Predict to Control, Test to Master: Integrating Predictive Processing and Control–Mastery Theory in Understanding How Psychotherapy Works

Elizabeth Li, James McCollum, Josh Krieger, Sophia E. Winter, Daniel Duane, and George Silberschatz

Predict to Control, Test to Master: Integrating Predictive Processing and Control–Mastery Theory in Understanding How Psychotherapy Works

Elizabeth Li[1, 2], James McCollum[3], Josh Krieger[3], Sophia E. Winter[3], Daniel Duane[3], and George Silberschatz[3, 4]

[1] Research Department of Clinical, Educational and Health Psychology, University College London (UCL)
[2] Anna Freud Centre, London, United Kingdom
[3] San Francisco Psychotherapy Research Group (SFPRG), San Francisco, California, United States
[4] Department of Psychiatry, University of California San Francisco (UCSF)

This article explores how integrating predictive processing framework and control–mastery theory offers a novel account of how psychotherapy works. Predictive processing proposes that the brain continually generates predictions about sensory inputs based on prior experiences to anticipate and respond effectively to its environment. When actual experiences differ, prediction errors prompt updates to the brain's internal model of the world, thereby improving future predictions and supporting adaptive behavior. In cases of psychopathology, however, entrenched negative predictions may persist, offering temporary reduction in uncertainty despite being ultimately harmful. This short-term predictability and stability may outweigh change processes, leading to persistent distress. Yet, predictive processing leaves open the question of why some individuals seek therapy to challenge such beliefs, despite their bringing a semblance of security. Control–mastery theory complements predictive processing by emphasizing the patient's intrinsic motivation and active role in challenging and revising pathogenic beliefs through patient testing in psychotherapy. We argue that effective psychotherapy depends not only on the therapist's provision of experiences that generate salient prediction errors, but also on the patient's motivation and readiness to engage in revising deeply ingrained beliefs. Drawing on a clinical case of a woman in her 30s, we illustrate how early trauma shapes maladaptive generative models, constraining relational expectations. In therapy, attuned responses generate salient prediction errors, disconfirming maladaptive priors and updating generative models. Through implicit testing and integration of disconfirmatory experiences, the patient revises her internal predictive model, reducing maladaptive prediction errors and fostering therapeutic change.

Public Health Significance Statement
Many people remain stuck in pathogenic patterns because they once provided a sense of predictability and controllability in a harmful environment, making change difficult. This article integrates predictive processing and control–mastery theory to explain how psychotherapy helps patients challenge these beliefs and regain mastery over their lives, not only through the therapist's provision of new experiences that generate prediction errors but also by recognizing and responding to the patient's intrinsic motivation to test and revise deeply ingrained expectations. Understanding these mechanisms can enhance psychotherapy and improve mental health outcomes.

A therapist worked for many years with a young male patient who had suffered terribly from the enduring effects of early childhood trauma and emotional neglect. This patient remembered his mother and father as cruelly indifferent to his emotional needs, a neglect so painful that he protected himself by never asking for any emotional support at all. Throughout adolescence and early adult life, this young man relied on the same strategy to protect himself from emotional injury at the hands of romantic partners and friends. Unsurprisingly, those relationships often ended poorly, as he would become furious with intimates over what he perceived as cruel indifference toward him. Over time, his therapist noticed that during moments of vulnerability the patient would speak with eyes downcast, as if determined to avoid looking at the therapist. One day, while recounting a troubling event, the patient became angry at the therapist and said that he could tell by her cold and blank facial expression that she did not care about him. Gently, the therapist pointed out that he had not looked at her for quite some time and therefore could have no idea of what facial expression she wore. The therapist encouraged the patient to look at her now. When he did, he saw the tenderness of her concern and reacted with astonishment, as if his mind had played a trick on him, which it had.

This example illustrates the profound relevance to psychotherapy of the emerging neuroscience theory of predictive processing (Clark, 2013). Predictive processing is the idea, supported by empirical research, that our brains maintain predictive models of the world, and, more importantly, what we perceive and experience as concrete reality is, in fact, a prediction generated by these models (Clark, 2013, 2023; Hohwy, 2013; Rao &

Ballard, 1999; Seth, 2019). In other words, our subjective reality is less a function of incoming sensory data—sights, sounds, smells, and so on—than of our brain's best anticipatory guess at what that sensory data is likely to be.

In this article, we begin with a brief overview of predictive processing model, emphasizing its relevance to psychopathology and psychotherapy. The predictive processing framework explains how maladaptive beliefs persist in psychopathology by providing a sense of predictability and controllability, even when ultimately harmful. Yet, it does not account for why individuals eventually seek therapy to challenge beliefs that once offered a sense of security. We then introduce control–mastery theory (CMT), a cognitive–psychodynamic–relational framework that explains how patients actively work to revise these beliefs in psychotherapy, driven by an intrinsic motivation for change. Within CMT, we introduce the concept of patient testing, an implicit strategy through which patients seek corrective emotional and interpersonal experiences. We argue that the neuroscience-based predictive processing model dovetails with psychoanalytically informed CMT, particularly in how patient testing serves as a mechanism for generating salient prediction errors, allowing for the revision of internal models and facilitating adaptive change. This integration offers novel clinical insights into the interplay between neuroscience and psychotherapy.

To be explicit, this article presents a theoretical integration (Norcross & Goldfried, 2005), a coherent explanatory model that links predictive processing and CMT to clarify mechanisms of psychotherapeutic change. Predictive processing provides a neuroscientific account of belief updating through prediction errors and model revision, while CMT offers a psychotherapeutic account of

patients' motivation to revise entrenched beliefs through their intrinsic drive to test and disconfirm pathogenic beliefs. Brought together, these complementary frameworks form a more comprehensive model of psychotherapy, illustrated with clinical material showing how attuned responses generate prediction errors that patients actively use to revise maladaptive priors. Our aim is not to merge or replace existing clinical methods or to prescribe a new manualized treatment. Rather, we propose a conceptual integration that generates testable hypotheses to guide future empirical and translational work.

A Brief Overview of the Predictive Processing Model

The origins of the predictive processing model can be traced back to the 19th-century work of Hermann von Helmholtz who posited that visual perception relies upon "unconscious inferences" drawn from sensory data and prior knowledge. According to von Helmholtz (1867), what we perceive is driven less by direct sensory input from the external world and more by the brain's best guess of what those inputs are likely to represent, informed by past experiences and learned knowledge. The contemporary iteration of this model, rooted in cognitive neuroscience, is known by various names, including predictive processing (Clark, 2013, 2023), predictive coding (Rao & Ballard, 1999), active inference (Friston, 2003, 2005, 2008), Bayesian brain (Knill & Pouget, 2004), and free-energy principle (Friston, 2010).

