Review

Biological Psychiatry: GOS

Understanding Cognitive Behavioral Therapy for Psychosis Through the Predictive Coding Framework

Julia M. Sheffield, Aaron P. Brinen, Brandee Feola, Stephan Heckers, and Philip R. Corlett

ABSTRACT

Psychological treatments for persecutory delusions, particularly cognitive behavioral therapy for psychosis, are efficacious; however, mechanistic theories explaining why they work rarely bridge to the level of cognitive neuroscience. Predictive coding, a general brain processing theory rooted in cognitive and computational neuroscience, has increasing experimental support for explaining symptoms of psychosis, including the formation and maintenance of delusions. Here, we describe recent advances in cognitive behavioral therapy for psychosis–based psychotherapy for persecutory delusions, which targets specific psychological processes at the computational level of information processing. We outline how Bayesian learning models employed in predictive coding are superior to simple associative learning models for understanding the impact of cognitive behavioral interventions at the algorithmic level. We review hierarchical predictive coding as an account of belief updating rooted in prediction error signaling. We examine how this process is abnormal in psychotic disorders, garnering noisy sensory data that is made sense of through the development of overly strong delusional priors. We argue that effective cognitive behavioral therapy for psychosis systematically targets the way sensory data are selected, experienced, and interpreted, thus allowing for the strengthening of alternative beliefs. Finally, future directions based on these arguments are discussed.

https://doi.org/10.1016/j.bpsgos.2024.100333

Persecutory delusions, strongly held beliefs that one is under threat, are present in over 70% of individuals with a primary psychotic disorder (1). Persecutory delusions are associated with poorer quality of life, more hospitalizations, and increased suicidality (2). In the United States and Europe, cognitive behavioral therapy for psychosis (CBTp) is the recommended psychotherapeutic modality for treating delusions (3,4).

However, the cognitive and neuroscientific mechanisms through which CBTp impacts delusions remain poorly understood. Early versions of CBTp focused broadly on developing alternative explanations for psychotic experiences and reducing maladaptive coping (5). These approaches were applied to individuals with schizophrenia without systematically targeting specific symptoms, and they ultimately demonstrated small effects (6,7). However, newer, formulation-driven CBTp approaches organized around an updated psychological theory target specific maintenance factors with promising results (2) and offer an opportunity for increased understanding of how psychotherapy can impact delusions. To date, these targets of CBTp have been described in personal-level psychological terms; however, examining effective psychological treatments in a cognitive neuroscience framework may open avenues for treatment advancement (8,9). We explore this possibility in this article.

We make our case via David Marr's levels of processing framework (10). Marr suggested that when we analyze a cognitive system, we should delineate a computational level (which reflects the goals of the system), an algorithmic level (the representations and manipulations required to meet those goals), and an implementation level (how neural systems instantiate those algorithms). In psychotherapy, the therapist and patient often interact via the shared therapeutic targets at the computational level: worry less, leave the house more. However, at the algorithmic level lies belief updating—the driver of these behavioral targets—which is in turn supported by neurobiological processes at the circuit and cellular levels (11,12).

In this article, we outline how newer, formulation-driven CBTp for persecutory delusions targets psychological processes that impede learning about safety. While clinically effective, the cognitive neuroscience mechanisms that underlie its efficacy have not been well described. Although extinction learning theory is often assumed to underlie cognitive behavioral interventions, we argue that the effects of CBTp for delusions can be understood better through the framework of predictive coding. Specifically, we outline how abnormal prediction error signaling disrupts hierarchical belief updating by increasing sensory uncertainty and strengthening delusional priors. Formulation-driven CBTp rebalances this system by shifting what sensory data is taken in, thereby influencing the perceived precision of that data and concurrently strengthening nondelusional beliefs. We end with future directions, including treatments that target specific aspects of belief updating (e.g., volatility expectations), application of predictive coding to other types of psychotic symptoms beyond safety learning, and clinical recommendations.

CBTp TREATMENT FOR PERSECUTORY DELUSIONS

In 1952, Beck published a case study of the successful treatment of a patient with a delusion that he was being targeted/monitored by the government (i.e., G-Men) using a treatment that resembled modern cognitive therapy (13). Beck asked the man to define what a G-Man looked like and then compare the definition of G-Man with all the people he feared were G-Men. Over time, the patient disqualified each person he feared, reducing the strength of his belief. Decades later, researchers more formally applied Beck's cognitive model of anxiety disorders to psychosis, yielding CBTp. Early work in psychosis found that directly challenging delusional beliefs does not improve delusion severity and can instead lead to their strengthening and elaboration (14–16). Therefore, CBTp promotes the notion of collaborative empiricism, in which the therapist adopts a curious, nonconfrontational stance to understand the belief system with the patient, and they learn together how evidence from the world supports it. The first wave of CBTp worked to conceptualize psychosis and schizophrenia within a unified and general model and focused broadly on alternative explanations for distressing experiences, with only small effects (6). However, newer, formulation-driven CBTp treatments specifically target a single symptom, such as persecutory delusions (2), thus providing focused treatment based on relevant psychological processes.

