



OPEN ACCESS

EDITED BY

Gustavo E. Tafet,
Texas A and M University, United States

REVIEWED BY

Asle Hoffart,
Modum Bad Psychiatric Center, Norway
Carlos Hesed Virto Farfan,
Andean University of Cusco, Peru

*CORRESPONDENCE

Hebert W. Harris
✉ herb.hw@gmail.com

RECEIVED 18 May 2025

ACCEPTED 24 July 2025

PUBLISHED 19 August 2025

CITATION

Harris HW (2025) Active inference and psychodynamics: a novel integration with applications to depression and stress disorders.
Front. Psychiatry 16:1630858.
doi: 10.3389/fpsyt.2025.1630858

COPYRIGHT

© 2025 Harris. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Active inference and psychodynamics: a novel integration with applications to depression and stress disorders

Hebert W. Harris*

Private Practitioner, Arlington, VA, United States

This paper introduces Active Intersubjective Inference (AISI), a novel framework that integrates psychodynamic theory with predictive processing to explain self-identity construction and psychopathology. AISI posits that the self emerges from recursive inferences about how others perceive us (second-order self), interacting bidirectionally with interoceptive processes. We map psychodynamic phenomena (e.g., transference, projection) onto neurocomputational mechanisms and apply AISI to Major Depressive Disorder (MDD) and Post-Traumatic Stress Disorder (PTSD), highlighting distorted second-order inference as a core dysfunction. Therapeutic implications include psychodynamic therapy, cognitive-behavioral approaches, and psychedelic-assisted treatments to enhance inference flexibility. AISI bridges psychodynamic insights with the NIMH Research Domain Criteria, offering a testable model for precision psychiatry and future clinical trials.

KEYWORDS

active inference, predictive processing, psychodynamic theory, second-order self, depression, PTSD, free energy principle, computational psychiatry

1 Introduction

Psychodynamic theories have long emphasized the centrality of relationships in shaping the self. However, they have lacked a neurobiological framework. Conversely, neuroscience and computational psychiatry have made substantial progress in modeling brain function as predictive inference. This paper aims to bridge these domains by introducing Active Intersubjective Inference (AISI), a framework that maps psychodynamic concepts onto a formal model of recursive prediction.

The AISI framework proposes that the self is constituted through active inferences about how others perceive us. We extract information about ourselves from the language and behavior of others, and we refer to the resulting predictive model as the second-order self. In contrast, the first-order model of self is derived directly from bodily sensations, including proprioception and physiological data. These direct, first-order models are termed interoceptive selves (1–4).

We view personal identity as arising from bidirectional interactions between the interoceptive and second-order selves. Interoceptive states influence our perceptions of social cues. At the same time, the threats, attitudes, and intentions of others, processed through AISI, shape the second-order self, which influences physiological responses (e.g., cortisol release) via top-down modulation of the interoceptive self. Conversely, bottom-up interoceptive signals (e.g., heart rate changes) update second-order self-models.

The prior interpersonal experiences that shape the second-order self are continuously updated through feedback from real and imagined interactions with others. Constructing second-order models from social, linguistic, and behavioral cues is complex and error-prone. It relies heavily on Bayesian priors derived from narrative memories that may date back to interpersonal interactions of early childhood.

AISI is a highly active inference process in which our language and behavior influence the inferences and behaviors of others in real time. We use the term intersubjective to emphasize that it involves the interactions of mental representations of self and other between two or more individuals.

The Free Energy Principle (FEP) holds that organisms behave in ways that minimize errors generated by active inference processes (5, 6). In this paper, we observe that applying the FEP to AISI may elucidate common psychodynamic processes such as projection and splitting while addressing many psychodynamic aspects of depression and stress-related disorders.

This paper focuses on MDD and PTSD to illustrate AISI's applicability to trauma and depression. The paper is structured first to define the AISI model and its neuroanatomical underpinnings, then to apply it to clinically relevant psychodynamic processes, as well as to major depressive disorder (MDD) and post-traumatic stress disorder (PTSD). Finally, we consider therapeutic implications and outline a future research agenda grounded in computational psychiatry and AI-enhanced psychodynamic assessment.

2 Theoretical framework

2.1 Overview of active inference

The active inference framework, developed by Karl Friston and colleagues, extends predictive processing through the free energy minimization principle (5). Here, “free energy” is a mathematical concept that quantifies the difference between the brain's internal model of the world and the actual state of the world as indicated by sensory data. In simpler terms, free energy is a measure of surprise or uncertainty.

According to this framework, biological systems are driven to minimize free energy through motor and sensory mechanisms. The brain constantly forms internal models that are updated to better match sensory input. Additionally, the brain initiates actions to alter sensory input in two ways: (1) changing perception, such as redirecting gaze to clarify ambiguous social cues (e.g., observing a facial expression), and (2) modifying the environment, such as

engaging in behaviors to elicit specific social responses (e.g., smiling to evoke a positive reaction) (7, 8).

Precision-weighting is a key mechanism in active inference. Prediction errors are assigned weights based on their estimated reliability or “precision.” High-precision errors drive model updates, while low-precision errors may be effectively ignored. This precision-weighting mechanism allows the brain to flexibly balance reliance on prior models versus new information, depending on context and levels of uncertainty. For example, in a noisy social setting (high uncertainty, e.g., a crowded party with ambiguous facial expressions), the brain may assign low precision to sensory data, relying more on priors (e.g., expecting friendliness from known peers). Conversely, in a clear context (low uncertainty, e.g., a one-on-one conversation with distinct cues), sensory data receive higher precision, driving model updates (e.g., revising a prior of rejection when a smile is clearly observed). Uncertainty reflects the brain's estimate of variability in sensory data or priors, distinct from but influencing the size of prediction errors, which are the mismatches between predictions and actual input (5). The relevance of information for reducing uncertainty is determined by its potential to minimize free energy, prioritizing reliable data (e.g., consistent social cues like repeated smiles) and contextually salient information (e.g., threat cues in dangerous settings or affiliative cues in bonding scenarios). Attention amplifies high-precision signals to optimize model updates (9).

The predictive hierarchy refers to layered brain processes where lower levels predict immediate sensory data (e.g., seeing a smile) and higher levels predict abstract states (e.g., others' approval). For instance, in a social setting, low-level predictions about facial expressions feed into high-level predictions about social acceptance. Precision allocation across this hierarchy is understood as a form of attention that directs the system's resources toward information considered most relevant for reducing uncertainty. Attention to one's own predictions is possible, as individuals can reflect on their expectations (e.g., ‘I expect rejection’) and adjust them based on new evidence, a process central to second-order inference. Disruptions in precision allocation are implicated in various psychological conditions, ranging from hallucinations (overweighted prior beliefs) to heightened anxiety (overweighted threatening prediction errors) (10, 11).

To ensure the AISI framework is empirically testable, its predictions can be evaluated using computational models such as the Hierarchical Gaussian Filter (HGF; 12), which quantifies belief updating in social contexts, and neuroimaging techniques like fMRI to measure activity in regions supporting second-order inference (e.g., medial prefrontal cortex, anterior insula). Virtual reality (VR)-based tasks can further simulate social interactions to assess inference flexibility, as detailed in Section 6.1.

2.2 From individual to intersubjective prediction

In social contexts, the brain's predictive architecture faces a unique challenge. We must model the mental states and beliefs of other agents who are actively forming predictions about us. This

creates a reciprocal system in which each person's predictions and actions provide sensory data for others' predictions. Moreover, our actions are shaped not only by our models of others but also by our models of how others model us, creating nested, recursive prediction hierarchies that are characteristic of human intersubjectivity.

2.3 First-order and second-order inference

2.3.1 First-order inference: modeling bodily states and observable behaviors

First-order inference operates on data derived from internal, physiological sources. It produces predictive models of autonomic, neuroendocrine, and proprioceptive states. At the highest levels, these models combine into a framework known as the interoceptive self (1–3). The interoceptive self integrates information about energy levels, appetite regulation, reward expectations, danger/anxiety signals, libido, and other biological drives that collectively shape mood states (13, 14). For example, in anxiety, first-order inference might amplify heart rate signals as predictions of imminent threat.

