On Valence: A Self-Predictive Processing Model of Emotion Regulation

Yuyue Jiang (yuyuejiang@ucsb.edu)

University of California, Santa Barbara, CA 93106 USA

University of California, Irvine, CA 92612 USA

Abstract

Emotion regulation is a fundamental process that shapes cognitive, affective, and behavioral responses to emotional stimuli. Traditional emotion regulation models conceptualize regulation as a sequential modulation of emotional responses. However, they do not fully explain how emotions are constructed in a way that allows such regulation to occur. Predictive processing (PP) provides a mechanistic framework for understanding emotion generation by proposing that the brain minimizes prediction errors (PEs) to optimize perception and behavior. Yet, standard PP accounts reduce valence to PE minimization, failing to explain how PEs can generate different subjective feelings. To address these limitations, we propose a valence-focused model of emotion regulation that integrates predictive processing with self-referential cognition. We incorporate emotional valence as interpretative processes of the self-model, which assigns emotional significance based on goals, values, and autobiographical context. This model bridges the gap between emotion generation and regulation, highlighting the dynamic interplay between prediction errors, subjective valuation, and self-referential processes. This approach not only advances theoretical understanding but also opens new avenues for computational modeling and empirical research into the adaptive and maladaptive aspects of emotional experience.

Keywords: emotion regulation; predictive processing; valence; self-referential process; cognitive reappraisal

Introduction

Emotion Regulation: The Process Model and the Predictive Processing Model

Emotions are central to human experience, shaping perception, decision-making, and behavior through both conscious and unconscious mechanisms (Panksepp, 1998). A key question in affective science concerns how emotions arise and how they are regulated by human endeavors. Individuals are able to modulate their emotional states to adaptively respond to internal and external demands. One influential framework, the process model of emotion regulation (Gross, 1998), characterizes this process of emotion regulation at various points along the timeline, including situation selection, attentional deployment, cognitive reappraisal, and response modulation (Gross & Thompson, 2007). While this framework has provided a widely used taxonomy of regulatory strategies, it leaves unresolved what is being regulated, namely, the emergence and modulation of affective valence. Without understanding how valence arises, effective regulation remains theoretically underdefined.

Recognizing this gap, Gross (2015) proposed an **extended process model** of emotion regulation, distinguishing between processes of identifying goal-incongruent emotions, selecting a regulatory strategy, implementing that strategy, and monitoring outcomes (Gross, 2015). This extension emphasizes the goal-sensitive evaluative process of alignment of emotional response in emotion regulation. Central to this model is the notion that emotions serve as goal-relevant evaluative signals. Valence, in this context, is conceived as the degree of alignment, or misalignment, between one's current and desired states (Gross, 2015; Lazarus, 1991). The definition of valence as a perceived gap between expectation and reality aligns with the functionalist theories of emotion, in which emotions function as error signals indicating progress (or lack thereof) toward goals.

This discrepancy-based account of emotional valence finds a striking conceptual parallel in predictive processing (PP) theories of the brain. According to the PP framework, the brain continually generates predictions about incoming sensory input and updates its internal models in response to prediction errors, discrepancies between expected and actual outcomes. Some theorists propose that affective valence corresponds to the *sign* or *trajectory* of prediction error, computationally represented by the first and second derivatives of PE. However, this perspective often fails to account for how identical PEs can lead to different emotional responses depending on individual goals and contexts. Newer accounts emerge arguing that goal congruency modulates prediction error and shapes emotional valence, suggesting that affect depends not just on surprise, but on whether outcomes align with current goals (Fromer et al., 2019; Molinaro & Collins, 2023).

Together, these advances reflect a growing convergence between emotion regulation theories and predictive processing frameworks, particularly in how they conceptualize valence as an emergent property of discrepancy, generated by an internal comparator that monitors how far the organism is from a settled, ideal, or homeostatic state (defined in motivational terms for Gross, or in predictive terms for PP). However, despite this shared foundation, these models remain narrowly outcome-focused,

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

anchored in predefined goals and prediction targets. This functionalist orientation, while computationally elegant, oversimplifies the nature of valence, reducing it to a signal of success or failure relative to external benchmarks(Carruthers, 2018). As a result, they overlook the deeper mechanistic and evaluative dimensions of emotional experience.

The Challenge of Valence Regulation

Understanding valence is especially important for emotion regulation because valence is often the very feature being regulated. Effective regulation aims to transform how an emotion feels at its core, often involves altering the meaning of emotional stimuli, an approach that is more consequential than modifying intensity alone (Ochsner et al., 2012).

