about different *actualizations* of structured perceptual worlds from a shared but undifferentiated substrate.

Bounded rationality. Simon's (1955) foundational analysis established that cognitive systems operate under resource constraints. The present framework extends Simon's analysis by specifying the mechanisms through which boundedness becomes self-reinforcing: it is not merely that the agent cannot process everything, but that the interaction of W, C, and M stabilizes the boundaries of processing against revision. The same resource-constraint logic applies to memory architecture: comprehensive temporal indexing of all memory states faces scaling difficulties as system complexity increases (Friedman, 1993; Howard & Kahana, 2002), which is why temporal information in memory tends to be encoded as reconstructible content (contextual cues, causal inferences) rather than maintained as structural metadata—a further instance of bounded resources producing systematic representational compression rather than mere quantitative limitation (Kriger, 2025). The SDP's trade-off between adaptive fitness and representational accuracy is itself a special case of a more general principle: when the costs of different error types are asymmetric, selection favors biases toward the less costly error (Haselton & Buss, 2000; Haselton et al., 2009). Under conditions where the cost of missed threats exceeds the cost of false alarms, or where the cost of delayed action exceeds the cost of acting on incomplete information, selection produces systems that are systematically biased rather than maximally accurate (Foster & Kokko, 2009). The SDP's $\sigma > 0$ is the perceptual expression of this general error-management principle, and the resource-rational analysis program (Lieder & Griffiths, 2020) provides converging evidence that many apparent biases are optimal given computational constraints—not deviations from rationality but instances of a different rationality, one calibrated to resource costs rather than to truth-tracking.

Global workspace theory. Baars' (1988) global workspace theory proposes that conscious access depends on information entering a global broadcasting network. The SDP account addresses what determines which information gains access to the workspace and which is excluded—a question that global workspace theory acknowledges but does not resolve at the level of cultural and moral constraints.

5.2 Novel Contributions

The individual components of this framework—bounded attention, cultural influence on perception, cognitive bias, the constructive nature of perception, motivated reasoning, sacred values—have extensive prior treatment. The contribution claimed here is architectural. The integration of these elements into a single closed-loop framework generates insights that do not emerge from any component considered in isolation.

This paper contributes three claims that, in their integration, constitute the framework's novelty.

1. **Attention as a co-primary inclusion/exclusion operator.** The dual function of attention is elevated from an empirical observation (inattentional blindness) to a structural principle embedded within a systemic architecture. This explains the stability and systematicity of perceptual bias—not merely its occurrence but its patterned structure.

2. **Temporal allocation as a structural signature of the experienced world.**

The distribution of $T(x)$ across domains provides a quantitative bridge between first-person experience and third-person measurement—an empirically measurable index of the effective boundaries of an agent's phenomenal world.

3. **The perceptual configuration cycle as a closed-loop architecture.** The self-reinforcing circular relationship among perception, attention, volitional regulation, culture, morality, temporal allocation, and worldview provides a unified, mechanistic account for phenomena often treated separately—cognitive entrenchment, worldview defense, sacred-value protection, cultural persistence of belief—and identifies specific conditions under which reconfiguration is possible.

5.3 Testable Predictions

The framework generates the following testable predictions. Table 1 summarizes where these predictions overlap with and diverge from existing frameworks.

Table 1: Comparative Prediction Matrix

Prediction	PP/AI	Cultural Psych.	Bias Research	SDP
1. Cross-cultural $T(x)$ divergence for culturally *novel* stimuli	No	Partially	No	Yes
2. $T(x)$ redistribution without information addition changes P	Not specific	Not specific	No	Yes
3. Moral ceilings on $T(x)$ dissociate from affective valence	No	Not specific	No	Yes
4. Moral shift produces $T(x)$ changes in non-moral domains	No	No	No	Yes

Prediction 1: Cross-cultural divergence in $T(x)$ distributions for culturally novel stimuli. If culture functions as a constraint on the attention function itself (not merely on the interpretation of already-perceived stimuli), then agents from different cultural backgrounds should show measurable differences in $T(x)$ distributions when exposed to identical stimulus arrays—and these differences should predict differences in subsequent perceptual reports, not merely differences in verbal interpretation. Critically, the SDP predicts that this divergence should persist even for stimuli that are culturally novel to both groups (where perceptual learning histories do not differ), because the cultural filter operates on domain categories rather than on specific stimulus familiarity. This condition distinguishes the SDP's structural account from the simpler alternative that cross-cultural differences in $T(x)$ merely reflect different perceptual learning histories.

Prediction 2: Attentional training modifies P through $T(x)$ redistribution, not through information addition. If the constructive consequence of the framework is correct, then interventions that increase $T(x)$ for previously excluded domains (e.g., sustained attention training, meditation, guided exposure) should produce measurable changes in perceptual sensitivity and subsequent perceptual reports—even when no new factual information is provided. This is the most tightly derived prediction: if the key variable is temporal allocation rather than information quantity, then equating information while varying only $T(x)$ should suffice to produce perceptual change. The reconstructive nature of memory (Nader et al., 2000; Schacter et al., 2012) implies a further observable consequence: attentional redistribution should alter not only future perceptual sensitivity but also the agent's reconstruction of prior perceptual experience, producing retroactive changes in the reported evidence base that are measurable through pre-and-post comparisons of autobiographical recall content.

Prediction 3: Moral-evaluative constraints impose ceilings on $T(x)$ that dissociate from affective avoidance. If M constrains $T(x)$ as a distinct architectural component (rather than as a subset of affective regulation), then domains marked as morally inadmissible should show truncated attentional engagement even when the specific stimuli within those domains are affectively neutral or positive. Conversely, domains marked as morally obligatory should sustain elevated $T(x)$ even when the specific stimuli are affectively aversive. The experimental design should manipulate moral framing of informationally and affectively equivalent stimuli and measure $T(x)$ under incentivized conditions, controlling for arousal and valence.