First-order inference is also employed to make predictive models of others' directly observable behaviors, including speech, facial expressions, and actions. At this level, we model others as complex agents, but do not attempt to understand the intentionality underlying their behavior or comprehend their mental states. This process predominates during early development before Theory of Mind (ToM) emerges.

2.3.2 Second-order inference: modeling others' perceptions of the self

Once ToM development begins, second-order inference allows us to model others' mental states, including how they perceive us—a process we refer to as recursive self-modeling. This generates predictive images of self and other (Figures 1A, B). In contrast to first-order inference's focus on observable actions (e.g., Person A predicts Person B's smile), second-order inference interprets underlying intentions (e.g., Person A infers that Person B's smile indicates approval). Details of how Person B models Person A are disclosed through language and behavioral cues. From this information, Person A can construct inferential models of themselves. First-order and second-order inference continue in parallel after ToM develops, with first-order handling direct sensory and behavioral modeling while second-order adds interpretive layers for mental states.

The socially constructed self (second-order self) must align with other sources of information, such as interoception, which provides data on mood states and biological drives (1–3). The interoceptive components of self, such as physiological states and biological drives, interact bidirectionally with the second-order self to produce mood and affect, and to elicit physiological responses appropriate to social contexts. For instance, a perceived social threat (second-order) might trigger cortisol release (interoceptive), while physical fatigue (interoceptive) could bias perceptions of social rejection (second-order). This bidirectional loop operates as follows: top-down modulation from the second-order self (e.g., negative social feedback strengthens priors of rejection, increasing cortisol via the

hypothalamus); bottom-up signals from the interoceptive self (e.g., physiological arousal signals danger, reinforcing negative second-order priors).

Self-experiences from activity, mastery, or situational performance (e.g., succeeding in a task) are integrated into the second-order self when interpreted through social feedback (e.g., praise from others) or internalized as self-evaluation, a process akin to Kohut's mirroring (Section 3.5). These experiences contribute to priors of competence, distinct from but interacting with interoceptive signals.

As illustrated in Figure 2, these shared priors create a dynamic feedback loop between bodily experiences and social self-representations. Social interactions have profound effects on interoceptive processes. Experiences of social threat, rejection, or affiliative bonding trigger specific neuroendocrine cascades that alter physiological parameters, including heart rate variability, cortisol levels, and immune function (15, 16). These bidirectional interactions, central to AISI, explain how social disruptions cascade into physical symptoms in disorders like depression (see Section 4). Actions to alter sensory input are central to AISI, as they test predictions about others' perceptions (e.g., seeking clarification in a conversation to resolve ambiguity in social cues).

2.4 The functional and evolutionary significance of higher-order inference

Cooperative activities—from hunting to childcare to cultural learning—depend on predicting others' actions and understanding their expectations of us. The emergence of second-order inference capabilities likely created powerful selection pressures throughout human evolution. Individuals who could accurately model how group members perceived them would gain significant advantages in managing their social reputation, forming alliances, and navigating hierarchies. These capabilities may have co-evolved with language and ToM (see Section 2.6 for developmental details). In psychopathology, disruptions in this process contribute to relational instability (see Section 3).

2.5 Neuroanatomical substrate

The neuroanatomical basis of AISI involves interconnected brain regions that support bidirectional information flow between interoceptive and second-order processes, as shown in Figure 2. This figure illustrates the recursive structure of AISI, where self-representations are continuously shaped by social and physiological feedback. The interoceptive self integrates signals related to drives, reward, and homeostasis (e.g., motivation, libido, sleep, energy, appetite), which in turn modulate the second-order self through neuroendocrine pathways. A key assumption is that these regions enable prediction generation and feedback integration to update models.

Several large-scale brain networks overlap with this architecture, including the Default Mode Network (DMN), the Salience Network, and components of the Mirror Neuron System. Each region contributes to one or more functions necessary for

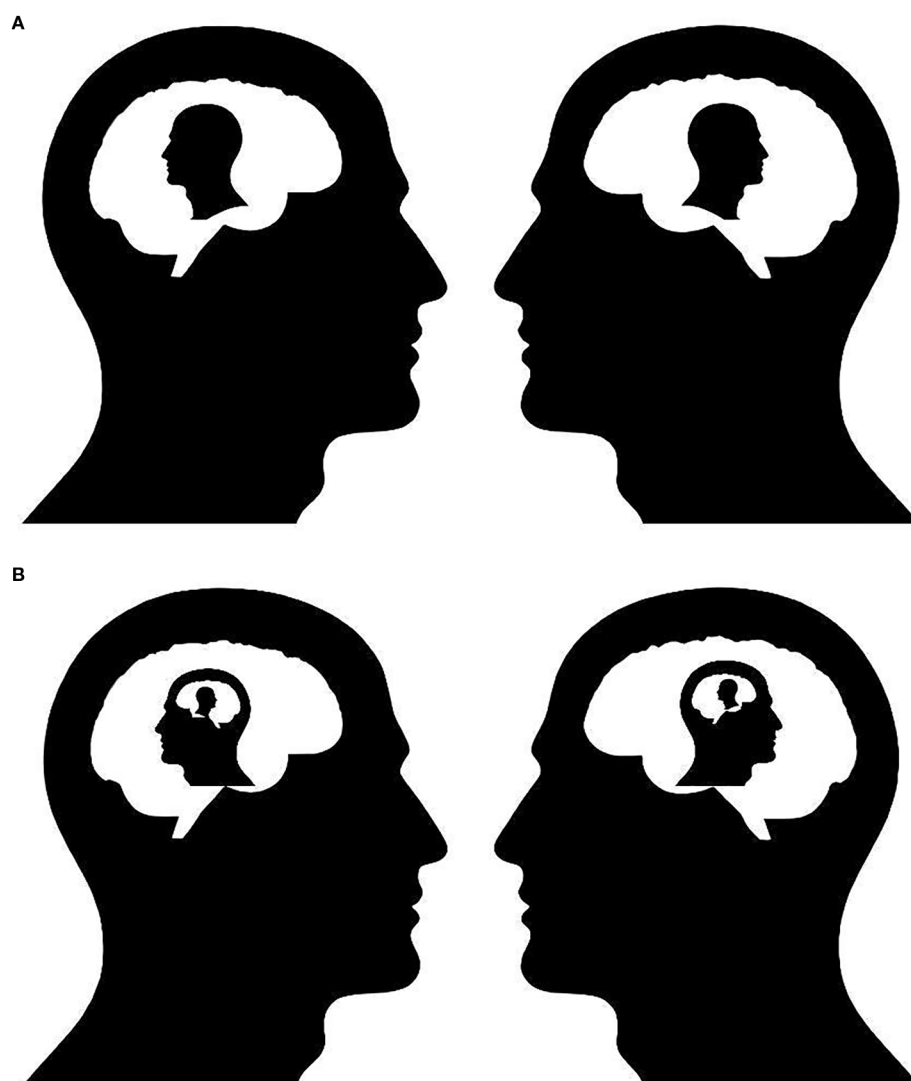


FIGURE 1

(A) First-Order Active Interpersonal Inference. This schematic illustrates an agent engaged in first-order active inference within a dyadic social interaction. The agent (left) uses sensory and contextual cues to generate predictions about the behavior, intentions, and emotional state of another person (right). However, the model remains egocentric; the agent does not represent how it is itself perceived by the other. In neurocomputational terms, this process recruits predictive coding hierarchies but does not yet involve higher-order recursive modeling. This form of inference supports basic empathy, behavioral synchronization, and social coordination but does not fully capture the reflexive dynamics of self-consciousness or interpersonal identity. **(B)** Second-Order Active Interpersonal Inference. This figure expands the model to depict second-order AISI, wherein the agent infers not only the other's mental states but also how the other perceives the agent in return. This recursive loop—I model how you see me—enables the emergence of complex interpersonal emotions such as shame, pride, embarrassment, and trust. The diagram highlights the closed-loop architecture of second-order social inference, essential for self-conscious affect and reflective identity. Such modeling requires more complex generative architectures and may depend on the integration of medial prefrontal, temporoparietal, and default mode networks in the brain.

intersubjective inference: modeling others' mental states, integrating social feedback, mapping bodily states onto emotion, and maintaining narrative continuity of identity.