One example of this type of formulation-driven CBTp treatment is the Feeling Safe Programme, which has shown large effects in improving persecutory delusion severity (17). This manualized protocol, based on the threat anticipation model (18), redefines persecutory delusions as unfounded threat beliefs that are maintained by psychological factors that block or impede learning about safety. Through rigorous causal inference testing (19), these aberrant psychological processes (which encourage and underlie the persecutory belief) have been shown to include worry, negative self-beliefs, anomalous experiences, safety behaviors, reasoning biases, and sleep disruption. Importantly, this approach does not involve confronting patients at the level of their delusional belief by directly arguing against it. Instead, it targets affective and cognitive accessories of the belief that the patient agrees are bothersome (20) and that keep the person feeling unsafe. Ultimately, the goal is to increase beliefs about safety that compete with the persecutory belief that one is under threat.

Fundamental interventions of CBT are applied to address these maintenance factors by targeting cognitions and behaviors to promote learning about safety. First, psychoeducation is provided about the psychological experience (e.g., worrying). Second, person-specific examples are elicited and examined with regard to how the maintenance factors impact the individual's daily life. Third, behavioral experiments are collaboratively developed with the patient to directly target the maintenance factor. For example, when targeting worry, the individual begins using worry periods to limit their worry during the day and plans alternative activities to support this worry reduction. Fourth, the therapist and patient conduct these behavioral experiments in session, with the therapist promoting alternative beliefs and explanations, which are written down by the patient to allow for out-of-session practice.

Critically, this type of formulation-driven CBTp treatment is active and involves garnering new experiences. Behavioral experiments are conducted in the real world to provide opportunities for more adaptive learning. Delusions are not contradicted or demeaned during the experiments, but rather alternative beliefs are promoted and reinforced for the patient. In the Feeling Safe Programme, for example, beliefs about being under threat (e.g., "the organization is out to get me") are placed in competition with new beliefs about safety ("maybe I am safer than I thought"). This is in contrast to habituation-based exposure therapy, where patients simply learn that fear can be tolerated if they stay in a situation long enough. Experimental data show that learning through trying new things has a stronger impact on reducing delusion conviction and distress than habituation-based exposure (21). While habituating to feared experiences can support future learning (22), it does not target the development of new safety beliefs. The development of new beliefs that directly compete with the delusion ("I am safe") is achieved through active exposure-based learning in which maladaptive safety behaviors (e.g., avoidance) are dropped and new experiences and explanations are sought out.

Extinction Learning and CBTp

It is notable that despite decades of work advancing CBTp-based treatment, the algorithmic mechanisms that underlie CBTp have not been described well. The cognitive and neuroscientific impact of CBT for other disorders, such as anxiety disorders, posttraumatic stress disorder, and obsessive-compulsive disorder, have been closely examined in terms of associative learning, with a particular focus on extinction learning (23–25); this reflects new (often Pavlovian) learning that a once-reinforced cue (conditioned stimulus [CS+]) (previously predictive of a salient outcome) now no longer portends something important or salient. Given that CBTp is based on the fundamentals of CBT for these disorders, the same foundational extinction learning mechanisms are assumed to be involved but have notable differences in individuals with delusions.

There is increasing evidence that fear and extinction learning, the key aspect of associative learning for updating threat beliefs, differs in individuals with schizophrenia and is associated with delusion severity (26–28). Individuals with schizophrenia demonstrate heightened fear response to neutral/safe cues (CS−) as measured by galvanic skin response. Elevated skin response to the neutral/safe cues is strongly and specifically associated with delusion severity (including persecutory delusions) (26), suggesting that heightened arousal in safe contexts may contribute to delusions (27). The heightened arousal to safe contexts may impair associative learning and drive new and aberrant associations with threat (i.e., the delusion metastasizes, incorporating spurious associations into the overarching threat belief). Furthermore, high delusion–prone individuals show elevated threat ratings to the previously threatening stimuli even after extinction learning compared with low delusion–prone individuals, again suggesting that individuals with more delusional thinking have difficulty learning safety cues (29). These fear conditioning paradigms offer foundational evidence that in laboratory experiments, individuals with delusions have a heightened fear response to neutral/safe stimuli that reflects aberrant salience and contributes to difficulty relearning safety (30). In the next section, we introduce predictive coding as a mechanistic account for this competition between adaptive and maladaptive beliefs.