These theories, collectively referred to here as predictive processing, propose that the human brain constructs an internal, multilevel sensory representation of the external world that is accumulated through lived, embodied experience. This internal representation is not a static template or a passive lens through which the world is viewed, but rather a dynamic generative model predicting the sensory inputs most likely to be encountered. These predictions continuously shape both perception and behavioral responses to incoming sensory data by comparing what actually arrives with what the generative model has anticipated. Discrepancies between the brain's predictions and the sensory data from physical reality produce error signals that drive the internal model to self-correct and adapt (Clark, 2013). Sensory experience, according to this view, serves primarily to update the brain's internal

representations, while the generative model itself is responsible for creating perception. What we "see" can be understood as a kind of "controlled hallucination" (Clark, 2013; Seth, 2019), where predictions produce perception yet are constrained and constantly revised by incoming sensory input. As such, we do not see the world as it truly is; instead, we see it as we are, filtered through the predictions and biases of our internal model (Seth, 2019).

One of the intriguing insights from the predictive processing theory is that the brain, when confronted with errors—that is, mismatches between its predictive model and sensory input—must constantly decide which information to prioritize. Feldman and Friston (2010) described this balancing act as an assessment of "precision," in which the brain continuously evaluates whether the error-producing sensory data are sufficiently precise and reliable to warrant updating the predictive model. In a situation where sensory data are deemed highly reliable and precise, a person may find themselves deeply attending to their sensory experience, allowing for rapid and accurate updates to their internal model. For example, an individual with social anxiety or rejection sensitivity might initially interpret a coworker's neutral remark as dismissive. However, if that same coworker consistently offers genuine compliments, maintains warm eye contact, and expresses sincere appreciation, saying things like,

> "*Your ideas really make a difference*"—these unmistakable signals become difficult to ignore, ultimately prompting the individual to re-evaluate their negative expectations about social interactions. By focusing on these reliable sensory cues, the individual's internal model is gradually updated, reducing anxiety and fostering a more accurate perception of social interactions. Conversely, if the error-producing sensory data is perceived as ambiguous and less reliable—prompting thoughts like, "*Who do you believe? Me or your lying eyes?*"

the preexisting internal model remains favored, and change is resisted. In the example above, this might involve the same coworker offering a fleeting smile without sustained eye contact or further verbal affirmation, lacking the clarity and consistency needed to challenge the belief[1] in inevitable rejection.

---

[1] There is some debate over whether the term "belief" is accurate within the predictive processing framework (Yon et al., 2020). We use the term here to refer broadly to predictions, cognitive schemas, and organizing frameworks for information in the brain.

The Relevance of the Predictive Processing Model to Psychopathology

The predictive processing model offers a compelling framework for understanding both the origins of psychopathology and the mechanisms through which psychotherapy can effect change. The predictive processing model explains psychopathology as emerging from a disruption in the brain's continuous cycle of prediction and error correction (Van de Cruys & Van Dessel, 2021). As mentioned, under normal conditions, the brain generates predictions about incoming sensory data based on internal models, compares these predictions with actual sensory input, and then updates the models to reduce prediction errors. However, when a person lives in a prolonged harmful environment and experiences chronic stress, trauma, or maladaptive learning, the brain may begin to generate consistently inaccurate or biased predictions. These maladaptive predictions lead to an internal model that is overly negative or rigid, creating a framework through which sensory input is interpreted in a way that reinforces these biases. Over time, this maladaptive internal model becomes self-perpetuating: misinterpreted sensory data fail to be perceived as errors to trigger the necessary updates, and the brain continues to expect negative outcomes. This vicious cycle—where biased predictions shape perception and behavior, which in turn reinforce the original maladaptive predictions—contributes to the development and maintenance of disorders such as anxiety, depression, and other forms of psychopathology.

Alternatively, when sensory data are too painful or overwhelming, the brain may adapt by reducing or ceasing its evaluation of new input. Instead of updating its internal model with distressing information, the brain may downregulate or "tune out" the signals, a protective mechanism that minimizes immediate emotional distress but may also reinforce an outdated or maladaptive internal model. In predictive processing terms, this can be understood as a shift in precision weighting (Friston, 2005, 2010), where new sensory input is assigned a lower reliability, effectively preserving the current model despite clear errors, much like what is seen in dissociative or avoidance responses in chronic trauma. Together, these pathways suggest that psychopathology arises from failures in the brain's adaptive learning process, either through persistent misinterpretations of sensory input that reinforce rigid and negative internal models or downregulation of sensory input to avoid distress, preventing necessary updates to one's internal model.

Fabry (2020) suggested that depression may result from an overemphasis on internal predictions, where the brain assigns too much weight to its own expectations at the expense of actual sensory feedback, thereby reinforcing negative moods and cognitive patterns. In this view, the brain's failure to update its internal model leads to a cycle where negative moods, impaired thinking, and a reduced sense of agency persist over time. Similarly, Herzog et al. (2022) suggested that borderline personality disorder develops when rigid, oversimplified predictive processing models of self and others combine with a poor ability to adjust beliefs, resulting in emotional instability and unhealthy interpersonal behaviors. Moreover, Kube et al. (2020) argued that a dysfunctional predictive processing model can shield individuals from even perceiving sensory input. In other words, entrenched, top-down predictions actively block or diminish the influence of new, potentially disconfirming evidence. Essentially, psychopathology can be seen as the result of a failure in the brain's adaptive learning process, where persistent errors in prediction lead to a distorted internal representation of reality (Van de Cruys & Van Dessel, 2021).

These entrenched, once-adaptive predictions are difficult to change because they shape the individual's very sense of reality. The brain may ignore or downplay new information that challenges these beliefs, often requiring a much greater amount of positive evidence to alter them (Baumeister et al., 2001). Overall, this rigidity in updating internal models represents a central challenge in understanding psychopathology within the predictive processing framework.

The Relevance of the Predictive Processing Model to Psychotherapy

A key goal of psychotherapy, within this framework, is to disrupt the maladaptive cycle of entrenched predictions by creating a safe and reciprocal environment where prediction errors can be both detected and integrated, facilitating the model updating. While different therapeutic models employ distinct techniques, they share the common aim of challenging and recalibrating inaccurate predictions arising from dysfunctional

generative models to achieve this disruption (Villiger, 2025). Krupnik (2019) similarly proposed that psychotherapy can be understood as a process of systematically recalibrating entrenched predictions to facilitate adaptive belief updating. By consistently providing reliable, unambiguous sensory input that contradicts the patient's maladaptive predictions, therapists can actively destabilize these entrenched expectations, facilitating the recalibration of generative models to align with present reality. Over time, such therapeutic interventions promote a more accurate perception of reality and, consequently, a reduction in symptomatic distress.