Table 1 summarizes key regions implicated in AISI, focusing on those relevant to depression and trauma (detailed in Section 4). These include the medial prefrontal cortex (mPFC) for self-modeling, temporoparietal junction (TPJ) and posterior cingulate cortex (PCC) for social feedback, anterior cingulate cortex (ACC) for error monitoring, and anterior insula (AI) for interoceptive bridging. Notably, these regions are densely interconnected with subcortical structures (e.g., hypothalamus, amygdala), allowing the

second-order self to modulate physiological states and vice versa. This substrate underpins the negotiation between internal drives and social expectations in AISI.

2.6 Developmental trajectory of active intersubjective inference

Building on the distinction in Section 2.3, AISI develops alongside ToM. Through observation and interaction, toddlers (1–2 years) build predictive models of caregivers' behaviors. However, these are

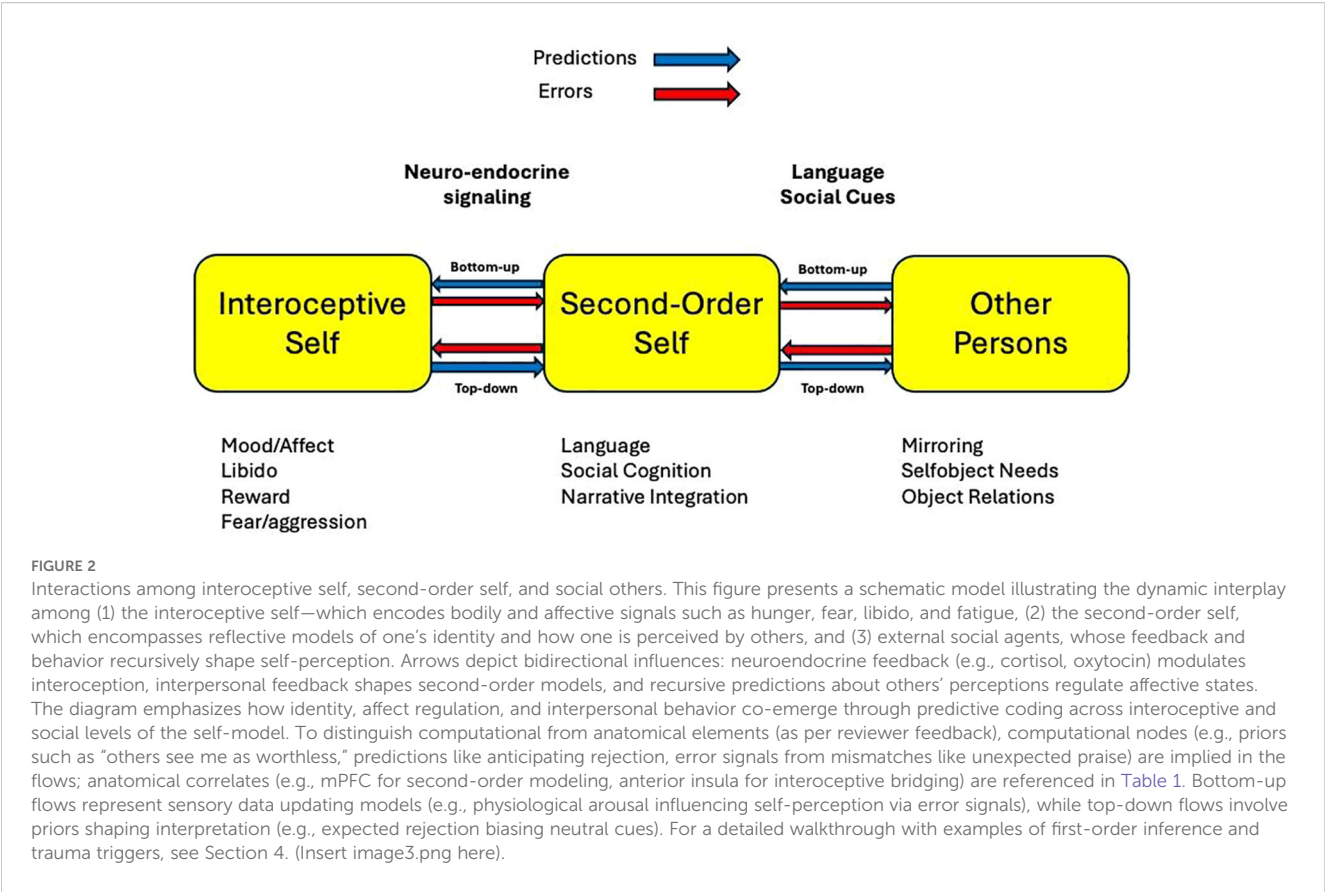


TABLE 1 Brain regions involved in AISI.

Region	Known function	Rationale for involvement in AISI
Medial Prefrontal Cortex (mPFC)	Self-referential processing; social cognition (17, 18)	Maintains hierarchical representations of self and others; supports second-order prediction models
Temporoparietal Junction (TPJ)	Social perspective-taking; predictive updating (19, 20)	Compares predicted and actual social feedback; essential for self-model revision
Posterior Cingulate Cortex (PCC)	Autobiographical memory; DMN integration (21, 22)	Integrates social feedback with narrative identity; maintains continuity of self
Anterior Cingulate Cortex (ACC)	Prediction error monitoring; conflict detection (23, 24)	Signals need for updating self-other models during interpersonal mismatch
Anterior Insula (AI)	Interoception; emotional salience (4, 25)	Links bodily states with emotional responses and social context
Precuneus	Mental imagery; social simulation (26, 27)	Simulates future social interactions and predicts others' responses
Hippocampus	Episodic memory; predictive mapping (28, 29)	Encodes social interactions and supports learning from social prediction errors

primarily first-order models. Theory of Mind (ToM)—the ability to attribute mental states to others—develops through predictable stages (30). The milestone of understanding false beliefs (ages 4–5) enables children to build second-order models of themselves based on caregivers’ inferences. ToM and second-order inference likely co-develop, with ToM enabling the shift to recursive modeling; more research is needed to establish the precise time course.

Winnicott’s (31) concept of play offers a psychodynamic parallel to active inference. Play serves as a low-risk space for testing social hypotheses. Mastery experiences in play contribute to second-order priors by providing feedback on competence (e.g., succeeding in a game elicits caregiver praise, reinforcing self-models of capability). Winnicott’s (32) “good-enough mother” aligns with AISI by providing empathic feedback to calibrate inference during development.

Over time, the second-order self acquires greater stability and autonomy as it becomes anchored by narrative memories. Bayesian priors favor continuity in self-models. However, constructing an autonomous self requires caregiver support; failures may lead to pathology (Section 3.5).

2.7 Integration of personal identity

The information we use to construct second-order models of ourselves comes from others, who may present the developing child

with divergent or even contradictory images that must be integrated into their self-model. The continuity and coherence of the child’s second-order self are relative to the frame of reference provided by their caretakers. The child’s sense of identity depends on the stable, consistent presence of others. This relativity makes the child vulnerable to changes in this frame of reference. While this may be an unavoidable part of development, when it becomes extreme, the child may experience distortions of self/other boundaries and displacement of identity. Distortions and displacements of the second-order self’s identity may manifest as projection, splitting, and other psychodynamic phenomena discussed in Section 3.

Over time, the second-order self acquires greater stability and autonomy as it becomes anchored by a growing body of narrative memories. The FEP and Bayesian inference drive us to predict that our next self will maintain continuity with our recent selves. However, the construction of a fully autonomous, enduring self may require considerable support from the child’s caregivers. If the child is allowed to test their autonomy while consistently receiving empathic, validating feedback, they can develop second-order models with increasing independence. Failures at this stage may lead to pathological development, as discussed in Section 3.5.

To some extent, the relativity of the second-order self persists throughout life. For example, the common experience of feeling like a different person in various social contexts or adopting collective identities of race, ethnicity, and nationality manifests the malleability of second-order identity (33). The nearly universal experiences of empathy and compassion may reflect displacements of identity with the transient, partial exchange of self with other. The integration of second-order models of the self into larger narratives is a process that continues throughout the entire lifecycle (34).

3 Psychodynamic phenomena as active intersubjective inference

AISI reframes psychodynamic processes as strategies for managing social uncertainty via free energy minimization. Table 2 summarizes key phenomena, their AISI mechanisms, and examples, reducing overlap in the text below.