Prediction 4: Disruption of the moral component (M) produces downstream changes in $T(x)$ for non-morally-loaded domains. If the perceptual configuration cycle operates as a genuine closed loop rather than as a set of independent parallel constraints, then a shift in moral commitments should produce measurable changes in $T(x)$ for domains far removed from the moral domain itself—because the reconfiguration of M propagates through the cycle to alter C, W, and consequently the overall $T(x)$ distribution. This prediction is distinctive to the full-loop architecture; simpler, single-mechanism accounts (habit formation, Bayesian prior strengthening, neural consolidation) do not predict downstream propagation from moral reconfiguration to perceptual engagement in unrelated domains. This can be tested by measuring $T(x)$ distributions before and after moral-commitment interventions (e.g., perspective-taking exercises, immersive ethical dilemmas) for both morally loaded and morally neutral stimulus domains.

5.4 Implications for Epistemology

The framework has implications for epistemological practice. If perceptual distortion is structural rather than accidental, then the conventional epistemic strategy of accumulating more information is insufficient. An agent that acquires additional data but processes it through the same attentional configuration will assimilate new information into existing perceptual patterns—a process documented in the confirmation bias literature (Nickerson, 1998) but here given a structural explanation as a consequence of the self-reinforcing perceptual configuration cycle.

The alternative epistemic strategy suggested by the framework is *attentional discipline*: the deliberate, sustained redirection of attention toward domains that the default configuration excludes. This is not a matter of acquiring more information but of modifying the regime through which information is processed—and, specifically, of increasing $T(x)$ for domains where the default $T(x)$ is near zero. In epistemic terms, attentional discipline is a *meta-credal* operation: it involves revising not the content of beliefs but the prior commitments about what is worth attending to—the framework within which learning occurs rather than the learning itself. This is why attentional discipline is harder than ordinary learning and why the will paradox resists it: ordinary learning operates within existing credal commitments (within the current precision-weighting regime), while attentional discipline requires modifying the framework that determines what counts as evidence in the first place (Mercier & Sperber, 2011; Kunda, 1990). The agent is not merely updating beliefs given evidence but revising the structure that determines what registers as evidence—a second-order operation that the cost-benefit logic of W systematically resists unless disruption conditions (Section 4.2) shift the computation.

5.5 Implications for Artificial Cognitive Systems

The Structural Distortion Principle applies, by its formulation, to any cognitive system with finite resources, including artificial ones. In biological systems, attentional selectivity is shaped by culture, morality, and volitional regulation; in artificial systems, functionally analogous roles are played by training data distributions (analogous to C), objective functions and reward signals (analogous to M), and optimization procedures (analogous to W).

This mapping is offered as a structural analogy, not an identity claim. Attention in transformer architectures is a specific computational operation (scaled dot-product similarity over key-query-value triples) that differs from biological attention in important respects. Nevertheless, the SDP's core claim—that any finite system performing selective processing will produce a systematically non-veridical representation shaped by its selection criteria—applies at a level of abstraction that encompasses both biological and artificial systems.

To develop one implication concretely: adversarial examples in machine learning (Goodfellow et al., 2015) can be interpreted through the SDP as exploitations of the system's structurally necessary exclusion set. A neural network trained on a given data distribution develops feature weightings—learned attention—that are highly selective along certain input dimensions and nearly null along others. Adversarial perturbations succeed precisely because they target directions in input space that fall within the system's exclusion set: they modify features that are nearly invisible to the system's learned attention, producing inputs that are perceptually identical to the system but categorically different to an observer with a broader attentional distribution. In SDP vocabulary, the adversarial perturbation exploits the high σ of the system's learned representation: the system's selectivity, which makes it efficient in its normal operating regime, creates a structured vulnerability in exactly the dimensions it has learned to suppress. This reframing does not replace existing analyses of adversarial robustness, but it connects the phenomenon to a broader structural principle and suggests that robustness improvements may require deliberate redistribution of attentional resources toward suppressed dimensions—an approach compatible with recent work on robust feature learning.

More broadly, the problem of value alignment can be reframed via the SDP as the challenge of installing an appropriate M-analogue to prevent an artificial system's attentional selection from systematically excluding the very environmental features—long-tail risks, subtle distributional shifts, underrepresented human values—that are most consequential for safe and beneficial operation. The convergence of biological and artificial systems on architectures with structurally similar properties—selective compression, reconstructive generativity, hallucination-like confabulation—is itself predicted by the SDP framework: any system complex enough to generalize must operate under a verification-complexity trade-off in which verification coverage decreases as decision-space complexity increases. Hallucination in generative AI, like false memory in biology, is the cost of an architecture that enables generalization and recombination—the same architecture that selection (evolutionary or market) converges on because the alternative (exhaustive verification before output) does not scale (Schacter et al., 2012; Wolpert & Ghahramani, 2000). The appropriate response is not elimination of the generative capacity but calibration—knowing when the architecture's outputs are reliable, which requires precisely the kind of meta-level awareness that the SDP's "recognition of structural limitation" prescribes for biological agents.

5.6 Limitations

This framework is conceptual and integrative. Its notation provides a formal conceptualization that specifies relationships and identifies parameters for empirical falsification, but does not permit simulation or quantitative prediction in its present form. Future work should pursue either a computational implementation (e.g., a toy agent in an information-selection task that demonstrates the trade-off curve posited by the SDP) or a program of empirical testing that operationalizes the constructs through the measurement paradigms identified in Section 3.5. A promising direction for computational modeling would be to formalize the perceptual configuration cycle as a fixed-point problem: if the cycle converges on a stable worldview, that worldview corresponds to a fixed point of the joint perception-attention-memory-retrieval operator. Classical fixed-point results (Banach contraction, Knaster-Tarski) provide sufficient conditions for the existence and uniqueness of such equilibria, and disruption conditions (Section 4.2) would correspond to perturbations that shift the system to a different basin of attraction. The formal treatment of memory attractors as fixed points of retrieval-reconstruction operators in Kriger (2025) offers a starting point for this computational program. A complementary formalization would model the SDP's trade-off curve using a competing-architectures framework analogous to that developed for memory systems (Kriger, 2019): defining fitness components for resource cost, retrieval speed, and representational flexibility, and demonstrating formally that the fitness advantage of compressed (structurally distorted) perception grows without bound as system complexity increases, while the advantage of veridical representation is bounded—yielding a complexity threshold above which structural distortion is the evolutionarily stable strategy.