This process of disrupting maladaptive predictions and recalibrating generative models is particularly relevant to trauma-focused therapies, where deeply entrenched expectations often prevent the integration of new, corrective experiences. Kindt et al. (2007) investigated mechanisms of change in cognitive behavioral therapy for posttraumatic stress disorder using a treatment that combined imaginal exposure and imagery rescripting. Their analysis of trauma processing focused on shifts in memory representations, demonstrating that effective treatment first required engaging with perceptual and sensory aspects of traumatic memories before transitioning to a more conceptual understanding. This progression—starting with reprocessing stored sensory and emotional experiences before integrating them into a broader cognitive framework—was associated with better treatment outcomes. Findings from Kindt et al. (2007) indicated that simply thinking about trauma on a conceptual level is insufficient for lasting change; instead, individuals must first reconnect with and reprocess the raw sensory and emotional components of their traumatic experiences before they can meaningfully update their beliefs and interpretations.

This conforms to Chamberlin (2019), who theorized that eye movement desensitization and reprocessing (EMDR) facilitates trauma resolution by first allowing individuals to reconnect with the felt experience of trauma, thereby helping them update maladaptive priors (i.e., preexisting beliefs or expectations) that otherwise constrain the integration of new, predictive models. Such maladaptive priors manifest as a resistance to modifying entrenched negative beliefs because the brain prioritizes previously learned, maladaptive predictions about safety, trust, and control. EMDR works by presenting the individual with new sensory input that contradicts these rigid predictions, thereby enabling prediction errors to recalibrate entrenched memory representations and foster adaptive belief updating. In other words, by revisiting the original sensory and emotional imprint of trauma, the brain becomes more receptive to incorporating new, corrective experiences, making it possible to reframe the meaning of past events in a way that reduces distress. Similarly, Putica et al. (2022) emphasized the role of predictive coding in posttraumatic stress disorder treatment, highlighting value-based learning—the process through which the brain updates its predictions and behaviors based on the perceived value of an outcome—as a mechanism for recalibrating maladaptive generative models and enhancing emotional regulation. Thus, trauma-focused interventions like EMDR and exposure therapy facilitate change by presenting repeated opportunities for the brain to detect and integrate prediction errors—mismatches between expectations and present sensory experiences—leading to a gradual reshaping of how trauma is encoded and interpreted.

However, a safe and reciprocal therapeutic environment may be a prerequisite for prediction errors to be detected and integrated. Connolly (2022) emphasized that before individuals can engage with new information that challenges entrenched priors, they must first establish a secure relationship with their therapist. This sense of safety modulates the precision of prediction errors, allowing previously filtered-out information to be processed. This aligns with Van de Cruys and Van Dessel (2021), who emphasized the importance of structured learning environments in updating generative models. A secure therapeutic relationship enables patients to engage with surprising information rather than filtering out contradictions to entrenched beliefs. Without first establishing the therapist as a trusted information source, prediction errors may be dismissed or misinterpreted, impeding generative model revision. These studies together highlighted that therapeutic effectiveness depends not only on the interventions themselves but on whether the patient perceives the therapist as a salient and reliable source of prediction error minimization.

How CMT Complements the Predictive Processing Framework

While Villiger (2025), Krupnik (2019), and others have either offered an integrative model for

conceptualizing psychotherapeutic interventions within the predictive processing framework or applied this framework to understanding specific therapeutic models and techniques, they often place the therapist—or the therapeutic models or techniques—at the center of the change process. Notably, absent is a consideration of the patient's agentic efforts to revise their predictive models in psychotherapy.

Ahn and Wampold (2001) estimated in their meta-analysis that about 13% of total outcome variance is explained by treatment (alliance, therapist, model or technique, placebo, and allegiance), leaving 87% of the variance attributed to patient or extratherapeutic factors as well as unexplained and error variance. Constantino et al. (2018) concluded in their meta-analysis that patients' early treatment outcome expectations were significantly associated with posttreatment outcomes, emphasizing the importance of patients' beliefs and contributions from the outset of therapy. Bohart and Tallman (2022) argued that clients are active participants who utilize their own strengths and expertise to drive healing, a view echoed by Carey (2010), who found that a patient-led approach in psychological interventions improved engagement and treatment outcomes. Together, these studies suggested that therapeutic change is not simply a product of therapist interventions but requires the patient's agentic efforts and engagement in revising deeply held priors through error detection and integration.

This section aims to shift the focus from the therapist or therapy to the patient's active role in updating preexisting beliefs or expectations, as viewed through the lens of CMT, a framework that emphasizes patient agency. This perspective enriches our understanding of how prediction errors are minimized in psychotherapy and broadens the discussion of the underlying mechanisms of therapeutic effectiveness by highlighting the patient's role in this process.

CMT is an integrative cognitive–psychodynamic–relational model that explains how maladaptive beliefs develop and how psychotherapy works by helping patients overcome these beliefs and reconfigure their internal dynamics (Silberschatz, 2005; Weiss, 1993). According to the CMT model, humans are continuously updating their internal models based on interactions with their immediate emotional environment, particularly within significant relationships like those with parents and family members. When a child is

subjected to adverse or traumatic experiences, such as experiencing a parent who is abusive or neglectful, the child must somehow adapt. One common way of doing so is for the child to develop an internal model or set of beliefs under which the child perceives themselves as bad or undeserving. While such beliefs can be adaptive in early childhood in that they can help the child cope with a dysfunctional environment, the same beliefs become maladaptive and pathogenic later in life when they interfere with the individual's ability to pursue goals or form meaningful relationships.

In the CMT framework, control refers to the unconscious strategies individuals use to manage their internal conflicts and relational dynamics, essentially defenses formed to cope with unresolved, distressing past experiences. Instead of reflecting a desire for dominance, control represents the adaptive strategies people use to create predictability and a sense of safety in their lives. In contrast, mastery describes the process by which individuals actively work through these conflicts, not merely coping with or avoiding them, but by overcoming and integrating challenging experiences to foster personal growth and pursue their life goals. From a predictive processing perspective, control can be seen as an attempt to minimize uncertainty by relying on established priors, even if those priors are no longer adaptive in the present reality. Mastery, on the other hand, involves revising and updating these priors in response to new information and corrective experiences in therapy, thereby reducing prediction errors and fostering a more adaptive sense of self. At its core, CMT is a theory of adaptation that aligns well with predictive processing frameworks.

While both CMT and the predictive processing framework agree that patients come to psychotherapy to update their internal models, we argue that they emphasize different aspects of this process. The predictive processing framework focuses on the brain's drive to minimize free energy, that is, to reduce uncertainty by lowering prediction errors through adjusting internal priors (Feldman & Friston, 2010; Friston, 2010). From this perspective, maladaptive beliefs persist because they provide a temporary reduction in uncertainty: if these beliefs fit the individual's current (even if distorted) model of the world, the brain may continue to rely on them to avoid the immediate costs of revising ingrained priors. These costs may include overwhelming pain and

anxiety, cognitive effort, or the destabilization of a familiar sense of self and relationships. Thus, while such beliefs may ultimately be harmful, the short-term "benefit" of stability and predictability can outweigh the motivation to change.

Paradoxically, if these beliefs buffer against uncertainty, making change feel risky or overwhelming, why would individuals seek psychotherapy for change? The predictive processing framework explains how maladaptive beliefs persist but does not fully account for why people choose to challenge them. CMT addresses this gap by emphasizing the patient's intrinsic motivation and active role in revising pathogenic predictions, adding a crucial motivational dimension that predictive processing alone does not explain.