3.1 Transference as Bayesian inference

Transference redirects past relational priors to current interactions (35). In AISI, these act as Bayesian priors for second-order models. Mismatches generate errors, but ingrained priors resist updates due to: (1) strong developmental weighting, (2) inflexibility, or (3) maladaptive error thresholds. For example, a patient might transfer abandonment fears onto the therapist, minimizing short-term uncertainty but perpetuating cycles. These inflexible priors are rigid beliefs that fail to account for alternative interpersonal scenarios (e.g., expecting hostility in all interactions, unable to consider positive alternatives). Therapeutic co-creation (36) minimizes dyadic free energy, sometimes reinforcing patterns.

TABLE 2 Psychodynamic phenomena in AISI terms.

Phenomenon	AISI mechanism	Clinical example
Transference	Applying Bayesian priors from past relationships to new contexts; resistant updating due to high prior precision	Patient expects therapist criticism based on parental models, ignoring neutral cues
Projection	External attribution of ego-dystonic content to minimize self-model errors	Attributing anger to a colleague to avoid internal conflict
Splitting	Partitioning self-models into binary categories for computational efficiency	Viewing self/others as all-good or all-bad in borderline states
Projective Identification	Active induction of projected content in others via intersubjective loops	Inducing helplessness in therapist to confirm disowned vulnerability
Self Psychology (Narcissism)	Reliance on external mirroring due to inflexible second-order priors	Needing constant validation to sustain fragile self-coherence

3.2 Projection

Projection attributes unacceptable impulses externally (37). In AISI, it resolves ego-dystonic errors by displacing them to others’ models, reducing internal free energy. For example, a depressed individual projects self-loathing as others’ judgment, maintaining self-coherence at interpersonal cost.

3.3 Splitting

Splitting divides experiences into extremes (38). AISI views it as partitioning second-order self-models to avoid integration costs. Under stress, binary categories reduce uncertainty short-term but lead to oscillations (e.g., idealizing/devaluing in relationships). Neurobiologically, it involves poor valence network integration.

3.4 Projective identification

Projective identification induces others to enact projections (38). AISI breaks it into: disavowal, projection, active induction via intersubjective loops, and identification. This blurs self-other boundaries, as in trauma where victims induce rejection to confirm priors.

3.5 Self psychology in the AISI framework

Kohut’s (39) self psychology emphasizes self object needs in narcissism. In AISI, developmental failures lead to reliance on mirroring for second-order stability. Empathic therapy updates priors, enhancing flexibility. Neuroimaging shows altered insula/mPFC in subjects with narcissistic traits (40, 41).

4 Major depressive disorder and post-traumatic stress disorder through the lens of AISI

Both major depressive disorder (MDD) and post-traumatic stress disorder (PTSD) involve distortions in the modeling of the second-order self, the predictive representation of how others perceive one. As illustrated in Figure 2 and supported by the neural structures listed in Table 1, the second-order self and interoceptive self are engaged in recursive, bidirectional inference loops. Dysfunction in either domain can cascade to the other, producing the range of affective, somatic, and cognitive symptoms observed in these conditions.

To clarify the model’s structure (as illustrated in Figure 2), let’s walk through its key steps. The interoceptive self generates bottom-up signals from physiological data (e.g., heart rate variability as error signals indicating arousal), which update priors in the second-order self (e.g., interpreting arousal as social threat). Conversely, top-down flows from the second-order self propagate predictions (e.g., “others reject me”) to shape interoceptive interpretation, minimizing overall prediction errors. First-order social inference operates in parallel, such as a smile eliciting a reciprocal smile via direct behavioral mirroring (e.g., mirror neuron activation) without engaging second-order mental state attribution. In trauma, a trigger like a sound can elicit both predictive processing (reactivating rigid second-order priors of vulnerability, leading to re-experiencing) and direct neuroendocrine changes (e.g., immediate amygdala-driven cortisol release, bypassing full inference). These bidirectional, recursive processes—grounded in free energy minimization—highlight how distortions cascade across levels in MDD and PTSD.

Table 3 summarizes these interactions in normal versus pathological states, providing a quick reference for how first-order, second-order, and physiological responses align or diverge.

4.1 MDD and the self

From a predictive coding perspective, MDD involves maladaptive models of the interoceptive self, including dysfunctions in systems governing sleep, appetite, energy, libido, and reward-seeking (42). We extend this account to incorporate distorted second-order self-models, informed by psychodynamic theory.

Classic psychodynamic models of depression emphasize the role of internalized object relationships. Freud’s “Mourning and Melancholia” (43) described depression as a response to unconscious loss and identification with a punitive object. Klein, Bowlby, and Jacobson further elaborated on how object loss and ambivalent attachment contribute to depressive states. Blatt (44) distinguished between introjective depression, focused on self-criticism and guilt, and anaclitic depression, centered on feelings of abandonment and helplessness.

TABLE 3 AISI processes in social and traumatic contexts.

Process	Normal state	Pathological state (e.g., MDD/PTSD)
First-Order Inference (Observable Behaviors/ Physiological Signals)	Direct mirroring of behaviors (e.g., smile elicits reciprocal smile via mirror neurons); flexible updating of physiological signals (e.g., heart rate as temporary arousal).	Rigid behavioral responses (e.g., hypervigilance to neutral cues); overamplification of signals (e.g., persistent arousal misinterpreted as threat).
Second-Order Inference (Mental State Modeling)	Recursive updating of self-models based on social feedback (e.g., praise adjusts priors of self-worth); balanced precision weighting.	Distorted priors resist update (e.g., rigid negative self-view despite positive feedback); over precision on threat/rejection.
Physiological Responses (Neuroendocrine/ Interoceptive)	Bidirectional adaptation (e.g., social bonding reduces cortisol); error minimization maintains homeostasis.	Cascading dysregulation (e.g., trauma trigger causes direct amygdala-cortisol surge + predictive re-experiencing); chronic hyperarousal or shutdown.
Interaction Example	Neutral sound processed as benign; social cue integrates bottom-up data with top-down priors for adaptive response.	Sound triggers dual paths: direct neuroendocrine spike (e.g., fear response) and predictive reactivation (e.g., flashback), amplifying errors.

The AISI framework situates these perspectives within a computational model of second-order inference. Vulnerability to depression emerges from:

1. The loss of attachment figures or self objects who contributed positive second-order representations leads to a deficit in affirming self-models, producing symptoms associated with anaclitic depression (44, 45). These include feelings of abandonment, helplessness, and interpersonal dependency.
2. Internalization of critical or punitive object representations generates persistently negative second-order priors, manifesting as introjective depression with themes of guilt, inadequacy, and self-criticism.

These distorted second-order priors exert top-down suppressive effects on interoceptive prediction, leading to diminished energy, motivation, appetite, and reward-seeking. This state may serve an adaptive function—minimizing energy expenditure in anticipation of social futility, akin to a metabolic “hibernation” (46).

Mood emerges as a complex construct reflecting predictive modeling across multiple domains. Depressive mood states appear to represent the integration of predictions about energy availability, sleep-wake cycles, appetite regulation, libido, reward anticipation, anhedonia, and crucially, predictions of social rejection and social defeat (47–49). The social defeat stress model in rodents provides a compelling parallel to AISI dysfunction, where

repeated social subordination leads to lasting alterations in both social behavior and physiological markers of depression (50).

Disruptions in interoceptive inference—due to chronic illness, inflammation, or dysregulated neurochemistry—may influence second-order models of the self. Low precision in interoceptive prediction increases the likelihood of misattribution (e.g., interpreting fatigue or somatic discomfort as signs of inadequacy or failure), reinforcing maladaptive second-order predictions (51; Gilbert et al., 2022). These bidirectional prediction processes, outlined in Section 2.3, demonstrate how disruptions in one domain cascade to affect the other, creating self-perpetuating cycles of dysfunction between the interoceptive self, the second-order self, and relations with others.

The AISI framework provides a unifying account for various depressive presentations. Melancholic depression, characterized by severe anhedonia and psychomotor changes, can be understood as pathologically rigid negative second-order priors combined with substantially disrupted interoceptive prediction. Atypical depression, with its altered mood reactivity and rejection sensitivity, reflects hypervigilant second-order inference systems that overweight potential social threats. These diverse presentations emerge from different patterns of dysfunction in the recursive loops between interoceptive and second-order inference systems.