The framework abstracts away from the rich internal structure of each component. Attention, volitional regulation, culture, and morality are each complex, multidimensional constructs that resist simple functional characterization. The present framework trades internal complexity for systemic architecture—a trade-off appropriate for a first formulation but requiring refinement.

The scope of the framework—spanning phenomenology, cognitive science, cultural psychology, self-regulation, moral psychology, and AI—invites breadth but risks diffuseness. While the breadth is intentional, as it reveals systemic relationships that narrower treatments miss, future work may profitably narrow the scope—for example, by concentrating on moral constraints on attention in politically polarized domains, or on attentional discipline as an epistemic virtue in scientific practice.

The information-theoretic formulation of σ and D_{KL} requires an operationalization of $H(R)$ that depends on experimental context. As noted in Section 3.4, the framework's predictions concern the patterning of divergence rather than its absolute magnitude, and R-specification should be treated as a design parameter in empirical tests.

Finally, the construct of W oscillates between a mechanistic reading (cost-benefit computation, per Kurzban et al., 2013) and the language of agentive deliberation ("directed perceptual revision," "discipline of attention"). The framework commits to the mechanistic reading: what is described as "recognition" of structural distortion and "deliberate" reconfiguration is understood as a regime shift in the cost-benefit computation—a state in which reconfiguration costs are assessed as justified by expected returns, typically triggered by one of the disruption conditions identified in Section 4.2. The phenomenology of deliberation accompanies this regime shift but does not constitute a separate causal mechanism.

6 Conclusion

Perceptual distortion is not a defect of cognition but a structural condition of adaptive systems operating under finite resource constraints. Any system that must act in an environment richer than its processing capacity must select, and selection entails exclusion. The excluded always exceeds the included.

The Structural Distortion Principle formalizes this condition and embeds it within the perceptual configuration cycle—a closed-loop architecture in which perception, attention, volitional regulation, culture, morality, temporal allocation, and worldview form a self-reinforcing system. The interaction of W, C, and M stabilizes the current perceptual configuration against revision. The convergent evidence from cognitive effort avoidance, motivated reasoning, cognitive dissonance, and sacred-value protection supports the will paradox: the mechanism that could counteract structural distortion is, under default operating conditions, the mechanism that most reliably reinforces it.

The redistribution thesis sharpens this analysis into a prescriptive claim: perceptual change is achieved not by adding information to the agent's environment but by redistributing attentional duration toward domains that the current configuration of W, C, and M is structured to exclude. This redistribution must overcome the triple resistance of volitional effort costs, cultural norm enforcement, and moral-evaluative prohibition—and the reconstructive nature of memory ensures that the agent's own evidential history is reshaped by the current configuration at each cycle traversal, deepening the difficulty of the task.

The central paradox is also the central opportunity. Complete veridicality is structurally unattainable, but the recognition of structural limitation is the precondition for directed revision. The path to greater perceptual reliability runs not through the accumulation of information but through the discipline of attention—the sustained reallocation of temporal engagement toward domains that the default perceptual configuration is structured to exclude.

A Practical Protocol for Attentional Reconfiguration

The following protocol operationalizes the redistribution thesis (Section 4.3) as a structured procedure for cognitive self-intervention. It is intended for any agent capable of reflective self-monitoring and follows directly from the architecture of the perceptual configuration cycle (Section 4). The protocol does not require the acquisition of new information at any stage. Its operative mechanism is the deliberate redistribution of $T(x)$ toward domains that the agent's current configuration of W, C, and M is structured to exclude.

A critical methodological specification distinguishes this protocol from simple concentration training. Sustained attentional duration is a necessary but insufficient condition for perceptual reconfiguration. If the agent merely extends $T(x)$ while the constraining operations of W, C, and M remain transparent—that is, while these mechanisms continue to regulate attention without themselves becoming objects of perceptual awareness—the extended duration produces habituation, not reconfiguration. The protocol therefore requires that each exercise combine sustained attention to the excluded domain with concurrent meta-cognitive observation of the regulatory forces that ordinarily interrupt, redirect, or suppress that attention. The operative principle is as follows: as long as attention moves rapidly across domains under its habitual allocation regime, the mechanisms governing that movement (W, C, M) remain invisible—they are the medium through which the agent perceives rather than objects within perception. When the agent deliberately sustains attention on a single domain against the habitual regime, these mechanisms surface as experiential phenomena: boredom (the cost signal of W), the judgment "this is pointless" (a norm evaluation imposed by C), moral discomfort (a prohibition enforced by M), and the impulse to disengage (the aggregate stabilization response of the cycle). The practical goal of each exercise is therefore not to "look longer" but to observe, in real time, how looking is regulated—to render W, C, and M visible as objects of perception rather than invisible as conditions of perception. Once a regulatory mechanism becomes an object of perceptual awareness, it ceases to operate as an automatic determinant of A. This is the mechanism through which the protocol produces perceptual reconfiguration rather than mere attentional endurance.

A.1 Phase 1: Diagnostic—Mapping the Current $T(x)$ Distribution

Objective

To produce an empirical record of the agent's effective attentional allocation across domains over a representative sampling period, yielding a quantitative approximation of the current perceived world as defined by the $T(x)$ distribution.