In real-world contexts, emotional experiences frequently diverge from the predictions of outcome-centric or error-based models. Individuals may feel distress after achieving a desired goal, or relief in the wake of failure, phenomena that simple discrepancy computations cannot account for. Empirical studies have documented such cases across domains, including reward learning, self-evaluation, and decision-making under uncertainty (e.g., Wood et al., 2003; Wilson & Ross, 2001). These findings point to the involvement of higher-order appraisal processes, particularly those tied to self-relevance, coherence, and meaning-making. This calls for a deeper understanding of the *appraisal mechanisms* that shape emotional experience. In particular, it suggests that valence emerges not just from low-level computations of mismatch, but from higher-order evaluations that integrate self-related information. LeDoux's Higher-Order Theory of Emotional Consciousness (HOTEC) offers a compelling account: emotions become consciously experienced and acquire valence only when they are represented in relation to the self. Affective consciousness depends on higher-order self-models that interpret lower-level bodily or cognitive signals. Neuroimaging studies evidenced that regions involved in self-referential processing (e.g., medial prefrontal cortex, posterior cingulate cortex) are consistently engaged during the evaluation of emotional significance (D'Argembeau, 2013; Northoff et al., 2006). Moreover, affective prediction is modulated by autobiographical memory and trait-level self-knowledge (Kelley et al., 2002; Andrews-Hanna et al., 2014), suggesting that emotional valence is filtered through internal models of identity and coherence. Together, these insights suggest that incorporating self-referential processes into predictive models of emotion may offer a promising path toward a more complete and mechanistic account of valence construction.

In what follows, we present a **Self-Predictive Processing (Self-PP) model** that accounts for the generation of emotional valence and its regulation. The Self-Predictive Processing (Self-PP) model extends predictive processing by incorporating the self-model as a crucial layer in evaluating emotional events and their regulation. In this framework, the brain doesn't simply compute discrepancies between external outcomes and goal states, it also monitors mismatches between events and internal self-predictions rooted in one's identity, beliefs, and personal narrative. Emotional experiences are thus appraised on two levels: (1) Outcome value, is the event consistent with my goals and needs? and (2) Self-model coherence, does this align with how I predict the world and myself to be? The Self-PP model also provides a unified account of the regulation of emotional valence and its regulation: regulatory strategies act at multiple hierarchical levels,from raw sensory prediction to high-level beliefs about the self. Rather than rejecting prior accounts such as Gross's extended model, Self-PP offers a mechanistic substrate that complements them, showing converging reasoning that certain strategies, like reappraisal or self-distancing, are more effective. These strategies work by modifying self-referential predictions that imbue an experience with valence, and their success varies across contexts and individuals precisely because the content and structure of the self-model vary.

A Self-PP Model of Emotional Valence

The Self is the central agent for evaluating relevance, setting regulatory goals, integrating autobiographical context, enabling metacognitive reflection, assigning valence, and dynamically updating the self-model (Northoff et al., 2011; Seth, 2013). The self-model serves as the mechanism through which emotional significance is assigned, allowing individuals to reconstruct valence based on personal relevance and goal. Emotion regulation often involves revising the self-model to reduce emotional distress. This process requires the Self to integrate information relating to subjective values and recalibrate predictions in a way that aligns with one's identity and goals (Lane et al., 2015). If valence arises from the interpretive functions of the self-model, we should expect individual differences in self-referential cognition to predict variability in emotion regulation efficacy. Supporting this claim, studies have shown that self-referential processing in the medial prefrontal cortex (mPFC) and default mode network (DMN) modulates emotional responses by integrating autobiographical context (Denny et al., 2014; Roy et al., 2012). Additionally, research on cognitive reappraisal suggests that reinterpreting an event in self-relevant terms enhances regulatory success (Kross et al., 2014). By combining PP's hierarchical error minimization with self-processing, we developed a model that explains emotion processing as a dynamic, hierarchical process that integrates sensory perception, predictive modeling, and self-referential cognition. With such a model, we can explain how valence is actively reconstructed (e.g., reappraisal updates self-relevant predictions) and gives rise to conscious emotional experience.