According to CMT, individuals enter therapy with an intrinsic drive to disconfirm pathogenic beliefs or schemas, beliefs that were once protective in response to traumatic experiences and their psychological processing but now perpetuate distress and hinder personal growth (Weiss, 1993, 1997). This theory posits that the desire to achieve mastery over past traumas and internal conflicts serves as a powerful force, prompting patients to test and update their internal models during therapy. This motivation is intrinsic and operates unconsciously, humans have an inherent drive to heal and grow, which shapes their behavior in therapy. Patients engage in unconscious testing of their beliefs with the therapist, seeking corrective experiences that facilitate belief revision and conflict resolution.

Thus, while the predictive processing framework explains the mechanisms of belief updating—how prediction errors lead to internal model revision—CMT adds a motivational perspective by explaining why individuals seek change despite the temporary stability of maladaptive beliefs. Even when these beliefs provide a sense of safety by reducing uncertainty, the drive to alleviate suffering, pursue personal goals, and form meaningful relationships eventually compels individuals to engage in therapy. Together, these two frameworks provide a more comprehensive account of psychotherapeutic change, with CMT highlighting the patient's active role in guiding the therapeutic process.

For clarity, in this article we use two tradition-specific terms that overlap but emphasize different aspects. In predictive processing, maladaptive priors refer to hierarchically encoded probabilistic expectations that bias perception and action. In CMT, pathogenic beliefs denote coherent, often unconscious expectations about the self and others that organize behavior and symptom formation. Both terms describe enduring internal templates that were once adaptive but now mislead perception and action. We retain both because predictive processing highlights the computational form (priors, precision, prediction error), whereas CMT highlights the motivational organization and the patient's unconsciously guided testing of those templates.

Patient Testing in Psychotherapy

A central tenet of CMT is the active role that patients play in psychotherapy, working to address their deep-seated problems and lead more fulfilling lives; that is, they come to therapy with the goal of challenging and disconfirming their pathogenic beliefs, and they work to test these beliefs within therapy sessions. In the CMT framework, patient testing refers to the process where patients, either unconsciously or consciously, elicit reactions from their therapists to determine whether they align with or disconfirm the patients' pathogenic beliefs (Weiss, 1993, 1997). This process serves as a mechanism for assessing whether it is safe to revise deeply ingrained, maladaptive expectations about themselves, others, and relationships. Patient testing occurs in two primary forms: transference testing, where patients reenact past relational patterns to see if the therapist reinforces or disconfirms their beliefs, and passive-into-active testing, where patients treat the therapist as they were once treated to observe if they receive a different response (Silberschatz, 2005; Weiss, 1993). For example, a patient who unconsciously believes "People will abandon me if I express my needs" may test this belief by becoming emotionally vulnerable with their therapist. If the therapist remains attuned and supportive rather than withdrawing, the patient may begin revising their belief, realizing that expressing needs does not necessarily lead to abandonment. When therapists respond in ways that disconfirm pathogenic beliefs—by providing attuned, corrective experiences—patients can begin updating their internal models, increasing their openness to change and adaptive growth (Silberschatz, 2008; Silberschatz & Curtis, 1993; Silberschatz et al., 1989).

As an illustration, a young woman sought psychotherapy because she felt excessively self-contained and wanted to develop a greater ability to rely on others for support. Growing up as an only child in a dysfunctional family, she recalled early experiences of seeking help from her parents, only to be met with frustration and rejection. When she asked for guidance as a child, her parents angrily responded, "We don't know what to do—why don't you figure it out yourself?" Over time, she internalized the belief that she could not rely on others for help, reinforcing a pathogenic belief that self-sufficiency was necessary for survival. This belief became a core part of her generative model, shaping how she approached relationships and leading to difficulties in seeking and accepting support.

In therapy, she unknowingly engaged in patient testing to determine whether the therapist would reinforce or disconfirm this belief. One day, she began a session by saying, "I'm really not sure what to talk about today and would like to get some guidance from you. Do you have any suggestions?" This was a subtle but crucial test: was it safe to rely on the therapist for guidance, or would they respond as her parents once did? A therapist adhering to a nondirective stance might reflexively reply, "What would you like to talk about?", an intervention that, while neutral, fails to account for the relational context of the patient's history and unintentionally reinforces her belief that she must rely solely on herself. A more attuned response, however, would recognize this as a test of trust and provide the guidance she explicitly requested, offering a corrective emotional experience that challenges her belief that seeking help will only lead to rejection. This example illustrates how patient testing serves as a mechanism for belief revision, allowing patients to elicit responses from the therapist that either confirm or disconfirm their pathogenic assumptions. When the therapist successfully "passes" these tests by responding in a way that challenges maladaptive priors, the patient begins to assign greater precision to new relational experiences, fostering the gradual revision of their generative model.

In predictive processing terms, patient testing can be understood as a process in which patients generate predictions about the therapist's responses based on their maladaptive priors. When the therapist's actual responses differ from these expectations, the resulting prediction errors provide critical feedback, enabling the patient to update their internal generative model. Patient testing thus functions as a central mechanism for eliciting informational discrepancies that heighten the likelihood of salient prediction errors. These errors signal a mismatch between the patient's entrenched expectations and reality, creating opportunities for belief revision. Building on the information revealed through patient testing, the therapist can offer interpretations that directly challenge these pathogenic predictions. By reframing the patient's experiences or introducing alternative perspectives, the therapist deliberately amplifies prediction errors that disrupt rigid generative models, increasing the likelihood of their revision and integration into a more adaptive framework.

The therapist's responses—such as emotional attunement, consistency, and supportive nonverbal cues—play a critical role in this process. When these responses violate the patient's maladaptive expectations—for example, demonstrating reliability and care in moments when rejection is anticipated—they generate salient prediction errors that facilitate belief updating. These corrective experiences, often elicited through patient testing, gradually shift precision weighting: over time, confidence in maladaptive priors weakens, while the precision assigned to new, more adaptive sensory inputs increases. This clinical insight echoes the long-standing corrective-experience literature: first described by Alexander and French (1946) and Loewald (1960) as corrective emotional experiences in psychoanalytic therapy, and later argued by Goldfried (1980, 2012) that such experiences as a core, transtheoretical process.

Additionally, the patient's emotional and interoceptive responses contribute to the detection of prediction errors. Feelings of unexpected safety, trust, or relief act as internal signals that contradict deeply ingrained negative expectations. As the therapist continues to pass the patient's tests by consistently disconfirming their pathogenic beliefs, the patient becomes more open to revising their internal models. Each instance in which the therapist responds in a nonpunitive, attuned, and supportive manner reinforces the patient's ability to trust new information. Over time, this process of implicit belief testing and learning supports the gradual revision of generative models, enabling more flexible and adaptive ways of engaging with the self and others. As a result, therapy provides not just exposure to new relational experiences but a safe

environment in which the patient can actively test, engage with, integrate, and stabilize these disconfirmatory signals, ultimately revising maladaptive beliefs.