Functional neuroimaging studies of MDD support this bidirectional model (52). Altered activity and connectivity have been consistently observed in regions central to AISI, including the mPFC, anterior insula, and ACC (52–54). In depression, positive social feedback often fails to update negative second-order self-perceptions, as they generate insufficient prediction error to override established negative priors.

In severe depression, suicide may represent a terminal outcome of unbearably negative second-order inference. When an individual's negative self-models become so rigid that no corrective experiences can update them, the psychological pain may become intolerable.

4.2 Trauma and the second-order self

Post-traumatic stress disorder (PTSD) is characterized by exposure to actual or threatened death, serious injury, or sexual violence, followed by distinct symptom clusters: intrusive memories, avoidance behaviors, negative alterations in cognition/mood, and heightened arousal/reactivity (55). Complex PTSD (C-PTSD) includes additional disturbances in self-organization, emotion regulation, and interpersonal functioning (56).

Psychodynamic theories have long emphasized how trauma alters self-experience, noting the internalization of the perpetrator's degrading view of the victim, the collapse of representational boundaries, and the inability to integrate traumatic experiences into a coherent self-narrative (57–59).

4.2.1 PTSD and predictive models of the self

The interoceptive self plays a central role in PTSD symptomatology. Hyperarousal, sleep disturbances, and exaggerated startle responses reflect dysregulated predictions at the interoceptive level. These manifest physiologically through heightened sympathetic activity, elevated baseline cortisol, and disrupted hypothalamic-pituitary-adrenal axis functioning (60). Fear-potentiated startle—the exaggerated acoustic startle response in threatening contexts—provides a particularly well-documented example of altered predictive processing, with robust evidence from both human and animal studies (61, 62).

Within the AISI framework, trauma also produces enduring distortions in second-order self-models. In many cases, trauma involves a perpetrator who imposes a hostile and degrading image of the victim. This model, transmitted through coercion, fear, or shame, may override the victim's preexisting second-order self-model, particularly if the trauma occurs during critical developmental stages. Once internalized, this hostile second-order model becomes a rigid prior, generating persistent predictions of vulnerability, worthlessness, and social rejection.

Triggering events in the present may reactivate these priors, leading to an intrusion of traumatic memories and associated affective states. Avoidance behaviors can be understood as active inference strategies to minimize exposure to stimuli that would generate high prediction errors against these trauma-informed models. The chronic hypervigilance in PTSD reflects an overprecision of threat-related priors in second-order inference, where even neutral social cues are interpreted as confirmatory evidence of danger or rejection.

The bidirectional link between second-order and interoceptive selves explains PTSD's somatic symptoms. A second-order model of the self as 'damaged/dangerous/worthless' maintains chronic stress system activation, perpetuating the physiological dysregulation characteristic of PTSD.

Neuroimaging studies support this model. PTSD has been associated with dysregulation in the medial prefrontal cortex (mPFC), anterior cingulate cortex (ACC), hippocampus, and anterior insula—regions implicated in both interoceptive inference and self-referential processing (63, 64). Reduced mPFC-amygdala connectivity may impair updating of trauma-related second-order priors, perpetuating hypervigilance and maladaptive threat perception in social contexts (24, 65).

4.2.2 Complex PTSD and identification with the aggressor

Complex PTSD (C-PTSD) stems from prolonged, repeated trauma exposure, particularly during developmental periods (56, 66). Its distinctive features include profound disturbances in self-organization manifesting as persistent negative self-concept, affective dysregulation, and enduring difficulties in relationships. These identity disturbances often involve feelings of emptiness, chronic shame, and fundamental alienation from one's authentic self (67, 68). Therapeutic

approaches like prolonged exposure, while effective for uncomplicated PTSD and some C-PTSD cases (69), may require adaptation for C-PTSD due to its distinct disturbances in self-organization and interpersonal functioning.

The disturbances in identity found in C-PTSD can be understood through AISI as the internalization of the aggressor's degrading second-order model of the victim, a process Ferenczi (57) termed "identification with the aggressor." In prolonged trauma, the victim's second-order self may incorporate the perpetrator's hostile perceptions to minimize free energy in an unpredictable environment. This adaptive strategy during trauma becomes maladaptive post-trauma, leading to fragmented self-models and relational instability.

5 Integration with research domain criteria

The NIMH Research Domain Criteria (RDoC) initiative aims to reconceptualize mental disorders in terms of brain-behavior dimensions (70). While RDoC has advanced biological psychiatry, it has struggled to incorporate the subjective experience and meaning-making processes central to psychodynamic theories. We propose that the AISI framework can bridge these perspectives, demonstrating how traditional psychodynamic concepts can be reformulated in terms of modern computational neuroscience.

5.1 RDoC and AISI: complementary frameworks

RDoC organizes psychopathology along functional domains (negative valence, positive valence, cognitive systems, social processes, arousal/regulatory systems) examined across multiple levels from genes to behavior. The AISI framework's emphasis on bidirectional interactions between interoceptive processes and social inference maps naturally onto these domains:

- **Negative Valence Systems:** AISI reinterprets sustained threat and loss constructs as maladaptive second-order inference patterns, where overly precise prior beliefs about rejection or worthlessness drive affective disturbances.
- **Positive Valence Systems:** Reward processing disruptions are viewed as predictive consequences of negative second-order models, where anticipation of social rejection diminishes reward-seeking behavior.
- **Social Processes:** Social communication and perception difficulties reflect distortions in recursive second-order modeling.
- **Arousal/Regulatory Systems:** Emotion regulation impairments may reflect maladaptive precision weighting in predictive hierarchies, a computational reframing of defensive mechanisms.

Recent work by Banaraki et al. (71) has similarly proposed integrating predictive processing with RDoC domains, focusing on cognitive systems. Our approach extends this integration to encompass psychodynamic insights about self-organization and interpersonal functioning.

5.2 Depression through an integrated lens

As detailed in Section 4.1, depression involves distorted modeling of both interoceptive states and social perceptions. The AISI-RDoC integration offers new insights into traditional psychodynamic depression subtypes, where positive- and negative valence systems are involved:

Anaclitic depression—characterized psychodynamically by dependency and abandonment fears (44)—can be reframed as hypersensitive second-order inference about social acceptance coupled with low precision of interoceptive prediction. This computational profile explains both the intense need for external validation and difficulty regulating emotions independently.

Introjective depression—marked by self-criticism and perfectionism in psychodynamic theory—represents rigid high-precision second-order models of social judgment that resist updating despite contradictory evidence. This corresponds to RDoC's sustained loss construct, where loss-related neural circuits maintain persistent activation patterns (81).

Neurovegetative symptoms reflect the interoceptive consequences of these maladaptive second-order models. Anhedonia emerges not simply from reward circuit dysfunction (RDoC's positive valence systems) but from active inference processes where:

1. Second-order predictions of social rejection increase the expected free energy of reward-seeking behaviors.
2. This leads to adaptive withdrawal to minimize prediction errors.
3. Consequently, dopaminergic signaling, which encodes the precision of reward prediction, is downregulated.

This integrative view explains why pharmacological approaches targeting monoaminergic systems may restore hedonic capacity without addressing the underlying social prediction models driving depression (72).

5.3 PTSD through an integrated lens

Building on Section 4.2, the AISI framework reinterprets PTSD across multiple RDoC domains:

In the negative valence domain, sustained threat activation represents rigid trauma-informed priors that chronically bias perception toward danger detection. The "overgeneralization of conditioned fear" described in RDoC can be reinterpreted through

AISI as overgeneralization of trauma-related second-order inference to non-threatening interpersonal contexts.

The distinction between acute stress disorder and PTSD reflects different stages of inferential adaptation:

1. Acute responses feature low precision of prior beliefs and high precision of sensory/interoceptive errors—an adaptive state that facilitates rapid model updating after unexpected danger.
2. Chronic PTSD develops when trauma-related models become entrenched as high-precision priors that resist revision, biasing perception toward threat detection across contexts.

6 Future directions

The AISI framework opens up numerous opportunities for exploratory research and clinical innovation. Here, we elect to emphasize a significant unmet need: the development of technologies that can facilitate efficacy studies of psychodynamic treatments.

Despite a long tradition of clinical observation and case-based research, psychodynamic interventions remain underrepresented in large-scale clinical efficacy trials. Several challenges hinder progress: the high degree of personalization in treatment, the lack of consensus on therapeutic mechanisms, and the limited availability of sensitive outcome measures.