PREDICTIVE CODING MODEL OF PERSECUTORY DELUSIONS

Predictive Coding Model

Associative learning models that form the basis of behavior therapy interventions, such as exposure therapy, offer a critical starting point for understanding the learning mechanisms targeted by CBTp. However, these models are limited. Algorithmically, the learning models that underlie exposure therapy, such as Rescorla-Wagner, assume that changes in a belief occur through the weakening of weights associated with that belief (31). What is increasingly being recognized in both the cognitive neuroscience and treatment literatures, however, is that beliefs are not unlearned but are instead outcompeted by alternative beliefs (32). Bayesian models are more suitable for capturing this. In addition, as noted, individuals with delusions experience neutral or safe stimuli as more threatening, which limits the translation of extinction learning paradigms to the modeling of delusions. Finally, associative learning lacks complex consideration of 2 clinical realities: 1) sensory experiences are often uncertain and therefore are open to interpretation, and 2) some beliefs are particularly resistant to change. These clinical realities can be placed within the predictive coding framework through the concepts of precision (33). We propose that predictive coding offers a framework for appreciating these complexities that can better explain the efficacy of CBTp for treating persecutory delusions.

According to predictive coding theory, the brain constantly makes predictions based on prior beliefs that are tuned by past experiences: for example, "when I turn the key in my car, it will turn on." When predictions are at odds with experiences ("my car did not turn on"), we are surprised and search for an explanation. By signaling that a mismatch between expectation and experience has occurred, the brain can adapt and minimize future surprise. The predictive coding model posits that beliefs form and are maintained by these types of surprising experiences, called prediction errors (34).

The goal of predictive coding is to build a consistent, reliable, and predictable model of the world (35). This goal is achieved in part through active inference, or interacting with the world by gathering information that supports prior predictions. The generative model space that develops from this active inference represents a range of prior beliefs about the world, which are continuously updated and optimized to minimize net prediction error (i.e., reduce uncertainty, surprise) (36).

In this framework, beliefs are updated hierarchically, with low-level sensory data influencing higher-level cognitive priors and vice versa (37). Critically, both sensory data and prior beliefs have a level of precision or certainty (inverse variance) that affects the magnitude of their influence on belief updating. More precise sensory data have a stronger influence on shifting a high-level prior than data that are noisy or uncertain. In contrast, a very strongly held belief is difficult to shift, even in the face of precise data. An empty gas gauge is clear data that my car did not turn on because it is out of gas; however, for someone with a strongly held (certain) belief that they have gas in their car already, an empty sign may suggest that the gauge had been tampered with. It may also support a higher-level belief that their coworkers are messing with them. This appreciation for the precision of prior beliefs and sensory data is limited in Rescorla-Wagner models of associative learning but is necessary for optimal inference and learning (33). Precision can be modeled for predictions (reflecting reliability of top-down expectations) or prediction errors (reflecting uncertainty of bottom-up signaling). Therefore, precision can influence the experience of salience in sensory data as well as the certainty with which beliefs are maintained (38).

One important implication of predictive coding is that sensory experiences and beliefs are inextricably linked. Sensory data are observed through the lens of high-level beliefs, thereby shaping perception (39). Therefore, very strong high-level priors are difficult to change in part because inferences about sensory experiences are perceived as evidence that support the strongly held belief. Prediction errors produced by low-level sensory data make their way up the hierarchy and, in the face of a very precise high-level belief, are more likely to strengthen or maintain that belief than change or update it (40).

At the level of implementation, prediction errors are encoded in dopamine signals in the striatum and midbrain (12,41). These signals are supported by a broad network of brain regions, including the amygdala, insula, prefrontal cortex, and cingulate, that promote domain-general prediction errors across reward, cognitive, social, or perceptual tasks (12). As has been detailed elsewhere (42), adaptive firing of prediction errors in response to dopamine signaling results in effective learning about the world and appropriate updating of one's prior beliefs to guide future behaviors.

Delusions as Overly Strong High-Level Priors

The predictive coding model suggests that delusional beliefs are strongly held high-level priors that develop in response to aberrant prediction errors or those that occur independent of meaningful cue and context (43). Because prediction errors drive learning, this aberrant salience promotes learning from stimuli that should be interpreted as neutral or meaningless (44). These stimuli are experienced as salient and perceived through the lens of the delusional prior, thus perpetuating false inference and strengthening the belief (45).

In support of this theory, experimental work has demonstrated that individuals in the early stages of psychosis rely more on priors higher in the hierarchy to resolve uncertain information (23,24). Interestingly, individuals at risk for psychosis show weaker low-level perceptual priors (23), consistent with the idea that higher-level priors (those more proximal to cognition than perception) may strengthen to compensate for prediction errors that ascend from lower levels, thus driving delusion formation (46,47). This occurs against the backdrop of elevated presynaptic dopamine within the striatum in people at risk for, and currently experiencing, psychotic disorders (48,49); this elevated presynaptic dopamine within the striatum is associated with positive symptom severity (48,50). Task-based functional neuroimaging studies have demonstrated responses that poorly distinguish surprising from unsurprising events in the midbrain, striatum, amygdala, and prefrontal cortex in psychotic disorders (43,51,52), which relate to delusion severity (43). This may be driven in part by an elevated response to neutral or unsurprising stimuli, which is consistently observed in schizophrenia (30). Aberrant prediction errors thereby promote learning from otherwise meaningless information, which fuels inappropriate belief updating.