A Clinical Example

This section draws on a clinical case of a woman in her 30s to illustrate how early trauma and neglect shape maladaptive generative models, constraining her predictions about self, others, and relationships. Through patient testing within a safe therapeutic relationship, repeated interactions generate salient prediction errors that disconfirm her maladaptive priors, facilitating their revision. As she tolerates and integrates disconfirmatory experiences, her expectations shift, enabling adaptive belief updating and therapeutic change.

A woman in her early 30s sought therapy to help with an acute depressive episode triggered by her discovery that her husband was having an affair. She had initially come across romantic email exchanges between her husband and another woman, but when she confronted him, he dismissed her concerns as "making much ado about nothing." This response led her to question her own judgment—"maybe I'm just being paranoid"—until she uncovered explicit sexual photos he had also shared with the other woman. The deep sense of betrayal she felt was compounded by his gaslighting, intensifying her distress and fueling her depressive symptoms.

Themes of betrayal and gaslighting permeated the patient's history. Her mother abandoned her and her father to join a commune, and her father later remarried a woman with a son several years older than the patient. When the patient entered adolescence, her stepbrother began entering her bedroom to masturbate while she was asleep. She reported this to her father and stepmother, but they both dismissed her concerns. The sexual abuse later escalated: the stepbrother routinely fondled her and demanded she masturbate him. When she confronted her parents again, her stepmother excused his behavior as typical of teenage boys, telling the patient to simply accept it.

Feeling vulnerable, confused, and unprotected by her parents, the patient turned to the pastor of her church's teen group. Though she did not know him well, he seemed kind and trustworthy. Their first meeting was brief, but he encouraged her to return to his office to discuss her situation further. Over several sessions, he asked detailed questions about the abuse, focusing on what had happened and how she felt. Despite his insistence on her returning to share more, he offered no advice or support. Over time, the patient began to feel these meetings were inappropriate and driven by his own interests. Leaving the church, she felt more alone and betrayed than ever.

According to CMT, the patient developed pathogenic beliefs that her feelings and needs did not deserve to be taken seriously, leading her to see herself as unworthy of protection and unable to trust her own perceptions. Her mother's abrupt departure—coupled with her father's preoccupation with a new partner—left her feeling confused, betrayed, and self-blaming. She struggled to make sense of why she had been abandoned and whether her distress was justified, reinforcing an early pattern of questioning her own emotions. These relational traumas likely shaped her unconscious expectation that her own perceptions were unreliable, leading her to habitually question whether her emotional responses were valid or exaggerated. When she sought help from her parents regarding the sexual abuse, their dismissal of her concerns reinforced her pathogenic belief that her experiences of the abuse were not valid, and her feelings were unworthy of acknowledgment. Her maladaptive generative model was once adaptive in childhood. As a child, attributing blame to herself provided a more stable and predictable model than accepting the chaotic and neglectful behavior of her caregivers. If she saw her caregivers as the problem, she would have had to confront the terrifying reality that no one was there to protect her. Instead, self-blame provided a "stable" model, allowing her to assume control over her suffering rather than experience complete helplessness.

Meanwhile, the pastor's inappropriate interest in hearing details of the abuse further destabilized her sense of reality, leaving her confused about whether her distress was justified. Feeling exposed and exploited rather than supported, she became increasingly uncertain about whether her emotions were an accurate reflection of reality, reinforcing the pathogenic belief that her experiences were unreliable and unworthy of acknowledgment. Her maladaptive generative model, once reinforced, became increasingly resistant to revision. Each new relational experience that aligned with her pathogenic beliefs

strengthened the high-precision weighting on those priors, making it harder to accommodate contradictory information.

This pattern repeated in adulthood when she discovered her husband's affair. His gaslighting response—dismissing her concerns as paranoia and exaggeration—mirrored her earlier experiences of having reality denied. Each time she questioned him, she wavered between anger and self-doubt, wondering if she was reading too much into the situation. Rather than integrating the conflicting evidence that her husband was deceiving her, her brain assimilated his response into her existing model, reinforcing her prior belief that she could not trust her own perceptions. The familiarity of being told she was "making much ado about nothing" deepened her confusion and reinforced her long-standing belief that her perceptions could not be trusted. As a result, she dismissed her own instincts and prioritized the stability of her maladaptive priors—"Maybe I am just paranoid" or "Maybe my perceptions really can't be trusted"—clinging to the belief that she was simply overreacting rather than risk confronting the painful truth.

How did she work in therapy to test her pathogenic beliefs? During the early months, the patient worked primarily by testing to see whether the therapist would take her perceptions and concerns seriously or dismissing them as overreactions, mirroring her past experiences with invalidation and relational neglect. For example, she frequently recounted instances where her intuitive suspicions about her husband—such as questioning his late arrivals when he claimed to have simply stopped for a drink—were dismissed by him as mere signs of paranoia. Her prior relational experiences of being manipulated and gaslit had led to a cognitive habit of downregulating contradictory evidence rather than using it to revise her beliefs. Under the high-precision influence of her maladaptive priors, she had learned to downregulate her confidence in her perceptions rather than assign greater weight to external contradictions. However, in sharp contrast to her husband's responses, the therapist validated her intuition, affirming that her concerns were not unfounded and that her tendency to question herself reflected long-standing relational conditioning rather than actual distortion of reality.

Similarly, she would test the therapist by seeking the therapist's response to self-doubt and conflicting interpretations of interpersonal events. She would say, "Maybe I was overreacting?" "I often feel that my sense of being ignored might just be my own interpretation," and "I sometimes wonder if I read too much into things." Having grown up in an environment where her feelings and perceptions were denied, dismissed, or manipulated, this testing was an implicit attempt to assess whether the therapist would mirror those past experiences or provide a disconfirmatory response.

Instead of reaffirming the familiar pattern of invalidation (e.g., outright dismissing her concerns or encouraging her to ignore her feelings)—as she had been conditioned to expect—the therapist provided consistent validation of her feelings and concerns, reinforcing that her emotional responses were legitimate rather than exaggerated. At the same time, the therapist gently helped her explore alternative evidence from recent interactions that contradicted her negative predictions as well as assess when her interpretations of sensory inputs were accurate versus when they were shaped by high-precision maladaptive priors. By repeatedly encountering salient prediction errors in a safe therapeutic environment—where her expectations of invalidation were disconfirmed—she gradually assigned greater precision to new relational experiences.

Over time, this weakened the precision weighting of her prior belief that her perceptions were unreliable, making space for new, more flexible interpretations of social interactions. As these repeated disconfirmatory experiences accumulated, she gradually developed a new generative model in which her perceptions were not inherently flawed, her emotional responses were meaningful, and she was capable of trusting her own interpretations rather than defaulting to self-doubt. As her generative models updated, she developed greater confidence in her emotional experiences, reducing the tendency to automatically doubt or distrust her own perceptions.