Method

The agent selects a sampling period of 5–7 consecutive days that includes both structured (work, study) and unstructured intervals. During this period, the agent records attentional engagement using a time-sampling protocol adapted from experience-sampling methodology (Csikszentmihalyi & Larson, 1987):

(i) At pseudorandom intervals (8–12 times per waking day, triggered by an external signal set to variable intervals averaging 75 minutes), the agent records:

- the domain of current attentional engagement (what is being attended to);
- the modality of engagement (visual, auditory, verbal-conceptual, proprioceptive, social-interactional, interoceptive);
- the estimated duration of uninterrupted engagement with that domain prior to the signal;
- a subjective effort rating (1–5: 1 = automatic/habitual, 5 = effortful/deliberate).

(ii) Domain categories should be inductively generated during the first two days and standardized for the remainder. A provisional taxonomy: professional/technical work, interpersonal interaction, media consumption (by type), internal planning/rehearsal, environmental monitoring, physical self-monitoring, emotional-state processing, moral/evaluative deliberation, aesthetic engagement, rest/disengagement.

(iii) At the end of the sampling period, the agent tabulates frequency and estimated duration per domain, producing a histogram that constitutes an empirical approximation of the current $T(x)$ distribution.

Expected output

A domain-frequency map revealing (a) which domains absorb the majority of attentional time, (b) which domains appear rarely or not at all, and (c) which domains carry consistently low effort ratings (indicating automatic, habituated allocation). Domains that appear rarely or not at all constitute the empirical analogue of $R \setminus P$: regions of the reference environment systematically excluded from the agent's perceived configuration.

A.2 Phase 2: Identification of Structurally Excluded Domains

Objective

To distinguish domains absent from the $T(x)$ distribution due to structural exclusion (by W, C, or M) from domains absent due to situational irrelevance, and to classify the operative exclusion mechanism for each structurally excluded domain.

Method

For each domain that received zero or near-zero $T(x)$ in the diagnostic map, the agent applies the following procedure:

(i) **Relevance test.** Would sustained attentional engagement with this domain plausibly alter the agent's understanding, perceptual sensitivity, or behavioral repertoire in any area the agent considers important? If no, the domain is situationally irrelevant and is set aside. If yes, proceed.

(ii) **Exclusion-source classification.** The agent examines the subjective quality of the avoidance:

- *W-exclusion (effort-based).* The domain is avoided because engagement feels effortful, boring, or cognitively demanding. The agent can articulate no moral or cultural objection. Characteristic phenomenology: restlessness, desire to switch tasks, the sense that "this is not worth the effort."
- *C-exclusion (norm-based).* The domain is avoided because engagement is not recognized as legitimate within the agent's reference community. The agent may experience the domain as "not serious," "irrelevant to real work," or "beneath attention." Characteristic phenomenology: sense of wasting time, anticipation of social incomprehension, feeling of being "off track."

- *M-exclusion (evaluative-prohibition).* The domain is avoided because engagement triggers moral discomfort, guilt, or the sense of violating one's own values. The agent experiences sustained attention to the domain not as boring or socially deviant but as *wrong.* Characteristic phenomenology: moral unease, contamination anxiety, the sense that even considering the domain constitutes a betrayal.

(iii) **Priority assignment.** The agent selects one to three excluded domains for Phase 3, prioritizing domains that (a) passed the relevance test, (b) produce the strongest avoidance response, and (c) are not currently addressed by any other structured practice. Domains producing the strongest avoidance are prioritized because the strength of the exclusion response is a proxy for the degree to which the perceptual configuration cycle has organized itself to prevent engagement with that domain—and therefore a proxy for the magnitude of perceptual reconfiguration that sustained engagement could produce.

A.3 Phase 3: Intervention—Meta-Cognitive Attentional Engagement with x_{excluded}

Objective

To produce a sustained increase in $T(x_{\text{excluded}})$ for the selected domain over a minimum period of 4–6 weeks, while simultaneously rendering the regulatory mechanisms (W, C, M) visible as objects of perceptual awareness. The intervention targets not the content of the excluded domain but the architecture of exclusion itself.

Session structure

The agent conducts daily sessions of 20–40 minutes. Each session has two concurrent tasks, executed simultaneously rather than sequentially:

Task 1: Sustained engagement. The agent directs and maintains attention on the selected excluded domain for the full session duration. The operative instruction is not "learn more about x" but "sustain perceptual contact with x." No new information is sought. The agent attends to the domain as it is already available—examining its structure, dwelling on features that are normally dismissed within seconds, allowing perceptual detail to develop through duration rather than through information acquisition.

Task 2: Meta-cognitive observation of regulatory interference. Concurrently, the agent monitors and records the internal events that arise during the attempt to sustain engagement. These events are not obstacles to be overcome but data to be observed—they are the experiential surface of W, C, and M as they operate in real time. The agent attends to:

- *Impulses to disengage*—the moment-by-moment urge to redirect attention elsewhere. These register the cost signal of W: the opportunity-cost computation signaling that attentional resources are being "misallocated" relative to the habitual regime.

- *Evaluative judgments*—spontaneous assessments such as "this is pointless," "this doesn't matter," "this is a waste of time." These register the norm-enforcement function of C: internalized cultural standards about what constitutes legitimate attentional engagement.

- *Moral reactions*—feelings of wrongness, guilt, contamination, or betrayal that arise from sustained contact with the domain. These register the prohibition function of M: the evaluative architecture marking the domain as inadmissible for sustained contemplation.

- *Automatic categorizations*—the rapid labeling of the domain's content ("already understood," "nothing new here," "simple/obvious") that functions to terminate further engagement. These register the joint operation of C and W: cultural categories that reduce the domain to a pre-processed summary, combined with the volitional system's preference for already-completed processing.

- *Somatic indicators*—physical restlessness, tension, fatigue, or arousal changes that accompany the attempt to sustain engagement against habitual allocation. These register the embodied dimension of the stabilization response.