Growing evidence suggests that paranoia and persecutory delusions are associated with one particular high-level prior: that the environment is volatile (frequently changing) (53–56). Individuals with elevated paranoia overestimate the volatility of their environments, meaning that they expect the hidden state of the world to change more frequently. The severity of persecutory delusions is related to elevated volatility priors during nonsocial (57) and social learning tasks (54). Recent work has shown impaired learning about environmental volatility in youths at clinical high risk for psychosis who self-report high levels of paranoia (58). Critically, a lower learning rate measured during 2 nonsocial tasks was associated with paranoia, but not with nonparanoid delusional ideation, suggesting that the prediction errors that govern learning are imprecise in individuals with high paranoia, which impairs associative learning.

Persecutory delusions are also related to excessive choice switching, a behavior consistently seen in schizophrenia (59–61), and is thought to reflect overly strong volatility priors (62). Because volatility priors interact in a hierarchical fashion with sensory evidence, excessive choice switching may reflect greater “noise” in their overall decision making when presented with more information (63,64), due to imprecise sensory data (37). Therefore, increased precision of higher-level predictions (e.g., about the volatility of the environment) (56,65) may actually compensate for reduced precision of lower-level predictions (66). To make matters worse, individuals with schizophrenia demonstrate overreliance on prediction errors during learning (67), which is also correlated with delusion severity (43,68,69). This could drive a more precise high-level belief that the world is volatile and unpredictable, contributing to the even higher-level belief that one is unsafe and under threat.

In summary, the predictive coding model expands upon stimulus-response fear conditioning paradigms by weighting learning by the strength (precision) of beliefs and sensory data and by appreciating the influence of strong top-down beliefs on the perception of sensory experiences. This hierarchical Bayesian model is more suitable for understanding belief updates related to delusions than simple associative learning models that have often been used to understand CBT for other disorders (45). A stronger prior requires a stronger violation to update a person’s internal model. Overly precise priors in individuals with delusions, such as a heightened expectation of environmental volatility, will therefore be more resistant to change in the face of contradictory evidence. In addition, noisy or imprecise sensory data, driven in part by excessive prediction error signaling, can strengthen the higher-order threat belief (70). This implies a hierarchical vulnerability to excessive belief updating that ultimately strengthens cognitive priors, thus making them more resistant to change.

CBTp TREATMENT EFFECTS IN THE PREDICTIVE CODING MODEL

In Table 1, we outline how each of the most effective components of Feeling Safe impacts belief updating. Because patients experience neutral stimuli as meaningful or threatening (aberrant salience), low-level sensory data can inadvertently reinforce their delusional belief. In addition, high-level delusional priors are excessively precise (certain) and therefore resistant to change. The power of Feeling Safe may be in the fact that it intervenes primarily at the level of sensory data rather than high-level beliefs. Alternative beliefs are gently offered but tend to focus on sensory beliefs (“what did you notice?”) as opposed to the delusion itself (“if that happened, the organization is probably not out to get you, right?”). Feeling Safe systematically focuses on improving the quality of sensory data that the patient takes in and adjusting low-level sensory beliefs, thereby increasing their precision. Over time, more precise sensory data compete with the delusional prior and shifts the posterior belief so as to be consistent with the person’s increasingly adaptive experiences. In Figure 1, we illustrate how, over time, targeting these factors reduces the weight of the threat belief and reinforces alternative nondelusional safety belief, even in the face of aberrant salience.

This can be outlined for each of the psychological maintenance factors. Worrying focuses sensory data selection on threatening images and ideas of the worst-case scenario. Reducing worry helps replace the sensory data of threatening images with sensory data associated with positive, meaningful experiences; this allows for information supporting safety to be taken in and strengthened. Interestingly, in our own data, worry was associated with elevated priors on volatility, an effect that was mediated by increased paranoia (57). This suggests that expecting more environmental volatility at the algorithmic level contributes to feelings of paranoia and the tendency to worry at the computational level. This is best tested using longitudinal data, where we would expect lower volatility priors to be related to less paranoia and worry over time.

Negative self-beliefs create a negative mental filter over data collection, as has been observed in depression (71). Improving self-confidence increases access to more positive sensory data. Savoring positive experiences increases the strength of that data. Engaging in meaningful activities that are congruent with values promotes experiences that reinforce positive self-beliefs.

Anomalous experiences, such as hallucinations, are strong reinforcers of threat beliefs (72), particularly when they are considered highly credible (73). Reducing the frequency of voice hearing limits exposure to this type of sensory data. Questioning the credibility of the voices and the quality of that data reduces their weight; anomalous experiences are therefore less likely to reinforce the threat belief.