CMT as a Motivational Complement to Predictive Processing

As mentioned earlier, while predictive processing explains how hierarchical generative models generate prediction errors that drive belief updates through precision modulation, it does not fully account for why individuals with psychopathology seek psychotherapy for intentional change,

especially when their maladaptive models provide a sense of stability and certainty. In predictive processing terms, belief revision is not automatic; individuals may resist change, even when confronted with contradictory evidence, because their high-precision priors offer a stable, predictable framework that minimizes uncertainty. Paradoxically, even distressing beliefs may persist because they serve an immediate psychological function, minimizing distress in the short term, even if they create suffering in the long run. A maladaptive but predictable model may feel safer than confronting an unknown alternative—which the brain registers as a threat—even if that alternative could be more beneficial in the long term. This process explains why people can remain stuck in dysfunctional relationships or harmful self-perceptions, despite experiencing emotional pain. A person in an emotionally abusive relationship, for example, may struggle to leave, not because they take pleasure in suffering, but because the uncertainty of leaving feels more threatening than the familiarity of staying.

While some individuals remain stuck in rigid, self-perpetuating beliefs, others work toward intentional change, despite the uncertainty involved. Why is this the case? CMT provides a crucial motivational dimension by positing that: first, people are intrinsically motivated to revise pathogenic beliefs to pursue goals, form meaningful relationships, and lead more fulfilling lives; second, through unconscious testing in a safe relationship, people can explore whether their priors hold true or can be disconfirmed. Psychotherapy provides a safe relational environment, offering people opportunities to structurally experience and integrate prediction errors. This process facilitates belief updating while managing psychological distress, allowing for the gradual revision of maladaptive generative models.

Moreover, as discussed earlier, while Villiger's (2025) and Krupnik's (2019) models emphasize how therapists can structure an environment to elicit discrepancies between top-down predictions and bottom-up sensory input (i.e., prediction errors signaling the need for model updating), they do not fully capture the patient's active role in testing and revising internal models. CMT fills this gap by positing that patients are not passive recipients of prediction errors but active agents with an intrinsic, goal-directed drive to test and disconfirm pathogenic beliefs. Extensive CMT research (e.g., Bugas & Silberschatz,

2000; Bugas et al., 2023; Gazzillo et al., 2025; Kealy et al., 2020, 2022) supports the notion that patients mobilize internal resources to challenge and revise entrenched, often unconscious, maladaptive schemata.

Passing the Test: How Therapists Create Conditions for Patient-Driven Model Updating

In this framework, the therapeutic process is viewed as a series of tests, wherein the patient, driven by a desire for mastery, actively engages with corrective experiences to validate new, more adaptive internal models. The therapist's role is to identify and respond to these testing processes, providing interventions that generate the necessary prediction errors while simultaneously empowering the patient to integrate these experiences as opportunities for change. Hence, CMT suggests that the drive for belief updating arises not only from the therapist's interventions but also from the patient's own repeated efforts to confirm or disprove pathogenic assumptions.

By recognizing patients as active agents in the therapeutic process, CMT clarifies that the therapist's role extends beyond simply generating prediction errors. The therapist's task is threefold. First, the therapist must create a safe and consistent relational environment that allows the patient to tolerate and integrate disconfirmatory experiences without being overwhelmed. Since maladaptive priors often persist because they provide a sense of stability and predictability, the therapist's attuned responses must help the patient feel secure enough to engage with new relational information rather than defensively rejecting it. A sense of relational safety reduces the perceived cost of prediction errors, enabling the patient to remain open to belief revision. This aligns with the concept of epistemic match in psychotherapy, where individuals can only update their generative models when they perceive the therapist as an attuned and trustworthy source of new information (Li et al., 2025). Early epistemic match creates a "window of opportunity" in which patients begin to assign higher precision to therapist-provided input, allowing for the gradual revision of entrenched priors. At this stage, patients engage in minimally active inference, passively integrating therapist-provided interpretations within existing generative models while remaining

within familiar interpretative frameworks. The therapist's marked mirroring—an attuned reflection of the patient's emotional state—modulates the salience of prediction errors, making them noticeable without overwhelming prior beliefs. This process ensures that early learning experiences in therapy remain within an optimal range, where new information is sufficiently surprising to generate model updating but not so imprecise as to trigger excessive uncertainty or resistance. Only when this foundation is established can patients begin assigning greater precision to novel inputs and actively revising their generative models of self and others.

Then, the therapist must create meaningful disconfirmatory experiences by subtly guiding the patient toward recognizing inconsistencies in their maladaptive models. Rather than simply contradicting the patient's pathogenic beliefs, which may be dismissed outright, the therapist helps the patient notice and process discrepancies in a way that makes belief revision more accessible. This involves modulating the precision of prediction errors, encouraging the patient to assign greater weight to new relational inputs while reducing the excessive precision of maladaptive priors. Through carefully paced interventions, the therapist helps the patient gradually integrate corrective experiences rather than overwhelming them with abrupt contradictions.

Finally, the therapist must recognize and respond optimally to patient testing. As patients unconsciously seek to verify whether their pathogenic beliefs hold true, they engage in subtle tests that elicit specific reactions from the therapist. These moments serve as opportunities for the therapist to either reinforce or disconfirm the patient's maladaptive priors. If a therapist fails the patient's test—for example, by acting dismissively, inconsistently, or unempathetically, or in a way that mirrors past invalidating experiences—the patient may interpret this failure as confirmation of their maladaptive prior, reinforcing the belief and making it even more resistant to change. This may to some degree explain why some patients appear to regress or disengage from therapy when the therapist inadvertently validates their pathogenic expectations. However, when the therapist consistently "passes" these tests—by responding in a way that disconfirms the pathogenic beliefs and providing validation, reliability, and corrective emotional experiences—the patient experiences a mismatch between their expectations and reality and gradually registers these disconfirmatory responses as evidence that their prior assumptions may not be accurate. Over time, this process weakens the rigidity of maladaptive models, allowing the patient to revise their expectations of self and others. By creating a secure therapeutic environment, creating controlled exposure to prediction errors, and optimally responding to patient testing, the therapist provides the conditions necessary for adaptive belief updating and psychological change.

Limits of the Predictive Processing–CMT Integration

Although CMT explains why many patients enter therapy with an implicit plan to test and revise pathogenic beliefs, this motivational account depends on enabling conditions that are not always present. Some priors may hold such excessive precision that they down-weight corrective evidence so strongly that patients cannot easily engage in effective testing. Moreover, the neurocomputational machinery (e.g., precision-weighting) must permit updating. Others lack the felt agency, emotion-regulation capacity, or epistemic openness needed to attempt or sustain tests. For example, patients who show a high level of general psychopathology (Caspi et al., 2014; a diffuse, transdiagnostic burden of symptoms) often experience pervasive anergia, cognitive fog, and overwhelming affect that together undermine the basic capacities required for testing. Such patients may be so emotionally dysregulated that the very idea of intentionally evoking a vulnerable moment feels intolerable: they cannot sit with the surprise or uncertainty that a test produces and so either avert the test altogether or respond in ways that collapse the opportunity (withdrawal, shutdown, or explosive reactivity).