AISI may provide a novel foundation for overcoming these barriers. By positing recursive inference, particularly second-order cognitive flexibility, as a transdiagnostic mechanism underlying various forms of psychopathology, the framework identifies a common therapeutic target that can be implemented across diagnoses and therapeutic modalities.

Several specific, testable hypotheses emerge from the AISI framework that could guide future research:

1. Individuals with depression will show reduced second-order inference flexibility, measurable through computational tasks that assess belief updating about how others perceive them, with greater rigidity correlating with depression severity.
2. PTSD symptoms will correlate with specific patterns of second-order prediction errors, particularly overestimation of negative social evaluation and underestimation of social support availability.
3. Successful psychotherapy should demonstrate quantifiable increases in second-order inference flexibility, detectable through pre- and post-treatment assessments using the Hierarchical Gaussian Filter or similar computational tool.
4. Precision-weighting abnormalities in second-order inference can be pharmacologically modulated, with serotonergic and dopaminergic agents showing differential effects on different aspects of social prediction updating.
5. Virtual reality-based social interaction paradigms will reveal distinct second-order inference signatures for

different personality disorders, providing objective biomarkers for treatment response.

Emerging technologies in computational psychiatry and AI provide promising tools for testing and refining this model. One such tool is the Hierarchical Gaussian Filter (HGF), a generative model that simulates belief updating under uncertainty (12). HGF can be applied to behavioral data from social inference tasks to estimate parameters like learning rates and precision weighting, offering quantitative measures of second-order flexibility.

We also propose that therapy can be understood as a coupled inference system in which the therapist and patient recursively update their models of each other. These models can be simulated to test how various therapeutic stances (e.g., empathic mirroring vs. interpretive confrontation) shape model convergence and belief revision.

Recent advances in AI-based psychodynamic assessment offer complementary approaches. Hoermann et al. (73) demonstrated that natural language processing algorithms can identify patterns of interpersonal reasoning and second-order perspective-taking in clinical interviews. Similarly, Liébana et al. (74) used interactive AI systems to elicit and quantify social inference flexibility in real-time simulated dialogues. Their findings showed a strong correlation between AI-generated metrics and clinician ratings of interpersonal functioning. Future research must address ethical challenges, such as ensuring equitable access to AI-driven assessments and minimizing risks in psychedelic trials, as discussed in Section 5. Additionally, while AISI emphasizes social and interoceptive contributions to self-experience, it could be extended to incorporate non-social sources like mastery or performance in future work.

These developments suggest that the AISI framework may help integrate diverse research programs—ranging from classical psychoanalysis to contemporary computational modeling—into a unified empirical project. By identifying second-order cognitive flexibility as a mechanistic target and linking it to quantifiable parameters, AISI provides the groundwork for precision psychodynamic psychiatry.

6.1 Empirical approaches to testing AISI

The AISI framework generates testable hypotheses that can be evaluated using computational, behavioral, and neuroimaging methods. Below, we outline specific experimental paradigms to assess AISI's predictions about second-order inference flexibility and its role in psychopathology:

1. **Neuroimaging Studies:** Functional MRI (fMRI) can be used to measure activity and connectivity in AISI-related brain regions (e.g., medial prefrontal cortex [mPFC], anterior insula, temporoparietal junction [TPJ]) during social feedback tasks. For example, participants could engage in a task where they receive positive or negative evaluations from virtual agents, allowing researchers to examine how

- prediction errors in second-order inference correlate with activity in these regions. Reduced mPFC-anterior insula connectivity in MDD or PTSD patients would support AISI's hypothesis of disrupted social inference (63).
2. Computational Modeling with Hierarchical Gaussian Filter (HGF) 80: The HGF (12) can quantify belief updating in second-order inference. In a behavioral task, participants could predict others' perceptions of them based on ambiguous social cues (e.g., neutral facial expressions). HGF parameters, such as learning rate and precision weighting, could reveal reduced inference flexibility in MDD (e.g., rigid negative priors) or PTSD (e.g., overestimation of threat). Comparing these parameters across clinical and control groups would validate AISI's mechanistic claims.
 3. Virtual Reality (VR)-Based Paradigms: VR environments can simulate dynamic social interactions to test AISI's predictions about recursive self-modeling. For instance, a VR task could present participants with avatars delivering ambiguous feedback, measuring how quickly participants update their second-order self-models in response to new information. Such paradigms, inspired by Liébana et al. (74), could quantify inference flexibility and identify biomarkers for treatment response in depression or trauma-related disorders.

These paradigms, summarized in Table 4, offer feasible approaches to test AISI's hypotheses. While large-scale neuroimaging studies may be resource-intensive, pilot studies using HGF or VR tasks provide accessible starting points for empirical validation.

6.2 Treatment implications

While the therapeutic applications of AISI are promising, they remain speculative and will require further empirical research to validate their clinical utility. The AISI framework identifies three independent axes along which patients can be classified: second-order inference flexibility, interoceptive-social coupling strength, and prior precision weighting. These axes, when integrated with RDoC constructs, offer a mechanistic basis for diagnosis, severity

assessment, and treatment selection tailored to underlying psychopathological processes.

For instance, psychodynamically oriented treatments—such as empathic attunement, object relational approaches, transference analysis, or interpersonal therapy—may be particularly indicated for deficits in second-order inference flexibility or maladaptive prior precision weighting. These interventions provide corrective interpersonal experiences that generate prediction errors, facilitating the updating of rigid self-models.

In contrast, modalities like cognitive-behavioral therapy (CBT), exposure therapy, or dialectical behavior therapy (DBT) could be selected for disruptions in interoceptive-social coupling strength, as they target the integration of bodily signals with social cognition through structured behavioral experiments and emotion regulation techniques.

Pharmacotherapies present additional opportunities. Psychedelics, such as psilocybin and MDMA, show potential for addressing second-order inference flexibility and prior precision weighting by temporarily reducing the rigidity of high-level priors (75–77). Other pharmacotherapies, including serotonergic and dopaminergic agents, should be investigated for their differential effects on these three axes, potentially modulating precision weighting in predictive hierarchies (78, 79).

For outcome prediction, AISI parameters—such as baseline second-order inference rigidity—could forecast treatment response. These might be measured through computational tasks (e.g., HGF-based assessments) or AI-driven tools that quantify interpersonal reasoning in clinical interviews (73). This approach supports precision psychiatry by enabling mechanism-based interventions, though prospective studies are needed to confirm these predictions.

7 Conclusion

The Active Intersubjective Inference framework provides a novel and integrative understanding of mental function and dysfunction. By suggesting that the brain constructs not only first-order models of bodily and environmental states but also second-order models regarding how one is perceived by others, it links phenomenological experiences of self, affect, and identity with the neurocomputational principles of predictive coding.

TABLE 4 Proposed experimental paradigms for AISI.

Method	Hypothesis	Task description	Expected outcome	Relevant disorders
fMRI Neuroimaging	AISI-related regions (mPFC, anterior insula, TPJ) show altered activity during second-order inference in MDD/PTSD.	Participants receive positive/negative social feedback in a task, with fMRI measuring regional activity and connectivity.	Reduced mPFC-anterior insula connectivity in MDD/PTSD, reflecting impaired inference flexibility.	MDD, PTSD
Hierarchical Gaussian Filter (HGF)	Reduced second-order inference flexibility correlates with symptom severity.	Behavioral task where participants predict others' perceptions based on ambiguous cues, modeled with HGF.	Lower learning rates and higher prior precision in MDD/PTSD patients vs. controls, indicating rigid self-models.	MDD, PTSD, Personality Disorders
VR-Based Social Interaction	Inference flexibility predicts treatment response in social inference tasks.	VR task with avatars delivering ambiguous feedback, measuring belief updating speed.	Faster belief updating in healthy controls vs. slower updates in MDD/PTSD, with improvements post-therapy.	MDD, PTSD

This framework sheds light on the pathogenesis of conditions such as MDD and PTSD, where recursive self-models and interoceptive processes become rigid, maladaptive, or fragmented. It also explains the relational instability observed in personality disorders, where second-order predictions lack coherence or resilience.