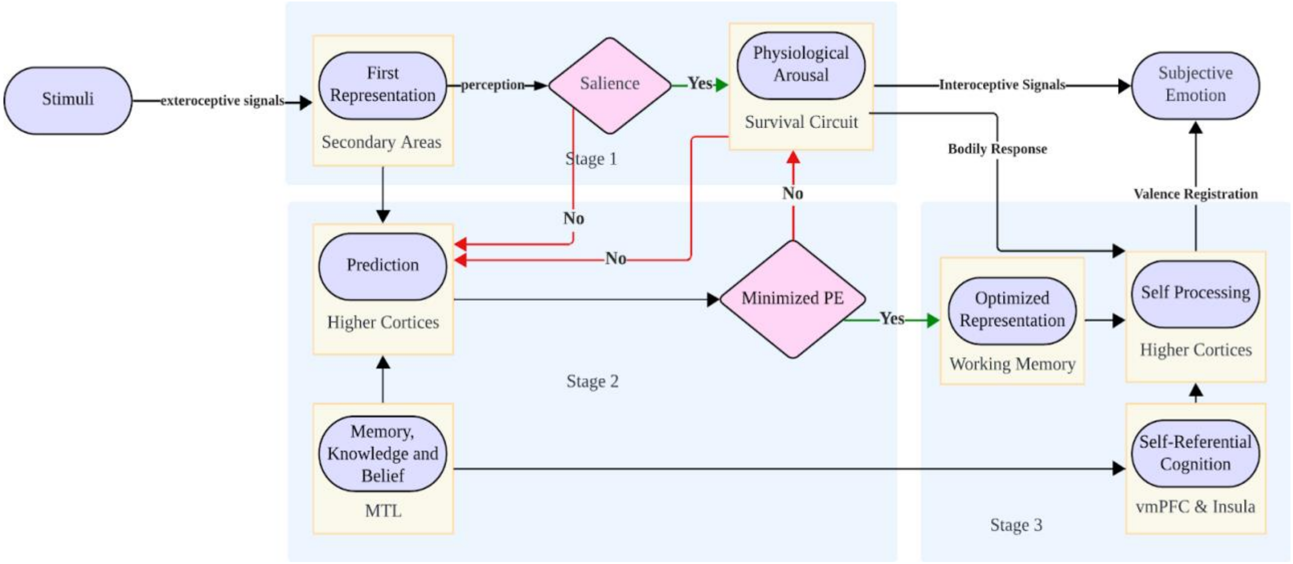

Figure 1: A three-stage predictive processing framework for emotion generation and valence assignment. **Stage 1 (left):** Raw interoceptive and exteroceptive signals produce a *First Representation* of stimuli in secondary sensory areas. A quick check for Salience determines if a stimulus should trigger an immediate Physiological Arousal via a survival circuit (pink path, “Yes” branch). If highly salient (threat or urgent), a rapid autonomic response is initiated (bypassing detailed predictions). If not (red “No” paths), processing continues to Stage 2. **Stage 2 (center):** Higher cortices generate a context-based Prediction for the stimulus and compare it to the input. Mismatches produce Prediction Errors (PEs). The system iteratively updates its predictions (red feedback arrows) using memory and knowledge (medial temporal lobe, “MTL”) until PEs are minimized. If PEs cannot be minimized to a satisfactory level (persistent surprise), this signals the need for further interpretation or model adjustment. Once an Optimized Representation is achieved (minimal PE; green arrow), it is loaded into working memory. **Stage 3 (right):** The optimized representation is integrated with Self-Processing in higher-order cortices (mPFC, etc.) and Self-Referential Cognition (ventromedial PFC & insula) to evaluate personal relevance. If deemed self-relevant, Subjective Emotion is generated with a registered valence (positive or negative) and conscious feeling.

Stage 1: Perception, Prediction, and Salience Detection.

Emotion construction begins with raw sensory input, both interoceptive and exteroceptive, processed by early cortical and subcortical systems. First-order representations are formed in primary and secondary sensory areas, supported by interoceptive regions such as the posterior insula (Barrett & Simmons, 2015). Simultaneously, the represented signals are being compared at multiple levels with top-down predictions generated from memory and contextual models in higher cortical areas, including the hippocampus and prefrontal cortex. When a stimulus is deemed urgent or survival-relevant, such as a threatening object or high-value reward, it may trigger fast autonomic responses via subcortical survival circuits, notably the amygdala and brainstem (LeDoux, 2014; Pessoa & Adolphs, 2010). These "low road" responses bypass extended inferential processing and generate physiological arousal and primitive affective states (e.g., fear, alarm) as rapid heuristics for action. However, if immediate salience is not detected, the system proceeds to further inference and interpretation.

Stage 2: Prediction, Comparison, and Error Minimization.

Mismatches between expectation and input produce prediction errors (PEs), engaging salience-monitoring and error-signaling networks such as the ACC, insula, and amygdala (Uddin et al., 2014; Schultz et al., 1997). These signals prompt iterative model refinement until an optimized representation with a minimized PE is reached. The result of this process is an optimized representation: a contextually appropriate, semantically coherent understanding of what is happening. Crucially, however, the representation remains affectively indeterminate. The brain now "knows" what the stimulus is, but not yet what it means for the self. The interpretive work of this stage is necessary but not sufficient for emotional valence. The optimized representation must be evaluated in light of the self, its goals, beliefs, and values, before it can be labeled

Stage 3: Self-Processing and Valence Assignment.