Finally, safety behaviors, such as avoidance, limit new data collection by prohibiting belief updates (74). Dropping safety behaviors involves going out into the world and seeking out new data while also noticing how this data is perceived and

Table 1. Impact of Psychological Processes on Sensory Data Selection and High-Level Priors

CBTp Treatment Target	Sensory Data Taken in That Maintains Persecutory Delusion	Sensory Data Taken in With CBTp Intervention	Impact of CBTp on Belief Updating
Worry	Thoughts and images of worst-case scenario	The usefulness of worry is questioned, and worry thoughts are reduced. This allows for increased nonworrisome thoughts about everyday life and meaningful experiences.	Worries are not treated as credible (precise) data. Instead, a broader range of sensory data is selected that are less threatening.
Negative Self-Beliefs	Negative mental filter focused on being incapable and vulnerable	Savoring and rehearsing data that are consistent with personal strengths and values	Data collection is biased toward positive experiences that reinforce beliefs about capability and strength. A more positive perspective is inferred from sensory experiences.
Anomalous Experiences	Hallucinations are meaningful, credible, and important data.	Hallucinations are experienced less frequently and are not perceived as credible sources of information about oneself and the world.	Because they are considered less credible, hallucinations are weighted as less precise and have less impact on belief updating.
Safety Behaviors	No new sensory data taken. Reinforces that the situation was dangerous, and danger was only averted because of avoidance/safety behavior	Sensory data that are perceived as safe is sought out, noticed, and reinforced. Feared outcome is tested. Internal experiences are reattributed to being reactive vs. a sign of danger.	An alternative, posterior belief is strengthened through gathering sensory data that support it. Repeated exposures with new information strengthen the alternative belief and weaken the threat belief. New data are perceived in support of the alternative belief, further strengthening it.

A description of how each psychological process that is targeted in formulation-driven CBTp for persecutory delusions impacts sensory data selection and high-level priors. CBTp changes the way sensory data are taken in and interpreted, reducing the strength of the delusional prior and providing the opportunity for an alternative nondelusional posterior belief to develop.

CBTp, cognitive behavioral therapy for psychosis.

interpreted. Together with the therapist, sensory data from both the internal and external milieu are closely noticed, and interpretations of it are interrogated. The alternative belief that "I am safe enough" is reinforced based specifically on incoming data.

More direct tests of this integrated model could determine whether psychotherapy normalizes maladaptive priors, for example by lowering expectations of environmental volatility. Distinguishing the hierarchical level at which CBTp has its effect could also be determined through testing whether changes in the experience of aberrant salience precede changes in high-level cognitive priors, as predicted by the model.

FUTURE DIRECTIONS

Therefore, if predictive coding is going to be a useful framework for treatment advancement, where can we go from here? We offer a few ideas.

1. Predictive coding experiments strongly suggest that paranoia is associated with overestimation of environmental volatility. Current CBTp formulations do not explicitly target volatility, focusing on safety instead. While predictability and safety are strongly related, they are not the same, and their co-occurrence may be critical for shaping emotional health (75). Combining the algorithmic level of volatility priors with computational goals that can be targeted directly in therapy may identify novel treatment targets. For example, perhaps searching out safety, as is done in Feeling Safe, could be further enhanced by searching out predictability. This could be done by examining how the sense of volatility is experienced and reinforced in daily life and then seeking out predictable and consistent sensory data that also feels safe.
2. While our focus has been on persecutory delusions, we believe that this framework extends to other delusional themes. This is in fact a strength of the predictive coding framework—it is not simply focused on fear conditioning and threat beliefs but rather describes how beliefs about oneself, others, and the world are updated and maintained, regardless of content. Recent work on grandiose delusions, for example, has described psychological maintenance factors of grandiose beliefs that can be similarly targeted by psychological treatments at the computational level (76). As