Crucially, AISI suggests that treatment may enhance second-order cognitive flexibility—whether through empathic attunement in psychodynamic therapy, corrective experiences in CBT, or pharmacological modulation of precision weighting. Psychedelics and AI-driven assessments serve as powerful adjuncts for promoting plasticity and personalizing interventions.

By grounding psychodynamic insights in the formal logic of inference, AISI creates a common language for dialogue between clinicians, neuroscientists, and computational modelers. AISI advances psychodynamic understandings of trauma and depression by providing testable computational mechanisms, bridging neurobiological and psychological perspectives in ways that can inform both therapeutic practice and empirical research. This may finally allow for a truly integrative psychiatry—one that respects the depth of subjective experience while embracing the rigor of mechanistic explanation.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

References

1. Seth AK. Interoceptive inference, emotion, and the embodied self. *Trends Cogn Sci.* (2013) 17:565–73. doi: 10.1016/j.tics.2013.09.007
2. Seth AK, Tsakiris M. Being a beast machine: The somatic basis of selfhood. *Trends Cogn Sci.* (2018) 22:969–81. doi: 10.1016/j.tics.2018.08.008
3. Clark A. *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford: Oxford University Press (2016).
4. Craig AD. How do you feel? Interoception: The sense of the physiological condition of the body. *Nat Rev Neurosci.* (2002) 3:655–66. doi: 10.1038/nrn894
5. Friston K. The free-energy principle: A unified brain theory? *Nat Rev Neurosci.* (2010) 11:127–38. doi: 10.1038/nrn2787
6. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active inference: A process theory. *Neural Comput.* (2017) 29:1–49. doi: 10.1162/NECO_a_00912
7. Hohwy J. Attention and conscious perception in the hypothesis testing brain. *Front Psychol.* (2012) 3:96. doi: 10.3389/fpsyg.2012.00096
8. Hohwy J. *The predictive mind*. Oxford: Oxford University Press (2013).
9. Feldman H, Friston K. Attention, uncertainty, and free-energy. *Front Hum Neurosci.* (2010) 4:215. doi: 10.3389/fnhum.2010.00215
10. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. *Front Psychiatry.* (2013) 4:47. doi: 10.3389/fpsy.2013.00047
11. Corlett PR, Honey GD, Fletcher PC. From prediction error to psychosis: Ketamine as a pharmacological model of delusions. *J Psychopharmacol.* (2010) 24:1337–48. doi: 10.1177/0269881109367960
12. Mathys CD, Lomakina EI, Daunizeau J, Iglesias S, Brodersen KH, Friston KJ, et al. Uncertainty in perception and the hierarchical gaussian filter. *Front Hum Neurosci.* (2014) 8:825. doi: 10.3389/fnhum.2014.00825
13. Barrett LF, Simmons WK. Interoceptive predictions in the brain. *Nat Rev Neurosci.* (2015) 16:419–29. doi: 10.1038/nrn3950
14. Seth AK, Friston KJ. Active interoceptive inference and the emotional brain. *Philos Trans R Soc B.* (2016) 371:20160007. doi: 10.1098/rstb.2016.0007
15. Eisenberger NI, Lieberman MD. Why rejection hurts: A common neural alarm system for physical and social pain. *Trends Cogn Sci.* (2004) 8:294–300. doi: 10.1016/j.tics.2004.05.010
16. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *psychol Bull.* (2004) 130:355–91. doi: 10.1037/0033-2909.130.3.355
17. Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: Anatomy, function, and relevance to disease. *Ann New York Acad Sci.* (2008) 1124:1–38. doi: 10.1196/annals.1440.011
18. Gusnard DA, Akbudak E, Shulman GL, Raichle ME. Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proc Natl Acad Sci.* (2001) 98:4259–64. doi: 10.1073/pnas.071043098
19. Carter RM, Huettel SA. A nexus model of the temporal-parietal junction. *Trends Cogn Sci.* (2013) 17:328–36. doi: 10.1016/j.tics.2013.05.007
20. Young L, Camprodon JA, Hauser M, Pascual-Leone A, Saxe R. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc Natl Acad Sci.* (2010) 107:6753–8. doi: 10.1073/pnas.0914826107
21. Buckner RL, Carroll DC. Self-projection and the brain. *Trends Cogn Sci.* (2007) 11:49–57. doi: 10.1016/j.tics.2006.11.004
22. Leech R, Sharp DJ. The role of the posterior cingulate cortex in cognition and disease. *Brain.* (2014) 137:12–32. doi: 10.1093/brain/awt162
23. Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ. The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat Rev Neurosci.* (2011) 12:154–67. doi: 10.1038/nrn2994