The critical instruction is: the agent does not act on these regulatory signals. The agent observes them. Each impulse, judgment, moral reaction, categorization, and somatic signal is noted as a datum—a manifestation of a specific component of the perceptual configuration cycle—and attention is returned to the excluded domain. The agent does not suppress the regulatory signals (suppression would itself be a deployment of W in service of a new goal, reproducing the cycle at a different level); the agent *perceives* them while continuing to perceive the excluded domain. The regulatory mechanisms thereby shift from being the transparent medium through which the agent perceives to being objects within the agent's perceptual field. This shift is the operative mechanism of reconfiguration: a mechanism that is perceived cannot regulate perception with the same automatic efficacy as a mechanism that remains invisible.

Procedural constraints

(i) **No information-seeking.** The agent does not research, read about, or seek new facts concerning the excluded domain during the session. If the domain is the perspective of a political opponent, the agent does not read new arguments; the agent sits with the position as already understood and sustains contact with its internal structure. If the domain is an emotional state, the agent does not seek explanations; the agent attends to the state as it presents itself.

(ii) **Duration commitment.** The agent sustains engagement for the full session duration including the periods when regulatory interference is strongest. The intensity of regulatory interference is informative: it indexes the strength of the exclusion architecture at that moment.

(iii) **Session log.** At the end of each session, the agent records: (a) total duration achieved; (b) the number and type of regulatory interruptions observed (classified as W-, C-, or M-type); (c) any features of the excluded domain that became perceptually available during the session that were not available prior to it; (d) any shifts in the quality or intensity of regulatory interference compared to previous sessions.

(iv) **Non-evaluative stance.** The agent is not required to agree with, endorse, or approve of the excluded domain. The protocol concerns perceptual access, not evaluative revision. An agent who sustains 6 weeks of engagement with a morally excluded domain and concludes that the domain remains morally objectionable has nonetheless achieved reconfiguration: the objection is now grounded in sustained perceptual contact with the domain's structure rather than in reflexive dismissal, and the regulatory mechanisms that formerly rendered the domain invisible have been observed and partially de-automatized.

Expected resistance sequence

The framework predicts the following temporal pattern during the first weeks of intervention:

Days 1–4: W-dominated interference. The primary experience is effort and boredom. The cost-benefit computation of W generates persistent signals that attentional resources are misallocated. The meta-cognitive task renders these signals visible as signals rather than as accurate assessments of the domain's value. Characteristic observation: "I notice that the sense of pointlessness arises before I have actually attended to the domain long enough to assess whether it contains structure."

Days 5–12: C-dominated interference. As W-resistance partially habituates, cultural-norm interference becomes prominent. The agent experiences the practice as illegitimate, unproductive, or embarrassing. These judgments do not arise from direct assessment of the practice's utility but from internalized norms governing what constitutes appropriate attentional engagement. The meta-cognitive task renders these norms visible as norms. Characteristic observation: "I notice that the judgment 'this is a waste of time' reproduces the evaluative standards of my professional community rather than reflecting an assessment I have independently derived."

Days 12–28: M-dominated interference. For domains excluded primarily by moral-evaluative constraints, the deepest interference layer emerges after W- and C-interference have been partially de-automatized. The agent experiences moral discomfort, guilt, or the sense that sustained attention constitutes a violation of deeply held values. The meta-cognitive task renders the moral constraint visible as a constraint. Characteristic observation: "I notice that the moral prohibition operates categorically—it prevents engagement with the domain as a class rather than evaluating specific features within it. The prohibition precedes assessment." This observation does not dissolve the moral commitment; it separates the commitment from its function as an automatic regulator of A, allowing the agent to maintain the moral commitment *while* perceiving the excluded domain's structure.

Illustrative applications

The following six examples demonstrate the protocol's application across distinct domains. In each case, the operative mechanism is not information acquisition but the combination of sustained $T(x)$ increase with meta-cognitive observation of the regulatory interference that sustained engagement elicits.

Example 1: Scientific research. A computational neuroscientist whose diagnostic map reveals zero $T(x)$ for the phenomenological dimension of the cognitive processes she models. The exclusion is jointly C- and W-based: her disciplinary culture treats first-person experiential data as methodologically inadmissible (C), and engagement with phenomenological literature requires effortful processing of an unfamiliar conceptual vocabulary (W). *Protocol application.* Daily 30-minute sessions of sustained attention on the experiential character of a single cognitive process she studies computationally (e.g., attentional shifting), using her own introspective access rather than published reports. *Meta-cognitive observations recorded over weeks 1–3:* "I notice the immediate impulse to translate the experience into computational terms—to re-encode it in my habitual vocabulary" (C-interference: disciplinary norm requiring formal translation); "I notice the judgment that this exercise is 'unscientific' and that I would be embarrassed if colleagues knew I was doing it" (C-interference: professional identity protection); "I notice that the experience is richer and more temporally structured than my models represent, but that acknowledging this generates anxiety about the adequacy of my models" (W-interference: cost of destabilizing existing theoretical investment). *Perceptual reconfiguration.* By week 5, the computational models she encounters in the literature spontaneously generate a concurrent experiential dimension—she perceives the models as approximations of something she can now access independently rather than as complete descriptions.

Example 2: Interpersonal relationships. An individual whose diagnostic map reveals near-zero $T(x)$ for the experiential perspective of an estranged family member. The exclusion is primarily M-based: the thought of empathically engaging with the family member's viewpoint generates moral indignation—the sense that the other person's perspective *does not deserve* sustained consideration. *Protocol application.* Daily 25-minute sessions of sustained attention on the family member's probable experiential state during a specific past conflict, using only information already available (no new inquiry). *Meta-cognitive observations:* "I notice the impulse to rehearse grievances rather than to attend to the other's experience" (W-interference: habituated cognitive routine reasserting itself); "I notice the moral judgment that attending to this perspective is equivalent to condoning the behavior" (M-interference: evaluative conflation of perceptual engagement with moral endorsement); "I notice that my reconstruction of the conflict becomes more complex when I sustain attention—motives and constraints I had not previously perceived become structurally available" (perceptual consequence of $T(x)$ increase); "I notice that the complexity produces discomfort, because it destabilizes the version of events that supports my current position" (W- and M-interference: joint resistance to reconfiguration that threatens the existing evidential structure). *Perceptual reconfiguration.* The agent does not necessarily revise the moral judgment; the agent perceives features of the situation that were structurally unavailable under the prior $T(x) \approx 0$ allocation. The conflict situation acquires structural depth.