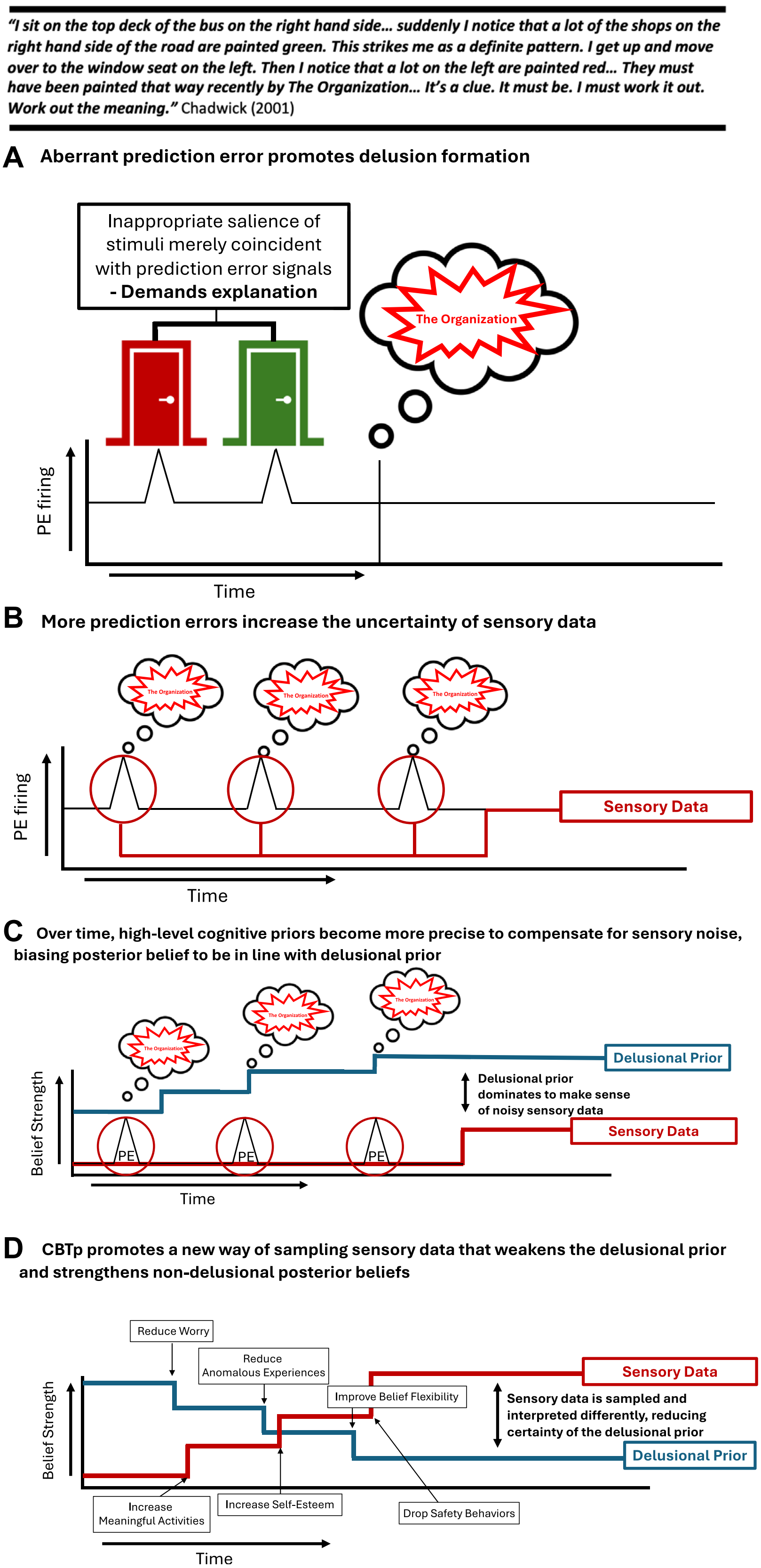

Figure 1. Schematic of how aberrant prediction error (PE) signaling impacts the development of overly strong delusional priors and how this can be influenced by psychotherapy. Panels **(A)** and **(B)** show the impact of aberrant PE firing on **(A)** increasing aberrant salience that demands explanation and **(B)** generating noisy sensory data. Panels **(C)** and **(D)** show the impact of this PE signaling on belief strength: **(C)** over time, delusional priors are strengthened to make sense of the noisy sensory data. For this individual, it may be a delusional prior that the organization is out to get him; **(D)** introducing cognitive behavioral therapy for psychosis (CBTp) promotes behaviors that change how sensory data are sampled and interpreted. This process results in increased certainty around sensory experiences and decreased certainty about the delusional prior, which strengthen nondelusional posterior beliefs.

with volatility expectations for paranoia, different maladaptive prior expectations may underlie grandiosity at the algorithmic level, which can be tested in behavioral paradigms.

3. Studies of fear conditioning in psychotic disorders have shown evidence of abnormalities at the level of implementation. During fear conditioning, there is reduced activation of the ventromedial prefrontal cortex in response to safety cues (77,78), particularly in patients with delusions (78). Reduced activation of the medial prefrontal cortex during the extinction phase of extinction learning is also observed in individuals with high delusion proneness (29). In addition, increased ventral striatum activity has been found in response to neutral (CS−) stimuli in schizophrenia (30). However, the impact of CBTp on these fear conditioning circuits, and on belief updating paradigms more broadly, has not been tested directly. In fact, only one known dataset has been used to examine changes in functional activation and connectivity pre-/postgeneric (i.e., not symptom-specific) CBTp. Interestingly, primary analyses from this study found 1) increased activation to neutral faces at baseline, and 2) a significant reduction in that activation with CBTp (79). Another analysis reported increased connectivity between prefrontal regions (dorsolateral prefrontal cortex) and regions of salience (e.g., insula) and threat (e.g., amygdala) following CBTp (80). Long-term changes in psychotic symptoms following CBTp were predicted by changes in prefrontal connections during an emotion processing task (81). These findings suggest that CBTp reduces neural activation to neutral stimuli, possibly through increased prefrontal control. Future studies examining neuroimaging markers of belief updating pre-/post-CBTp will be useful for further elucidating aspects of the predictive coding hierarchy that are most impacted by treatment.
4. We urge the following practices for clinicians working with delusional patients. First, resist the urge to confront the high-level delusional belief directly, because intervention at this level of hierarchy is not effective in weakening beliefs (14–16); it is more likely to harm the therapeutic alliance, which is critical for CBTp (82). Second, do not assume that you and the patient share a common experience of what stimuli are salient. Be curious about what the patient notices, how they perceive it, and what it means to them. Third, experiences are the strongest way to update a person's model of the world. Frequent, small behavioral experiments should be focused on or done in addition to 1 or 2 large weekly goals.