The third stage operates at a metacognitive layer with the self-model. Here, the brain takes the optimized representation from Stage 2 and evaluates it in relation to the self according to individual goals. Belief and propositional attitudes about oneself. This evaluative step occurs through interaction of higher-order cortical networks involved in self-referential processing (mPFC, PCC, etc.) with limbic and interoceptive areas (Roy et al., 2012; Amodio & Frith, 2006). The integration of autobiographical memory and self-related knowledge provides the context that frames the meaning of the situation for the individual. Specifically, the integration of abstract, self-relevant constructs (e.g., values, identity) enables the registration of valence. By the end of Stage 3, the emotional episode is fully constructed: the person has an emotional reaction with a certain valence (positive/negative), arousal level, and qualitative character (perhaps even categorized as a specific emotion like *anger* or *relief*). This three-stage framework is summarized in **Figure 1**. Without this final, self-referential layer, predictive processing cannot explain why the same stimulus might be experienced as exhilarating by one individual and threatening by another. The Self-PP model thus reframes emotion as a multilevel inferential process, in which hierarchical predictive computations are ultimately made subjectively meaningful through the lens of the self. It is at this final stage that emotion becomes not just a reaction, but a personalized and regulated experience, integrated into one's broader narrative of who they are and what matters to them.

A Self-PP model of Emotion Regulation

Following the aforementioned model, we further propose that emotion can be regulated during or after the generative stages. *Emotion regulation* is reconceptualized as altering the predictive processing stream at specific levels so as to change the resulting emotional experience. Because our model defines multiple stages (levels of representation) in emotion generation, we can categorize regulation strategies by **which stage of processing they target**. This provides a hierarchical view of regulation: from low-level physiological interventions to high-level reinterpretations of meaning.

Critically, each level of intervention has different properties in terms of **flexibility**, **context-dependence**, and **energetic cost**. Higher-order strategies, such as self-referential reinterpretation, allow individuals to modify emotional significance without directly altering previously formed predictive models. Modulations on self-referential and conceptual processes are more flexible than low-level sensory mechanisms because they reinterpret, reframe, and restructure emotional meaning (Etkin et al., 2015). In contrast, Early-stage interventions, such as physiological modulation, operate on pre-conscious mechanisms and require restarting the model in order to engage with autonomic responses. Such a pathway of regulation is constrained by the involuntary nature of physiological and survival-based responses, making it more challenging and computationally costly (LeDoux, 2014). Even if such physiological modulation is done successfully, the brain's cognitive and self-referential representation of the emotion may remain unchanged, creating a discrepancy between the lower-level signals and higher-order representations, leaving the individual in a dissonance. The model highlights how Stage 3 (self-focused) regulation provides a particularly adaptive top-down approach that is convergent with current consensus in research on emotion regulation strategies ' effectiveness (Nook et al., 2017).

Stage 1 Regulation: Arousal Modulation

The first stage of regulation operates at the level of first-order processing, where interoceptive and exteroceptive signals are initially integrated to form raw sensory representations (Seth, 2013). These representations serve as the foundation for higher-order predictions and are pre-conceptual, thus are absent of cognitive interpretation or self-relevance (Paulus & Stein, 2010). Interventions at this stage focus primarily on modulating physiological inputs rather than altering the cognitive interpretation or self-relevance of an emotional experience (Craig, 2009). Two primary mechanisms are central to regulation at this level: Interoceptive Recalibration and Sensory Attenuation, which are all subject to the strategy of suppression. Interoceptive Recalibration modulates internal bodily states in an attempt to reduce interoceptive prediction errors (e.g., through paced breathing or biofeedback). By stabilizing physiological signals (e.g., lowering heart rate or slowing respiration), this approach recalibrates the body's homeostatic setpoints, preventing excessive autonomic responses that could amplify emotional intensity (Khalsa et al., 2018; Schulz & Vögele, 2015). Similarly, Sensory Attenuation reduces exteroceptive prediction errors through repeated exposure to emotionally salient stimuli, gradually diminishing their capacity to generate surprise or hyperarousal (Craske et al., 2014). Despite their efficacy in regulating arousal, these strategies have inherent limitations. First-order regulation does not engage change in either the cognitive reappraisal (stage 2) or the self-referential processing (stage 3), meaning that while fear or distress may be physiologically mitigated, the subjective feeling of the experience remains intact (Farb et al., 2015). For instance, while paced breathing may lower heart rate during anxiety, it does not change the appraisal that a given situation is threatening (Gross, 2002). As a result, the regulation strategy of suppression, the act of inhibiting an emotional response, is computationally costly at this stage because it requires sustained top-down control to override autonomic responses while keeping the prediction and self-reference unaltered. This demand for ongoing inhibition can paradoxically increase physiological arousal due to the effort to change physiological responses while higher-order predictions remain unchanged. Thus, prediction errors increase beneath the surface, awaiting resolution at higher processing stages (Webb et al., 2012).