Example 3: Moral and political disagreement. A policy analyst whose diagnostic map reveals zero sustained engagement with the substantive arguments of the opposing political position. The exclusion is jointly M- and C-based: M marks the opposing position as morally inadmissible, C reinforces this through professional norms that treat engagement with the opposing position as a form of legitimation. The analyst can *summarize* the opposing position (the information is available) but has never sustained attentional engagement long enough for the position's internal structure—its coherence, empirical claims, and value commitments—to become perceptually available. *Protocol application.* Daily 30-minute sessions of sustained attention on a single argument from the opposing position, examining its internal logic without evaluating its conclusion. *Meta-cognitive observations:* "I notice the automatic categorization 'I already know what they think' activating within the first 90 seconds" (C-interference: cultural knowledge claim terminating further engagement); "I notice moral arousal—the feeling that by engaging seriously with this argument I am granting it status it doesn't deserve" (M-interference: sacred-value protection); "I notice that when I sustain attention past the categorization, structural features emerge that my summary does not contain—assumptions I hadn't identified, internal tensions the position's adherents themselves debate, empirical predictions I hadn't considered testable" (perceptual consequence of $T(x)$ increase); "I notice that the emergence of this structural complexity is itself uncomfortable, because it complicates the dismissal that my current configuration requires" (M- and W-joint interference: the cycle resisting the reconfiguration that would follow from perceiving the opposing position as structurally complex rather than categorically defective). *Perceptual reconfiguration.* The political landscape acquires structural differentiation. The opposing position ceases to be a monolithic category and becomes a space with internal variation, tensions, and empirical commitments—some of which the agent may continue to reject, but which are now rejected on the basis of perceptual contact rather than on the basis of structural invisibility.

Example 4: Everyday perception of environment. An urban professional whose diagnostic map reveals near-zero $T(x)$ for the non-utilitarian features of her daily commute environment. The exclusion is W-based: the commute is cognitively framed as transit time, and W allocates attention to internal planning and media consumption rather than to the physical environment. *Protocol application.* Daily 25-minute sessions during the commute in which the agent sustains visual and auditory attention on the environment—architectural detail, vegetation patterns, pedestrian behavioral micro-dynamics, ambient sound structure—without the goal of arriving efficiently. *Meta-cognitive observations:* "I notice persistent internal planning reasserting itself within 30 seconds of each redirection to the environment" (W-interference: habitual allocation defending its territory); "I notice the judgment 'there is nothing to see here' arising reflexively" (C-interference: cultural norm that commuting environments are perceptually empty); "I notice that when I sustain attention past the planning impulse, the environment contains temporal structure—sequences of events, recurrent patterns, spatial relationships—that I have traversed daily for years without perceiving" (perceptual consequence of $T(x)$ increase). *Perceptual reconfiguration.* By week 4, the commute environment spontaneously generates perceptual detail without effortful redirection. The environment has not changed; P has expanded to include features that were present in R throughout but excluded by the habitual $T(x)$ allocation. The meta-cognitive observations reveal that the exclusion was maintained not by the environment's poverty but by W's allocation strategy.

Example 5: Professional decision-making. A senior manager whose diagnostic map reveals that 80% of his decision-relevant attentional allocation falls within financial metrics, competitive positioning, and short-term deliverables. The domain "long-term consequences of current decisions for the organization's junior personnel" receives $T(x) \approx 0$. The exclusion is C-based: his professional culture treats attention to junior-personnel impact as "soft," peripheral to strategic work, and characteristic of managers who lack analytical rigor. *Protocol application.* Daily 30-minute sessions in which the agent sustains attention on the downstream effects of recent decisions on specific individuals, using only information already available. *Meta-cognitive observations:* "I notice the immediate categorization 'this is HR's domain, not mine' " (C-interference: professional role boundary enforcing attentional exclusion); "I notice the judgment that this exercise is sentimental rather than analytical" (C-interference: cultural hierarchy of cognitive styles); "I notice that when I sustain attention, specific individuals emerge from the aggregate—people with particular constraints, career trajectories, and responses to organizational changes that the aggregate metrics do not represent" (perceptual consequence of $T(x)$ increase); "I notice resistance to this specificity, because it complicates decisions that are currently experienced as clean and data-driven" (W-interference: cost of increased decision complexity). *Perceptual reconfiguration.* Decision-making contexts begin to spontaneously generate considerations that previously required external prompting. The downstream human consequences enter the agent's default P, altering the perceived content of what a "decision" contains.

Example 6: Internal emotional states. An individual whose diagnostic map reveals near-zero $T(x)$ for the domain of sustained attention to one's own emotional states during professional interactions. The exclusion is jointly C- and M-based: the agent's professional culture treats emotional self-monitoring as indulgent (C), and the agent's moral framework associates sustained attention to one's own feelings with selfishness or weakness (M). *Protocol application.* Daily 20-minute sessions following a professional interaction in which the agent sustains attention on the affective states experienced during the interaction—not analyzing their causes, not evaluating their appropriateness, but attending to their phenomenological character as presented. *Meta-cognitive observations:* "I notice the impulse to immediately narrativize the emotional state—to convert it into an explanation ('I felt that way because...')—rather than to remain in perceptual contact with the state itself" (W-interference: cognitive routinization converting perception into pre-processed narrative); "I notice the judgment 'I shouldn't feel this way' arising as an automatic categorization that truncates further attention" (M-interference: moral evaluation terminating perceptual engagement with the state's structure); "I notice that the emotional state, when attended to for more than 60 seconds without narrativization, has internal differentiation—it is not a single feeling but a complex with spatial, temporal, and qualitative structure that the label 'frustration' or 'anxiety' does not capture" (perceptual consequence of $T(x)$ increase); "I notice that the rapid labeling functioned not as accurate description but as a mechanism for terminating attention—the label is a compression that permits disengagement" (C- and W-joint interference: cultural emotion categories serving as attention-termination devices). *Perceptual reconfiguration.* The agent's internal emotional landscape acquires structural depth. Subsequent professional interactions are accompanied by a richer interoceptive field—not because the agent has learned new information about emotions, but because the attentional architecture now permits sustained perceptual contact with affective complexity that was previously compressed to categorical labels.