While we hope that the integration of these conceptual frameworks is useful, its potency must be determined with data. A limitation of our proposed model is that direct evidence for it is sparse. Whether treatment with CBTp directly influences belief updating processes in individuals with delusions will offer critical insights into the utility of the proposed framework. This will require conversation across disciplines, i.e., between those who develop psychological interventions and those who model cognitive neuroscience–based mechanisms of delusions.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the National Institutes of Health (Grant Nos. K23MH126313 [to JMS] and R01MH127018 [to BF]).

PRC is the cofounder and board member of Tetricus Labs, a computational psychiatry company. He owns equity in the company. The company did not fund the current research or any other work in PRC's laboratory. All other authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, Tennessee (JMS, APB, BF, SH); and the Department of Psychiatry, Clinical Neuroscience Research Unit, Yale School of Medicine, New Haven, Connecticut (PRC).

Address correspondence to Julia M. Sheffield, Ph.D., at Julia.sheffield@vumc.org.

Received Oct 11, 2023; revised May 2, 2024; accepted May 4, 2024.

REFERENCES

1. Coid JW, Ullrich S, Kallis C, Keers R, Barker D, Cowden F, Stamps R (2013): The relationship between delusions and violence: Findings from the East London first episode psychosis study. JAMA Psychiatry 70:465–471.
2. Freeman D (2016): Persecutory delusions: A cognitive perspective on understanding and treatment. Lancet Psychiatry 3:685–692.
3. National Collaborating Centre for Mental Health (2014): Psychosis and Schizophrenia in Adults: Treatment and Management. London: National Institute for Health and Care Excellence.
4. Substance Abuse and Mental Health Services Administration (2021): Routine Administration of Cognitive Behavioral Therapy for Psychosis as the Standard of Care for Individuals Seeking Treatment for Psychosis: State of the Science and Implementation Considerations for Key Stakeholders. Center for Mental Health Services, Substance Abuse and Mental Health Services Administration. Available at: https://store.samhsa.gov/sites/default/files/pep20-03-09-001.pdf. Accessed July 1, 2023.
5. Mander H, Kingdon D (2015): The evolution of cognitive–behavioral therapy for psychosis. Psychol Res Behav Manag 8:63–69.
6. Laws KR, Darlington N, Kondel TK, McKenna PJ, Jauhar S (2018): Cognitive Behavioural Therapy for schizophrenia – Outcomes for functioning, distress and quality of life: A meta-analysis. BMC Psychol 6:32.
7. Van der Gaag M, Valmaggia LR, Smit F (2014): The effects of individually tailored formulation-based cognitive behavioural therapy in auditory hallucinations and delusions: A meta-analysis. Schizophr Res 156:30–37.
8. Morrison AP, Law H, Carter L, Sellers R, Emsley R, Pyle M, *et al.* (2018): Antipsychotic drugs versus cognitive behavioural therapy versus a combination of both in people with psychosis: A randomised controlled pilot and feasibility study. Lancet Psychiatry 5:411–423.
9. Morrison AP (2019): Should people with psychosis be supported in choosing cognitive therapy as an alternative to antipsychotic medication: A commentary on current evidence. Schizophr Res 203:94–98.
10. Marr D (1982): Vision. San Francisco: Freeman and Company.
11. Gershman SJ, Uchida N (2019): Believing in dopamine. Nat Rev Neurosci 20:703–714.
12. Corlett PR, Mollick JA, Kober H (2022): Meta-analysis of human prediction error for incentives, perception, cognition, and action. Neuropsychopharmacology 47:1339–1349.
13. Beck AT (1952): Successful outpatient psychotherapy of a chronic schizophrenic with a delusion based on borrowed guilt. Psychiatry 15:305–312.
14. Milton F, Patwa VK, Hafner RJ (1978): Confrontation vs. belief modification in persistently deluded patients. Br J Med Psychol 51:127–130.
15. Greenwood VB (1983): Cognitive therapy with the young adult chronic patient. In: Freeman A editor. Cognitive Therapy with Couples and Groups. Springer, Boston, MA, 183–198.