Stage 2 Regulation: Cortical Prediction Optimization

The second stage of emotion regulation plays a crucial role in shaping how an emotional event is interpreted, forming the cognitive bridge between raw sensory input and self-referential evaluations (Etkin et al., 2015; Buhle et al., 2014). At this stage, emotional experiences are constructed through higher-order representations (HORs) that integrate sensory information, prior experiences, and contextual knowledge to form a refined emotional prediction (Kober et al., 2008; Brown & Ledoux, 2017). This process is supported by a network of brain regions, including the prefrontal cortex

(PFC), anterior cingulate cortex (ACC), amygdala, and hippocampus, which work together to update and optimize emotional predictions. Unlike Stage 3, this stage does not yet incorporate self-relevance. Instead, it focuses on modifying the conceptualization of the event itself, ensuring that the emotional response aligns with an individual's broader regulatory goals (Sheppes et al., 2014). The brain recalculates its predictions by introducing or focusing on previously overlooked contextual inputs that influence the categorization and interpretation of the emotional event (Barrett & Satpute, 2013; Lindquist et al., 2011). This process computes an updated prediction that explicitly incorporates the regulation goal as a constraint, ensuring that the revised emotion aligns with the individual's desired response (Denny et al., 2014). Consequently, prediction errors are recalibrated to reflect this newly optimized representation, and the revised prediction is relayed to Stage 3 for final self-relevance evaluation and integration into conscious emotional awareness (Smith et al., 2018). Key regulatory strategies at this stage include cognitive reappraisal, attention deployment, and mindfulness practices, each supported by distinct neural mechanisms. Cognitive reappraisal, which involves reinterpreting the meaning of an emotion-evoking stimulus, engages the dorsolateral prefrontal cortex (dlPFC) and ventrolateral prefrontal cortex (vlPFC) to modulate amygdala activity, reducing prediction errors and emotional intensity (Buhle et al., 2014). The ACC functions as a hub that monitors the reappraisal and signals the need for further adjustments, ensuring that the updated prediction aligns with regulatory goals. Attention deployment is associated with vlPFC activation, which is thought to serve the function of disengaging with the current emotional state, and thus effectively lowers the sensory precision by reducing the salience of the event (Gronau et al., 2017). Mindfulness practices reduce the weight of maladaptive priors (e.g., "I always fail") by engaging the dlPFC to inhibit activity in the default mode network (DMN), particularly in the PCC and medial PFC (mPFC). This inhibition allows for more flexible emotional responses and adaptive prediction updating (Hölzel et al., 2011; Lutz et al., 2014). Simultaneously, the hippocampus integrates contextual information and prior experiences that are relayed to higher cortices for emotional predictions. It helps recalibrate predictions by retrieving relevant memories and contextual cues, ensuring that the updated prediction aligns with the individual's regulatory goals (Barrett & Satpute, 2013). Together, these neural mechanisms enable the brain to refine emotional predictions, minimize prediction errors, and ensure that the emotional response aligns with the individual's desired regulatory outcome.

Stage 3 Regulation: Self-Relevance Evaluation

The third stage, self-processing, is particularly influential in determining the efficacy of regulation and is the most malleable. At this stage, individuals operate on higher-order representations (HORs) that have already been distilled from earlier processing stages. Rather than expending energy on encoding new sensory information, the brain primarily allocates resources to integrating these representations with autobiographical knowledge to assess their valence through self-relevant processing. Once the self-relevance of a stimulus is diminished, the emotional response can be readily altered. For example, while it may be difficult for an individual to reconstruct the view of having seen a repulsive image or seeking to find positive meaning from this event, it is relatively easy for them to conclude that the image is irrelevant to themselves, thereby quickly inhibiting the emotional reaction. At this stage, the prediction itself remains unchanged; instead, it is the relevance of the prediction's content that is recalibrated. This allows prediction errors (PEs) to remain at their lowest or most optimized level, eliminating the need for recomputation of the prediction and conserving cognitive resources. The neural mechanisms underlying these regulatory processes are anchored in key brain regions, including the ventromedial prefrontal cortex (vmPFC), default mode network (DMN), and anterior insula. The vmPFC, integral to updating self-referential priors and value-based decision-making, interacts with the DMN, a system governing autobiographical reflection and self-projection, to modulate the abstract, subjective dimensions of self-relevance. The anterior insula, sensitive to interoceptive and emotional salience, provides feedback that informs these metacognitive adjustments (Craig, 2009).