A.4 Phase 4: Stabilization—Emergence of Perceptual Reconfiguration

Mechanism

After 4–6 weeks of sustained intervention, the framework predicts a qualitative shift in the agent's perceptual configuration. This shift results not from information accumulation but from three structural changes in the perceptual configuration cycle:

(i) **Precision re-weighting.** The previously excluded domain has acquired non-zero precision-weighting within the agent's predictive processing hierarchy. The domain now generates prediction errors—discrepancies between the agent's implicit model of the domain and its actual structure—that register as informative rather than as noise. This is the computational signature of the domain's entry into P.

(ii) **De-automatization of regulatory mechanisms.** The sustained meta-cognitive observation of W-, C-, and M-interference has partially de-automatized these mechanisms with respect to the target domain. The mechanisms continue to operate, but they are now accompanied by perceptual awareness of their operation. A mechanism that is perceived as a mechanism no longer functions with the same transparent automaticity: the agent can notice the impulse to disengage, the cultural norm generating the judgment of irrelevance, or the moral prohibition truncating contemplation, and can choose not to act on the signal—not through suppression but through the simple fact of having perceived it. This is structurally identical to the mechanism by which metacognitive awareness of automatic thoughts reduces their behavioral influence in cognitive-behavioral paradigms, and by which mindfulness observation of impulses reduces their compulsive force in contemplative traditions—but applied here to the regulatory architecture of perception itself rather than to thoughts or impulses considered individually.

(iii) **Associative integration.** The sustained engagement has produced cross-domain connections: the previously excluded domain is now linked to the agent's broader representational structure, generating inferences, coherence pressures, and spontaneous associations that were absent when $T(x) \approx 0$.

Indicators of reconfiguration

The agent can assess reconfiguration by comparing the post-intervention $T(x)$ distribution (obtained by repeating the diagnostic protocol of Phase 1) with the pre-intervention baseline. Successful reconfiguration is indicated by: (a) a measurable increase in spontaneous (non-effortful) $T(x)$ for the previously excluded domain; (b) a decrease in the effort rating associated with engagement; (c) the appearance of the domain in the agent's $T(x)$ distribution at non-session times; (d) a qualitative increase in the perceptual detail available within the domain; and (e) a reduction in the intensity and frequency of regulatory interference during sessions, as recorded in the session logs.

A further indicator follows from the reconstructive nature of memory (Section 3.8): the agent may find, upon retrospective report, that prior experience in the reconfigured domain is recalled with greater structural detail than was available at the time of original experience. This retroactive enrichment is not confabulation but a predictable consequence of cycle reconfiguration: the new perceptual configuration provides a richer retrieval context that supports more differentiated reconstruction of past events in the target domain.

A.5 Phase 5: Constraints and Limits

Scope

The protocol does not aim at perceptual omniscience. The Constraint Lemma (Section 3.10) establishes that $P \neq R$ is structurally unavoidable for any finite system. Attentional reconfiguration does not eliminate structural distortion; it deliberately shifts its distribution. Every increase in $T(x)$ for an excluded domain entails a corresponding redistribution of attentional resources from other domains. The protocol is therefore a tool for *targeted revision* of the $T(x)$ distribution in domains where the agent has identified structural exclusion as detrimental to perceptual reliability.

Limits

(i) The protocol addresses domains excluded by W, C, and M. It does not address perceptual limitations imposed by sensory-apparatus constraints or by information that is genuinely absent from R.

(ii) The protocol requires that the agent possess the meta-cognitive capacity to distinguish regulatory interference from accurate assessment of a domain's relevance. In cases where the agent cannot distinguish the two, the Phase 2 classification procedure may produce false negatives (structurally excluded domains misclassified as situationally irrelevant).

(iii) The protocol's efficacy depends on sustained practice. The perceptual configuration cycle is a self-reinforcing architecture (Section 4); partial intervention that is discontinued before stabilization (Phase 4) will result in the cycle restoring the prior configuration. The 4–6 week minimum is derived from the timescales documented in attentional training research (Lutz et al., 2008) and should be treated as a lower bound.

(iv) The meta-cognitive component requires that the agent observe regulatory mechanisms without acting on them. This capacity is itself subject to individual variation and may require preliminary training in meta-cognitive observation (e.g., through structured mindfulness protocols) before the full protocol can be effectively implemented.

(v) The protocol does not prescribe *which* excluded domains should be selected for reconfiguration. Domain selection (Phase 2, step iii) is left to the agent's judgment. The framework provides diagnostic tools for identifying excluded domains and classifying their exclusion mechanisms, but the evaluative question of whether a given reconfiguration is desirable lies outside the protocol's scope.

A.6 Conclusion

This protocol is a direct practical consequence of the Structural Distortion Principle. If perception is constituted by the distribution of attentional duration $T(x)$, and if that distribution is regulated by the joint operation of volitional, cultural, and moral constraints (W, C, M) that are normally invisible to the perceiving agent, then perceptual reconfiguration requires two simultaneous operations: the sustained redirection of $T(x)$ toward previously excluded domains, and the concurrent meta-cognitive observation of the regulatory mechanisms that ordinarily prevent that redirection. The first operation expands the perceptual field; the second de-automatizes the architecture that kept it contracted. Neither operation alone is sufficient. Sustained attention without meta-cognitive observation produces endurance training within the existing perceptual configuration. Meta-cognitive observation without sustained engagement with excluded domains produces self-knowledge without perceptual expansion. The protocol's efficacy depends on the conjunction: the agent begins to perceive what was excluded, and simultaneously begins to perceive *why* it was excluded—rendering the exclusion mechanism itself an object of awareness rather than an invisible condition of perception. This is the practical expression of the paper's central claim: the path to greater perceptual reliability runs not through the accumulation of information but through the discipline of attention directed at both the excluded domain and the architecture of exclusion.