Author contributions

HH: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

24. Etkin A, Egner T, Kalisch R. Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends Cogn Sci.* (2011) 15:85–93. doi: 10.1016/j.tics.2010.11.004
25. Critchley HD, Harrison NA. Visceral influences on brain and behavior. *Neuron.* (2013) 77:624–38. doi: 10.1016/j.neuron.2013.02.008
26. Cavanna AE, Trimble MR. The precuneus: A review of its functional anatomy and behavioural correlates. *Brain.* (2006) 129:564–83. doi: 10.1093/brain/awl004
27. Northoff G, Bermpohl F. Cortical midline structures and the self. *Trends Cogn Sci.* (2004) 8:102–7. doi: 10.1016/j.tics.2004.01.004
28. Eichenbaum H. A cortical-hippocampal system for declarative memory. *Nat Rev Neurosci.* (2000) 1:41–50. doi: 10.1038/35036213
29. Schacter DL, Addis DR, Hassabis D, Martin VC, Spreng RN, Szpunar KK. The future of memory: Remembering, imagining, and the brain. *Neuron.* (2012) 76:677–94. doi: 10.1016/j.neuron.2012.11.001
30. Wellman HM, Cross D, Watson J. Meta-analysis of theory-of-mind development: The truth about false belief. *Child Dev.* (2001) 72:655–84. doi: 10.1111/1467-8624.00304
31. Winnicott DW. Transitional objects and transitional phenomena. *Int J Psychoanalysis.* (1953) 34:89–97.
32. Winnicott DW. Ego distortion in terms of true and false self. In: *The maturational process and the facilitating environment.* Hogarth Press, London (1960). p. 140–52.
33. Harris HW, Blue HC, Griffith EEH. *Racial and ethnic identity: Psychological development and creative expression.* New York: Routledge (1995), ISBN: .
34. Erikson EH. *Identity: youth and crisis.* New York: Norton (1968).
35. Freud S. The dynamics of transference. *Standard Edition.* (1912) 12:99–108.
36. Mitchell SA. *Relational concepts in psychoanalysis: An integration.* Cambridge, MA: Harvard University Press (1988).
37. Freud S. Psycho-analytic notes on an autobiographical account of a case of paranoia (dementia paranoides). *Standard Edition.* (1911) 12:1–82.
38. Klein M. Notes on some schizoid mechanisms. *Int J Psycho-Analysis.* (1946) 27:99–110.
39. Kohut H. *The analysis of the self.* New York: International Universities Press (1971).
40. Fan Y, Wonneberger C, Enzi B, de Greck M, Ulrich C, Tempelmann C, et al. The narcissistic self and its psychological and neural correlates: An exploratory fMRI study. *psychol Med.* (2011) 41:1641–50. doi: 10.1017/S003329171000228X
41. Schulze L, Domes G, Krüger A, Berger C, Fleischer M, Pohn K, et al. Neuronal correlates of cognitive reappraisal in borderline patients with affective instability. *Biol Psychiatry.* (2013) 69:564–73. doi: 10.1016/j.biopsych.2010.11.025
42. Gilbert P, McEwan K, Irons C, Bhundia R, Christie R, Broomhead C, et al. Self-harm in a mixed clinical population: The roles of self-criticism, shame, and social rank. *Br J Clin Psychol.* (2010) 49:563–76. doi: 10.1348/014466509X479771
43. Freud S. Mourning and melancholia. *Standard Edition.* (1917) 14:243–58.
44. Blatt SJ. Levels of object representation in anaclitic and introjective depression. *Psychoanalytic Study Child.* (1974) 29:107–57. doi: 10.1080/00797308.1974.11822616
45. Blatt SJ. Experiences of Depression: Theoretical, Clinical, and Research Perspectives. *American Psychological Association.* (2004). doi: 10.1037/10725-000
46. Clark JE, Watson S, Friston KJ. What is mood? A computational perspective. *psychol Med.* (2018) 48:2277–84. doi: 10.1017/S0033291718000430
47. Gilbert P, Allan S, Brough S, Melley S, Miles JNV. Relationship of anhedonia and anxiety to social rank, defeat, and entrapment. *Journal of Affective Disorders.* (2022) 74 (2):141–151. doi: 10.1016/S0165-0327(00)00192-3
48. Golden SA, Covington III, HE, Berton O, Russo SJ. A standardized protocol for repeated social defeat stress in mice. *Nat Protoc.* (2011) 6:1183–91. doi: 10.1038/nprot.2011.361
49. Björkqvist K. Social defeat as a stressor in humans. *Physiol Behav.* (2001) 73:435–42. doi: 10.1016/S0031-9384(01)00490-5
50. Krishnan V, Han MH, Graham DL, Berton O, Renthal W, Russo SJ, et al. Molecular adaptations underlying susceptibility and resistance to social defeat in brain reward regions. *Cell.* (2007) 131:391–404. doi: 10.1016/j.cell.2007.09.018
51. Kube T, Berg M, Kleim B, Herzog P. Rethinking post-traumatic stress disorder – A predictive processing perspective. *Neurosci Biobehav Rev.* (2020) 113:448–60. doi: 10.1016/j.neubiorev.2020.04.014
52. Drevets WC, Price JL, Furey ML. Brain structural and functional abnormalities in mood disorders: Implications for neurocircuitry models of depression. *Brain Structure Funct.* (2008) 213:93–118. doi: 10.1007/s00429-008-0189-x
53. Whitfield-Gabrieli S, Ford JM. Default mode network activity and connectivity in psychopathology. *Annu Rev Clin Psychol.* (2012) 8:49–76. doi: 10.1146/annurev-clinpsy-032511-143049
54. Drevets WC, Savitz J, Trimble M. The subgenual anterior cingulate cortex in mood disorders. *CNS Spectrums.* (2008) 13:663–81. doi: 10.1017/S1092852900013754
55. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders. 5th ed.* Washington, DC: American Psychiatric Publishing (2013).
56. Herman JL. Complex PTSD: A syndrome in survivors of prolonged and repeated trauma. *J Traumatic Stress.* (1992) 5:377–91. doi: 10.1002/jts.2490050305
57. Ferenczi S. Confusion of the tongues between the adults and the child—(The language of tenderness and of passion). *Int J Psychoanalysis.* (1949) 30:225–30.
58. van der Kolk BA, van der Hart O. The intrusive past: The flexibility of memory and the engraving of trauma. *Am Imago.* (1991) 48:425–54.
59. Bromberg PM. *Awakening the dreamer: Clinical journeys.* Mahwah, NJ: Analytic Press (2006).
60. Yehuda R, LeDoux J. Response variation following trauma: A translational neuroscience approach to understanding PTSD. *Neuron.* (2007) 56:19–32. doi: 10.1016/j.neuron.2007.09.006
61. Davis M, Walker DL, Miles L, Grillon C. Phasic vs sustained fear in rats and humans: Role of the extended amygdala in fear vs anxiety. *Neuropsychopharmacology.* (2010) 35:105–35. doi: 10.1038/npp.2009.109
62. Norrholm SD, Jovanovic T, Olin IW, Sands LA, Karapanou I, Bradley B, et al. Fear extinction in traumatized civilians with posttraumatic stress disorder: Relation to symptom severity. *Biol Psychiatry.* (2011) 69:556–63. doi: 10.1016/j.biopsych.2010.09.013
63. Lanius RA, Vermetten E, Loewenstein RJ, Brand B, Schmahl C, Bremner JD, et al. Emotion modulation in PTSD: Clinical and neurobiological evidence for a dissociative subtype. *Am J Psychiatry.* (2010) 167:640–7. doi: 10.1176/appi.ajp.2009.09081168
64. Patel R, Spreng RN, Shin LM, Girard TA. Neurocircuitry models of posttraumatic stress disorder and beyond: A meta-analysis of functional neuroimaging studies. *Neurosci Biobehav Rev.* (2012) 36:2130–42. doi: 10.1016/j.neubiorev.2012.06.003
65. Shin LM, Liberzon I. The neurocircuitry of fear, stress, and anxiety disorders. *Neuropsychopharmacology.* (2010) 35(1):169–191. doi: 10.1038/npp.2009.83
66. Courtois CA. Complex trauma, complex reactions: Assessment and treatment. *Psychotherapy: Theory Research Practice Training.* (2004) 41:412–25. doi: 10.1037/0033-3204.41.4.412
67. Cloitre M, Garvert DW, Brewin CR, Bryant RA, Maercker A. Evidence for proposed ICD-11 PTSD and complex PTSD: A latent profile analysis. *Eur J Psychotraumatol.* (2013) 4:20706. doi: 10.3402/ejpt.v4i0.20706
68. Karatzias T, Shevlin M, Fyvie C, Hyland P, Efthymiadou E, Wilson D, et al. Evidence of distinct profiles of posttraumatic stress disorder (PTSD) and complex posttraumatic stress disorder (CPTSD) based on the new ICD-11 trauma questionnaire (ICD-TQ). *J Affect Disord.* (2017) 207:181–7. doi: 10.1016/j.jad.2016.09.032
69. Sele P, Hoffart A, Bækkelund H, Økstedalen T. Comparing phase-based treatment, prolonged exposure, and skills-training for Complex Posttraumatic Stress Disorder: A randomized controlled trial. *J Anxiety Disord.* (2023) 100:102791. doi: 10.1016/j.janxdis.2023.102791
70. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am J Psychiatry.* (2010) 167:748–51. doi: 10.1176/appi.ajp.2010.09091379
71. Banaraki AK, Toghi A, Mohammadzadeh A. RDoC framework through the lens of predictive processing: focusing on cognitive systems domain. *Comput Psychiatry.* (2024) 8:178. doi: 10.1162/cpsy_a_00100
72. Harmer CJ, Duman RS, Cowen PJ. How do antidepressants work? New perspectives for refining future treatment approaches. *Lancet Psychiatry.* (2017) 4:409–18. doi: 10.1016/S2215-0366(17)30015-9
73. Hoermann S, McCabe KI, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: Systematic review. *J Med Internet Res.* (2023) 25:e29207. doi: 10.2196/29207
74. Liébana D, Roldán M, Morales JA, Muñoz Á, López-Schliching JT. AI-assisted psychometric assessment: Promise, challenges, and ethics. *PLoS One.* (2023) 18:e028104. doi: 10.1371/journal.pone.028104
75. Carhart-Harris RL, Friston KJ. REBUS and the anarchic brain: Toward a unified model of the brain action of psychedelics. *Pharmacol Rev.* (2019) 71:316–44. doi: 10.1124/pr.118.017160
76. Mitchell JM, Bogenschutz M, Lilienstein A, Harrison C, Kleiman S, Parker-Guilbert K, et al. MDMA-assisted therapy for severe PTSD: A randomized, double-blind, placebo-controlled phase 3 study. *Nat Med.* (2021) 27:1025–33. doi: 10.1038/s41591-021-01336-3
77. Gorman I, Belser AB, Jerome L, Hennigan C, Shechet B, Hamilton S, et al. Posttraumatic growth after MDMA-assisted psychotherapy for posttraumatic stress disorder. *J Traumatic Stress.* (2022) 35:236–52. doi: 10.1002/jts.22683
78. Friston K, Thornton C, Clark A. Free-energy minimization and the dark-room problem. *Front Psychol.* (2012) 3:130. doi: 10.3389/fpsyg.2012.00130
79. Cools R, Nakamura K, Daw ND. Serotonin and dopamine: Unifying affective, motivational, and decision functions. *Neuropsychopharmacology.* (2011) 36:98–113. doi: 10.1038/npp.2010.121
80. Friston K, Da Costa L, Hafner D, Hesp C, Ueltzhöffer K. Sophisticated inference. *Neural Comput.* (2022) 34:2150–70. doi: 10.1162/neco_a_01455
81. Drevets WC, Price JL, Furey ML. Neurocircuitry of mood disorders. *Neuropsychopharmacology.* (2011) 36(1):115–134. doi: 10.1038/npp.2010.104