Several regulation strategies target the self-model's predictions to recalibrate emotional valence, either by recalibrating self-relevance through positive reinterpretation or by diminishing self-relevance to reduce emotional intensity. To recalibrate self-relevance, strategies aim to restart the evaluation process with the goal of finding positive meaning as a constraint. For example, metacognitive Reframing functions by revising deeply-held self-model priors. By decoupling arousal from maladaptive self-relevance, it neutralizes emotions like shame (Lane et al., 2015). Similarly, narrative Reframing rewrites the self-model's interpretations by integrating new, adaptive meanings into autobiographical memory through interactions between the default mode network (DMN) and vmPFC (McAdams, 2001). A different type of strategy functions by weakening the binding of higher-order representations (HORs) to the self-model, thereby lowering emotional intensity. By adopting a third-person perspective, the strategy of Self-Distancing dilutes emotional intensity through cognitive detachment, effectively reducing the self-relevance of the experience (Kross et al., 2014). Cognitive Defusion disentangles thoughts from self-identity (e.g., "I have the thought that I am a failure" vs. "I am a failure"), reducing their emotional grip by reclassifying them as transient mental events rather than truths (Hayes et al., 2012).

The hierarchical structure of the three stages of emotion regulation suggests that effective regulation follows a top-down trajectory, with self-referential adjustments serving as the most adaptive regulatory mechanism. While emotion emerges from the bottom-up integration of sensory, cognitive, and self-referential processing, regulation occurs in a top-down cascade, starting with higher-order modulation before engaging lower-level systems. This inverse relationship highlights the efficiency of self-directed regulation strategies, where changing one's *mindset or narrative* can rapidly resolve emotional distress with minimal physiological struggle, whereas fighting one's bodily reactions while still believing the distressing interpretation is akin to swimming upstream, an results that aligns with empirical evidence (Butler et al., 2007). By consciously restructuring how emotion is conceptualized within the self-model, individuals can modulate emotional responses without necessitating an effortful, direct recalibration of lower-level signals via response modulation, thereby avoiding the cognitive dissonance and leading to more effective regulation (Gross, 2022).

While the Self-PP model underscores the potency of top-down strategies, it does not presume their universal superiority. Depending on the qualities of the context and stimuli, bottom-up responses (Stage 1) may be essential. Regulatory flexibility (Bonanno & Burton, 2013) suggests that optimal strategy use is context dependent. Our model clarifies this: higher-stage interventions recalibrate valence more deeply when time and cognitive resources allow; lower-stage tactics maintain stability when they do not. Individual differences in regulatory proficiency further shape which strategies are accessible or effective, shaped in part by the structure of the self-model.

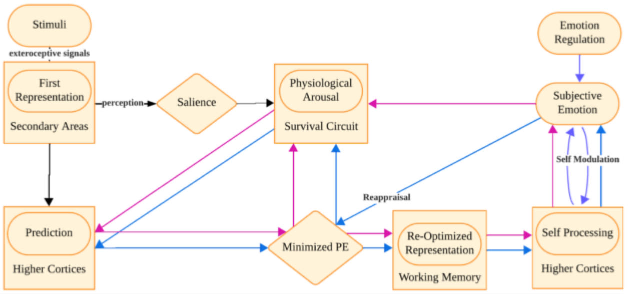

Figure 2: A model of emotion regulation pathways across the three generative stages. **Stage 1 – Physiological Modulation (pink pathways):** Regulation targets the initial sensory and bodily response. This includes modifying interoceptive signals or external stimuli (*pink arrows*) **Stage 2 – Cognitive Reframing (blue pathways):** Regulation influences the predictive and interpretive processes. Techniques like reappraisal or attentional deployment introduce new information or perspectives (*blue arrows*) into the Prediction and Re-Optimized Representation nodes (working memory and higher cortex), so that the Minimized PE reflects an updated understanding aligned with one's goals. **Stage 3 – Self-Relevance Reevaluation (purple pathways):** Higher-order regulation engages self-referential cognition. Strategies such as self-distancing or metacognitive reframing adjust the link between the Self Processing node and the Subjective Emotion (*purple arrows*), altering valence by changing how much the self-model is involved in the emotional appraisal (self-modulation).

Conclusion

In this paper, we have outlined a theoretical model that integrates predictive processing with self-referential cognition to explain both the generation and regulation of emotion. Our Self-Predictive Processing model proposes that emotional experiences are not merely passive readouts of bodily states or hard-wired responses, but active constructions that emerge from the brain's attempts to predict the world and interpret those predictions in light of the self. By embedding the self and valuation into PP, our account extends prior predictive models of emotion (Joffily & Coricelli, 2013; Seth, 2013) which lacked a mechanism for qualitative experience, and it complements constructivist theories (Barrett, 2017) by providing a process-level description of how core affect can be shaped by personal context. It also integrates with Gross's extended process model by offering a mechanistic basis for the identification and selection stages: the self-model's detection of goal incongruence could correspond to identifying the need to regulate, and the generative model's multi-level structure suggests which types of strategies (targets) might be selected.