At the same time, structural and cultural barriers (stigma, poverty, limited access) or a history of repeated interpersonal betrayals can produce deep epistemic mistrust—an implicit stance that others are not safe or reliable sources of information—which prevents patients from treating a new therapist's responses as informative even when the therapist is consistent and attuned. Therapist factors—poor attunement, inconsistency, or model mismatch—can also cause repeated test failures that

reinforce pathogenic beliefs. In these cases, the CMT pathway—where patient-led testing leads to corrective experience and then to model revision—is stalled not because the patient lacks motivation in any moral sense, but because the neurobiological and affective resources required to carry out and integrate tests are compromised. Thus, while CMT captures an important motivational mechanism, it is not universally sufficient: change requires both the patient's drive to test and a set of neurobiological, relational, and contextual conditions that enable testing to generate salient prediction errors and subsequent model revision.

While predictive processing and CMT are clearly complementary, there are important boundary conditions and potential tensions that merit careful consideration. First, predictive processing's mechanistic emphasis on prediction error and precision foregrounds computational descriptions of how the brain updates models, but this focus can understate the motivational and volitional factors that CMT places at the center of therapeutic change. For example, predictive processing models describe how a salient prediction error could, in principle, drive updating; they say less about why a patient would risk provoking that error or sustain the difficult behavioral experiments that testing requires. Conversely, CMT emphasizes unconscious plans, goals for mastery, and the patient's agency in eliciting corrective experiences, concepts that are theoretically rich but can be challenging to formalize in computational terms. This asymmetry raises both conceptual and operational questions: how do we map motivational states (felt agency, desire for mastery) onto precision parameters or decision policies within active-inference formalisms?

Second, there is a difference in timescale and tempo. Predictive processing often models relatively rapid perceptual inference and trial-by-trial learning dynamics, whereas CMT foregrounds longer-term, motivation-driven processes of testing, mastery and gradual reorganization of relational expectations. A single strong prediction error may shift perception in an experimental task, but mastery as CMT describes typically unfolds across numerous, spaced relational encounters. This suggests the need for multitimescale models and longitudinal methods that can capture both phasic neural

surprise signals and the slow accrual of corrective interpersonal experiences.

Third, heterogeneity in change pathways is substantial and consequential. Not all therapeutic gains will be produced primarily by in-session patient testing and therapist disconfirmation. For some people, powerful environmental shifts (e.g., leaving an abusive partner, obtaining stable housing, or forming a new supportive friendship) may supply the bulk of corrective experiences needed for model revision, with therapy playing a facilitative rather than central role. For others, bottom-up sensory or pharmacological interventions may alter affective salience and open a window for testing that would otherwise be impossible. These differences imply that the integrated account should be positioned as a mechanistic route to change rather than a single universal pathway.

Finally, these caveats point to clear empirical priorities rather than fatal flaws. Empirical work should (a) operationalize motivational constructs in ways that can be integrated with computational models, for example, by measuring felt agency, epistemic trust, and willingness to test alongside precision-related neural signals; (b) adopt multitimescale longitudinal designs that link momentary prediction-error markers with slow changes in belief conviction and interpersonal behavior; and (c) explicitly test heterogeneity by comparing patients who change primarily via in-session testing with those whose change is driven by external, nontherapeutic events. Doing so will reveal when the predictive processing—CMT synthesis provides a powerful explanatory lens and where alternative or additional mechanisms are required.

Conclusions

In sum, although both predictive processing and CMT involve mechanisms of belief updating, they differ in their underlying assumptions. Predictive processing describes how the brain adjusts its predictions based on environmental input, prioritizing uncertainty minimization as a survival mechanism, predicting to control one's environment. CMT emphasizes an intrinsic drive toward self-healing and mastery, viewing belief revision as a process shaped by unconscious goals that extend beyond error minimization, testing to master problems and reclaim agency over life. Integrating these perspectives provides a

richer understanding of how individuals revise their generative models in psychotherapy, not only to reduce uncertainty but also to foster psychological growth and self-transformation. We reiterate that this contribution represents a theoretical integration, offering a coherent conceptual account and generating testable hypotheses. It is not a clinical protocol or a substitution for existing therapeutic models.

Understanding the relationship between patient testing, prediction error minimization, and belief updating also presents new opportunities for empirical research. Future studies could integrate neurobiological measures of prediction error processing with behavioral indices of patient testing in psychotherapy. Multilevel data collection—including session recordings, computational modeling, and self-reported belief change—could provide a more precise picture of how the brain processes therapist responses during patient testing. Additionally, simulations that model the iterative interactions between therapist-induced prediction errors and patient-initiated tests could further refine our understanding of psychotherapeutic change and strengthen both theory and practice.

References

Ahn, H. N., & Wampold, B. E. (2001). Where oh where are the specific ingredients? A meta-analysis of component studies in counseling and psychotherapy. *Journal of Counseling Psychology*, *48*(3), 251–257. https://doi.org/10.1037/0022-0167.48.3.251

Alexander, F., & French, T. M. (1946). The principle of corrective emotional experience. In F. Alexander & T. M. French (Eds.), *Psychoanalytic therapy: Principles and application* (pp. 66–96). Ronald Press.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Bohart, A. C., & Tallman, K. (2022). Client expertise: The active client in psychotherapy. In J. N. Fuertes (Ed.), *The other side of psychotherapy: Understanding clients' experiences and contributions in treatment* (pp. 13–43). American Psychological Association. https://doi.org/10.1037/0000303-002

Bugas, J., McCollum, J., Kealy, D., Silberschatz, G., Curtis, J. T., & Reid, J. (2023). Identifying patient verbal coaching in psychotherapy. *Counselling and Psychotherapy Research*, *23*(1), 247–257. https://doi.org/10.1002/capr.12495

Bugas, J., & Silberschatz, G. (2000). How patients coach their therapists in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *37*(1), 64–70. https://doi.org/10.1037/h0087676

Carey, T. A. (2010). Will you follow while they lead? Introducing a patient-led approach to low-intensity CBT interventions. In J. Bennett-Levy, D. A. Richards, P. Farrand, H. Christensen, K. M. Griffiths, D. J. Kavanaugh, B. Klein, M. A. Lau, J. Proudfoot, L. Ritterband, J. White, & C. Williams (Eds.), *Oxford guide to low-intensity CBT interventions* (pp. 331–338). Oxford University Press. https://doi.org/10.1093/med:psych/9780199590117.003.0034

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The *p* factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Chamberlin, D. E. (2019). The predictive processing model of EMDR. *Frontiers in Psychology*, *10*, Article 2267. https://doi.org/10.3389/fpsyg.2019.02267

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2023). *The experience machine: How our minds predict and shape reality*. Pantheon Books.