References

Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bateson, G. (1972). *Steps to an ecology of mind*. University of Chicago Press.

Baumeister, R. F., & Vohs, K. D. (2004). *Handbook of self-regulation: Research, theory, and applications*. Guilford Press.

Bourdieu, P. (1977). *Outline of a theory of practice* (R. Nice, Trans.). Cambridge University Press.

Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.

Cisek, P. (2019). Resynthesizing behavior through phylogenetic refinement. *Attention, Perception, & Psychophysics*, *81*(7), 2265–2287.

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

Constant, A., Bervoets, J., Hens, K., & Steyaert, J. (2019). Precise worlds for certain minds: An ecological perspective on the relational self in autism. *Topoi*, *39*, 611–622.

Csikszentmihalyi, M., & Larson, R. (1987). Validity and reliability of the experience-sampling method. *Journal of Nervous and Mental Disease*, *175*(9), 526–536.

Dane, E. (2010). Reconsidering the trade-off between expertise and flexibility: A cognitive entrenchment perspective. *Academy of Management Review*, *35*(4), 579–603.

De Brigard, F. (2012). The role of attention in conscious recollection. *Frontiers in Psychology*, *3*, 29.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215.

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

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

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, *113*(1), 44–66.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded rationality: The adaptive toolbox.* MIT Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* Oxford University Press.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015).*

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046.

Gregory, R. L. (1970). *The intelligent eye.* Weidenfeld & Nicolson.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion.* Vintage Books.

Harmon-Jones, E., & Mills, J. (2019). An introduction to cognitive dissonance theory and an overview of current perspectives on the theory. In E. Harmon-Jones (Ed.), *Cognitive dissonance: Reexamining a pivotal theory in psychology* (2nd ed., pp. 3–24). American Psychological Association.

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

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.* Princeton University Press.

Hohwy, J. (2013). *The predictive mind.* Oxford University Press.

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, *35*(2), 183–204.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Husserl, E. (1982). *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy* (F. Kersten, Trans.). Nijhoff. (Original work published 1913)

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–682.

Kriger, B. (2019). Evolutionary selection for atemporal memory storage: Why three convergent pressures favor architectures where time belongs to retrieval, not to storage. *Zenodo*. https://doi.org/10.5281/zenodo.18381880

Kriger, B. (2021). The evolutionary architecture of human consciousness: Cognitive contradictions, biological duality, and the illusion of time. *Zenodo*. https://doi.org/10.5281/zenodo.18384277

Kriger, B. (2022). Evolutionary theory of credence: A conceptual framework with formal analogies for understanding generative modeling as a resource-theoretic consequence of complexity. *Zenodo*. https://doi.org/10.5281/zenodo.18379476

Kriger, B. (2025). Atemporality of mental memory space: A structural hypothesis grounded in resource constraints, cyclical closure, and reconstructive retrieval. *Zenodo*. https://doi.org/10.5281/zenodo.18381912

Kriger, B. (2026a). Differentiation as the ontological condition of actualization. *Zenodo*. https://doi.org/10.5281/zenodo.18268520

Kriger, B. (2026b). The evolutionary inevitability of predictive processing: A physical constraint argument. *Zenodo*. https://doi.org/10.5281/zenodo.18444910

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661–679.

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: A meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*(4), 870–900.

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82.

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

Luhmann, N. (1995). *Social systems* (J. Bednarz Jr. & D. Baecker, Trans.). Stanford University Press.

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2008). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences*, *12*(4), 163–169.

Mack, A., & Rock, I. (1998). *Inattentional blindness*. MIT Press.

Maturana, H. R., & Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Reidel.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57–74.

Merleau-Ponty, M. (2012). *Phenomenology of perception* (D. A. Landes, Trans.). Routledge. (Original work published 1945)

Mylopoulos, M., & Lau, H. (2020). Naturalizing intentional action. In U. Apfelbaum & R. A. Mele (Eds.), *Advances in experimental philosophy of action* (pp. 131–158). Bloomsbury Academic.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722–726.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220.

Nisbett, R. E. (2003). *The geography of thought: How Asians and Westerners think differently... and why*. Free Press.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences*, *9*(10), 467–473.

Noë, A. (2004). *Action in perception*. MIT Press.

Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. *Philosophical Transactions of the Royal Society B*, *377*(1844), 20200531.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, *58*, 1–23.

Raichle, M. E., & Gusnard, D. A. (2002). Appraising the brain's energy budget. *Proceedings of the National Academy of Sciences*, *99*(16), 10237–10239.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758.

Sara, S. J. (2000). Retrieval and reconsolidation: Toward a neurobiology of remembering. *Learning & Memory*, *7*(2), 73–84.

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

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception*. Bobbs-Merrill.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

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

Simondon, G. (2020). *Individuation in light of notions of form and information*. University of Minnesota Press. (Original work published 1958)

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074.

Spencer-Brown, G. (1969). *Laws of form*. Allen & Unwin.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Blackwell.

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393.

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*(7), 320–324.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870.

Thompson, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Treisman, A. M. (1964). Selective attention in man. *British Medical Bulletin*, *20*(1), 12–16.

Varela, F. J., Thompson, E., & Rosch, E. (1991). *The embodied mind: Cognitive science and human experience*. MIT Press.

von Foerster, H. (1981). *Observing systems*. Intersystems Publications.

von Uexküll, J. (2010). *A foray into the worlds of animals and humans* (J. D. O'Neil, Trans.). University of Minnesota Press. (Original work published 1934)

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, *3*, 1212–1217.