There are several exciting directions to extend and empirically ground the Self-PP model. One direction is to investigate the neural interactions among the networks implicated in our framework: how the ventromedial prefrontal cortex (vmPFC), default mode network (DMN), and anterior insula interact to dynamically update self-relevant priors and assign emotional valence (Northoff et al., 2011; Seth, 2013). We hypothesize there is a bidirectional interplay such that activity in self-related regions can attenuate or amplify prediction-error signals depending on appraisal. This proposed framework also opens promising avenues for future computational studies of emotion. Due to constraints of scope, we did not elaborate on a formal computational architecture within this paper. However, an implementation of the Self-PP model, including a process-level simulation and network structure, is presented in Jiang & Luo (2024). A formal mathematical treatment of valence computation grounded in predictive error dynamics and self-relevant inference is described in Jiang & Luo (2025). Together, these developments aim to provide a computationally tractable and empirically testable foundation for understanding how affective experiences emerge, acquire meaning, and become subject to regulation. self-models, as it provides a promising avenue to advance theoretical understanding and enables the development of more sophisticated computational models of human affect, ultimately providing deeper insights into the adaptive and maladaptive aspects of emotional experiences.

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*(4), 268–277.

Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: Component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences, 1316*(1), 29–52.

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20–46.

Barrett, L. F. (2017). *How Emotions Are Made: The Secret Life of the Brain*. Houghton Mifflin Harcourt.

Barrett, L. F., & Satpute, A. B. (2013). Large-scale brain networks in affective and social neuroscience: Towards an integrative functional architecture of the brain. *Current Opinion in Neurobiology, 23*(3), 361–372.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16*(7), 419–429.

Bonanno, G. A., & Burton, C. L. (2013). Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on Psychological Science, 8*(6), 591–612.

Buhle, J. T., Silvers, J. A., Wager, T. D., Lopez, R., Onyemekwu, C., Kober, H., ... & Ochsner, K. N. (2014). Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cerebral Cortex, 24*(11), 2981–2990.

Butler, E. A., Lee, T. L., & Gross, J. J. (2007). Emotion regulation and culture: Are the social consequences of emotion suppression culture-specific? *Emotion, 7*(1), 30–48.

Carruthers, P. (2018). Valence and value. *Mind & Language, 33*(2), 130–147.

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

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

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10*(1), 59–70.

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy, 58*, 10–23.

D'Argembeau, A. (2013). On the role of the ventromedial prefrontal cortex in self-processing: The valuation hypothesis. *Frontiers in Human Neuroscience, 7*, 372.

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2014). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience, 26*(3), 411–427.

Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature Reviews Neuroscience, 16*(11), 693–700.

Farb, N. A., Anderson, A. K., & Segal, Z. V. (2015). The mindful brain and emotion regulation in mood disorders. *Canadian Journal of Psychiatry, 60*(6), 259–263.

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

Frömer, R., Dean Wolf, C. K., & Shenhav, A. (2019). Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making. *Nature Communications, 10*(1), 4926.

Gaál, Z. A., Simor, P., & Czigler, I. (2024). Working memory and emotion regulation: A review of the neural mechanisms. *Neuroscience & Biobehavioral Reviews, 151*, 104927.

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2017). A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: The case of felt power. *Comprehensive Results in Social Psychology, 2*(1), 123–138.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*(3), 271–299.

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology, 39*(3), 281–291.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry, 26*(1), 1–26.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*(2), 348–362.

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of Emotion Regulation* (pp. 3–24). Guilford Press.

Gross, J. J., Uusberg, H., & Uusberg, A. (2019). Mental illness and well-being: An affect regulation perspective. *World Psychiatry, 18*(2), 130–139.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and Commitment Therapy: The Process and Practice of Mindful Change*. Guilford Press.

Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist, 52*(12), 1280–1300.

Hölzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R., & Ott, U. (2011). How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on Psychological Science, 6*(6), 537–559.

Jiang, F., & Luo, D. (2024). Implementing self-models through joint-embedding predictive architecture. In *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*.

Jiang, Y., & Luo, D. (2025). Valence Computation as Higher-order Inference via Conceptual Self-Processing. OSF.

Joffily, M., & Coricelli, G. (2013). Emotional valence and the free-energy principle. *PLoS Computational Biology, 9*(6), e1003094.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study. *Journal of Cognitive Neuroscience*, *14*(5), 785–794.