Connolly, P. (2022). Instability and uncertainty are critical for psychotherapy: How the therapeutic alliance opens us up. *Frontiers in Psychology*, *12*, Article 784295. https://doi.org/10.3389/fpsyg.2021.784295

Constantino, M. J., Coyne, A. E., Boswell, J. F., Iles, B. R., & Vîslă, A. (2018). A meta-analysis of the association between patients' early perception of treatment credibility and their posttreatment outcomes. *Psychotherapy*, *55*(4), 486–495. https://doi.org/10.1037/pst0000168

Fabry, R. E. (2020). Into the dark room: A predictive processing account of major depressive disorder. *Phenomenology and the Cognitive Sciences*, *19*(4), 685–704. https://doi.org/10.1007/s11097-019-09635-4

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, Article 215. https://doi.org/10.3389/fnhum.2010.00215

Friston, K. (2003). Learning and inference in the brain. *Neural Networks*, *16*(9), 1325–1352. https://doi.org/10.1016/j.neunet.2003.06.005

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B:*

Biological Sciences, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), Article e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gazzillo, F., Mannocchi, C., Curtis, J., Biuso, G. S., De Luca, E., Fimiani, R., Fiorenza, E., Genova, F., La Stella, M., Leonardi, J., Rodomonti, M., & Silberschatz, G. (2025). Passing patients' tests and following patients' coaching communications in psychotherapy: An empirical study. *Counselling Psychology Quarterly*, *38*(1), 160–186. https://doi.org/10.1080/09515070.2024.2354283

Goldfried, M. R. (1980). Toward the delineation of therapeutic change principles. *American Psychologist*, *35*(11), 991–999. https://doi.org/10.1037/0003-066X.35.11.991

Goldfried, M. R. (2012). The corrective experience: A core principle for therapeutic change. In L. G. Castonguay & C. E. Hill (Eds.), *Transformation in psychotherapy: Corrective experiences across cognitive-behavioral, humanistic, and psychodynamic approaches* (pp. 13–29). American Psychological Association. https://doi.org/10.1037/13747-002

Herzog, P., Kube, T., & Fassbinder, E. (2022). How childhood maltreatment alters perception and cognition - The predictive processing account of borderline personality disorder. *Psychological Medicine*, *52*(14), 2899–2916. https://doi.org/10.1017/S0033291722002458

Hohwy, J. (2013). *The predictive mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Kealy, D., Gazzillo, F., Silberschatz, G., & Curtis, J. T. (2020). Plan-compatible termination in psychotherapy: Perspectives from control–mastery theory. *Psychotherapy*, *57*(4), 508–514. https://doi.org/10.1037/pst0000300

Kealy, D., McCollum, J., Curtis, J. T., Silberschatz, G., Aafjes-van Doorn, K., & Luo, X. (2022). Failure to respond to the patient's coaching: A case study of premature termination in psychodynamic psychotherapy. *Counselling Psychology Quarterly*, *35*(4), 789–813. https://doi.org/10.1080/09515070.2021.2000941

Kindt, M., Buck, N., Arntz, A., & Soeter, M. (2007). Perceptual and conceptual processing as predictors of treatment outcome in PTSD. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*(4), 491–506. https://doi.org/10.1016/j.jbtep.2007.10.002

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. https://doi.org/10.1016/j.tins.2004.10.007

Krupnik, V. (2019). Bayesian Approach to psychotherapy integration: Strategic modification of priors. *Frontiers in Psychology*, *10*, Article 356. https://doi.org/10.3389/fpsyg.2019.00356

Kube, T., Berg, M., Kleim, B., & Herzog, P. (2020). Rethinking posttraumatic stress disorder - A predictive processing perspective. *Neuroscience and Biobehavioral Reviews*, *113*, 448–460. https://doi.org/10.1016/j.neubiorev.2020.04.014

Li, E., Midgley, N., Campbell, C., & Luyten, P. (2025). A theory-building case study of resolving epistemic mistrust and developing epistemic trust in psychotherapy with depressed adolescents. *Psychotherapy Research*. Advance online publication. https://doi.org/10.1080/10503307.2025.2473927

Loewald, H. W. (1960). On the therapeutic action of psycho-analysis. *International Journal of Psycho-Analysis*, *41*, 16–33. https://psycnet.apa.org/record/1961-06593-001

Norcross, J. C., & Goldfried, M. R. (2005). *Handbook of psychotherapy integration* (2nd ed.). Oxford University Press.

Putica, A., Felmingham, K. L., Garrido, M. I., O'Donnell, M. L., & Van Dam, N. T. (2022). A predictive coding account of value-based learning in PTSD: Implications for precision treatments. *Neuroscience and Biobehavioral Reviews*, *138*, Article 104704. https://doi.org/10.1016/j.neubiorev.2022.104704

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Seth, A. K. (2019). From unconscious inference to the beholder's share: Predictive perception and human experience. *European Review*, *27*(3), 378–410. https://doi.org/10.1017/S1062798719000061

Silberschatz, G. (2005). *Transformative relationships: The control–mastery theory of psychotherapy*. Routledge. https://doi.org/10.4324/9780203955963

Silberschatz, G. (2008). How patients work on their plans and test their therapists in psychotherapy. *Smith College Studies in Social Work*, *78*(2–3), 275–286. https://doi.org/10.1080/00377310802114528

Silberschatz, G., & Curtis, J. T. (1993). Measuring the therapist's impact on the patient's therapeutic progress. *Journal of Consulting and Clinical Psychology*, *61*(3), 403–411. https://doi.org/10.1037/0022-006X.61.3.403

Silberschatz, G., Curtis, J. T., & Nathans, S. (1989). Using the patient's plan to assess progress in psychotherapy. *Psychotherapy: Theory, Research, Practice, Training*, *26*(1), 40–46. https://doi.org/10.1037/h0085403

Van de Cruys, S., & Van Dessel, P. (2021). Mental distress through the prism of predictive processing theory. *Current Opinion in Psychology*, *41*, 107–112. https://doi.org/10.1016/j.copsyc.2021.07.006

Villiger, D. (2025). An integrative model of psychotherapeutic interventions based on a predictive processing framework. *Journal of Contemporary Psychotherapy*, *55*(1), 39–49. https://doi.org/10.1007/s10879-024-09637-7

von Helmholtz, H. (1867). *Handbuch der physiologischen Optik* [Handbook of physiological optics]. Leopold Voss. https://doi.org/10.3931/e-rara-21259

Weiss, J. (1993). *How psychotherapy works: Process and technique*. Guilford.

Weiss, J. (1997). The role of pathogenic beliefs in psychic reality. *Psychoanalytic Psychology*, *14*(3), 427–434. https://doi.org/10.1037/h0079734

Yon, D., Heyes, C., & Press, C. (2020). Beliefs and desires in the predictive brain. *Nature Communications*, *11*(1), Article 4404. https://doi.org/10.1038/s41467-020-18332-9