Khalsa, S. S., Adolphs, R., Cameron, O. G., Critchley, H. D., Davenport, P. W., Feinstein, J. S., ... & Paulus, M. P. (2018). Interoception and mental health: A roadmap. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(6), 501–513.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage, 42*(2), 998–1031.

Kross, E., Ayduk, O., & Mischel, W. (2014). When asking "why" does not hurt: Distinguishing rumination from reflective processing of negative emotions. *Psychological Science, 22*(9), 107–111.

Lane, R. D., Ryan, L., Nadel, L., & Greenberg, L. (2015). Memory reconsolidation, emotional arousal, and the process of change in psychotherapy: New insights from brain science. *Behavioral and Brain Sciences, 38,* e1.

Lazarus, R. S. (1991). *Emotion and Adaptation*. Oxford University Press.

LeDoux, J. E. (2014). Coming to terms with fear. *Proceedings of the National Academy of Sciences, 111*(8), 2871–2878.

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences, 114*(10), E2016–E2025.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2011). The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences, 35*(3), 121–143.

Lutz, A., Slagter, H. A., Dunne, J. D., & Davidson, R. J. (2014). Attention regulation and monitoring in meditation. *Trends in Cognitive Sciences, 12*(4), 163–169.

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

McRae, K., Misra, S., Prasad, A. K., Pereira, S. C., & Gross, J. J. (2012). Bottom-up and top-down emotion generation: Implications for emotion regulation. *Social Cognitive and Affective Neuroscience, 7*(3), 253–262.

Molinaro, G., & Collins, A. G. E. (2023). A goal-centric outlook on learning. *Trends in Cognitive Sciences, 27*(12), 1150–1164.

Mobbs, D., Adolphs, R., Fanselow, M. S., Barrett, L. F., LeDoux, J. E., Ressler, K., & Tye, K. M. (2019). Viewpoints: Approaches to defining and investigating fear. *Nature Neuroscience*, *22*(8), 1205–1216.

Molinaro, G., & Collins, A. G. E. (2023). Intrinsic rewards explain context-sensitive valuation in reinforcement learning. *PLOS Biology, 21*(7), e3002201.

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, *1*(12), 881–889.

Northoff, G. (2016). Is the self a higher-order or fundamental function of the brain? The "basis model of self-specificity" and its encoding by the brain's spontaneous activity. *Cognitive Neuroscience, 7*(1-4), 203–222.

Northoff, G., Qin, P., & Nakao, T. (2011). Rest–stimulus interaction in the brain: A review. *Trends in Cognitive Sciences, 15*(6), 314–322.

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

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press.

Paulus, M. P., & Stein, M. B. (2010). Interoception in anxiety and depression. *Brain Structure and Function, 214*(5–6), 451–463.

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience, 9*(2), 148–158.

Pessoa, L., & Adolphs, R. (2010). Emotion processing and the amygdala: From a 'low road' to 'many roads' of evaluating biological significance. *Nature Reviews Neuroscience, 11*(11), 773–783.

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal–subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences, 16*(3), 147–156.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion, 23*(7), 1307–1351.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573.

Sheppes, G., Scheibe, S., Suri, G., & Gross, J. J. (2014). Emotion-regulation choice. *Psychological Science, 25*(4), 1150–1158.

Smith, R., Lane, R. D., & Nadel, L. (2018). Unconscious emotion: A cognitive neuroscientific perspective. *Neuroscience & Biobehavioral Reviews, 86*, 1–14.

Uddin, L. Q., Kinnison, J., Pessoa, L., & Anderson, M. L. (2014). Beyond the tripartite cognition–emotion–

interoception model of the human insular cortex. *Journal of Cognitive Neuroscience, 26*(1), 16–27.

Webb, T. L., Miles, E., & Sheeran, P. (2012). Dealing with feeling: A meta-analysis of the effectiveness of strategies derived from the process model of emotion regulation. *Psychological Bulletin, 138*(4), 775–808.

Wilson, A. E., & Ross, M. (2001). From chump to champ: People's appraisals of their earlier and present selves. *Journal of Personality and Social Psychology, 80*(4), 572–584.

Winecoff, A., Labar, K. S., Madden, D. J., Cabeza, R., & Huettel, S. A. (2013). Cognitive and neural contributors to emotion regulation in aging. *Social Cognitive and Affective Neuroscience, 6*(2), 165–176.

Wood, J. V., Taylor, S. E., & Lichtman, R. R. (2003). Social comparison in adjustment to breast cancer. *Journal of Personality and Social Psychology, 49*(5), 1169–1183.

Yih, J., Uusberg, A., Taxer, J. L., & Gross, J. J. (2019). Better together: A unified perspective on appraisal and emotion regulation. *Cognition and Emotion, 33*(1), 41–47.