16. Wincze JP, Leitenberg H, Agras WS (1972): The effects of token reinforcement and feedback on the delusional verbal behavior of chronic paranoid schizophrenics1. J Appl Behav Anal 5:247–262.
17. Freeman D, Emsley R, Diamond R, Collett N, Bold E, Chadwick E, *et al.* (2021): Comparison of a theoretically driven cognitive therapy (the Feeling Safe Programme) with befriending for the treatment of persistent persecutory delusions: A parallel, single-blind, randomised controlled trial. Lancet Psychiatry 8:696–707.
18. Freeman D, Garety PA, Kuipers E, Fowler D, Bebbington PE (2002): A cognitive model of persecutory delusions. Br J Clin Psychol 41:331–347.
19. Brown P, Waite F, Freeman D (2019): 'Twisting the lion's tail': Manipulationist tests of causation for psychological mechanisms in the occurrence of delusions and hallucinations. Clin Psychol Rev 68:25–37.
20. Freeman D, Taylor KM, Molodynski A, Waite F (2019): Treatable clinical intervention targets for patients with schizophrenia. Schizophr Res 211:44–50.
21. Freeman D, Bradley J, Antley A, Bourke E, DeWeever N, Evans N, *et al.* (2016): Virtual reality in the treatment of persecutory delusions: Randomised controlled experimental study testing how to reduce delusional conviction. Br J Psychiatry 209:62–67.
22. Benito KG, Walther M (2015): Therapeutic process during exposure: Habituation model. J Obsessive Compuls Relat Disord 6:147–157.
23. Putica A, Felmingham KL, Garrido MI, O'Donnell ML, Van Dam NT (2022): A predictive coding account of value-based learning in PTSD: Implications for precision treatments. Neurosci Biobehav Rev 138: 104704.
24. Graham BM, Milad MR (2011): The study of fear extinction: Implications for anxiety disorders. Am J Psychiatry 168:1255–1265.
25. Cooper SE, Dunsmoor JE (2021): Fear conditioning and extinction in obsessive–compulsive disorder: A systematic review. Neurosci Biobehav Rev 129:75–94.
26. Holt DJ, Lebron-Milad K, Milad MR, Rauch SL, Pitman RK, Orr SP, *et al.* (2009): Extinction memory is impaired in schizophrenia. Biol Psychiatry 65:455–463.
27. Tuominen L, Romaniuk L, Milad MR, Goff DC, Hall J, Holt DJ (2022): Impairment in acquisition of conditioned fear in schizophrenia. Neuropsychopharmacology 47:681–686.
28. Perez DL, Pan H, Weisholtz DS, Root JC, Tuescher O, Fischer DB, *et al.* (2015): Altered threat and safety neural processing linked to persecutory delusions in schizophrenia: A two-task fMRI study. Psychiatry Res 233:352–366.
29. Louzolo A, Lebedev AV, Björnsdotter M, Acar K, Ahrends C, Kringelbach ML, *et al.* (2022): Resistance to extinction of evaluative fear conditioning in delusion proneness. Schizophr Bull Open 3:sgac033.
30. Dugré JR, Bitar N, Dumais A, Potvin S (2019): Limbic hyperactivity in response to emotionally neutral stimuli in schizophrenia: A neuroimaging meta-analysis of the hypervigilant mind. Am J Psychiatry 176:1021–1029.
31. Rescorla RA, Wagner AR (1972): A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF, editors. Classical Conditioning: II. Current Research and Theory. New York: Appleton-Century-Crofts, 64–99.
32. Craske MG, Kircanski K, Zelikowsky M, Mystkowski J, Chowdhury N, Baker A (2008): Optimizing inhibitory learning during exposure therapy. Behav Res Ther 46:5–27.
33. Yu AJ, Dayan P (2005): Uncertainty, neuromodulation, and attention. Neuron 46:681–692.
34. Friston K, Kiebel S (2009): Predictive coding under the free-energy principle. Philos Trans R Soc Lond B Biol Sci 364:1211–1221.
35. Sprevak M, Smith R (2023): An introduction to predictive processing models of perception and decision-making. [published online Oct 29]. Top Cogn.
36. Friston K (2010): The free-energy principle: A unified brain theory? Nat Rev Neurosci 11:127–138.
37. Ashinoff BK, Singletary NM, Baker SC, Horga G (2022): Rethinking delusions: A selective review of delusion research through a computational lens. Schizophr Res 245:23–41.
38. Corlett PR, Honey GD, Fletcher PC (2016): Prediction error, ketamine and psychosis: An updated model. J Psychopharmacol 30:1145–1155.
39. Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, *et al.* (2015): Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci U S A 112:13401–13406.
40. Jern A, Chang KM, Kemp C (2014): Belief polarization is not always irrational. Psychol Rev 121:206–224.
41. Roeper J (2013): Dissecting the diversity of midbrain dopamine neurons. Trends Neurosci 36:336–342.
42. Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH (2010): Toward a neurobiology of delusions. Prog Neurobiol 92:345–369.
43. Corlett PR, Murray GK, Honey GD, Aitken MRF, Shanks DR, Robbins TW, *et al.* (2007): Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. Brain 130:2387–2400.
44. Nour MM, Dahoun T, Schwartenbeck P, Adams RA, FitzGerald THB, Coello C, *et al.* (2018): Dopaminergic basis for signaling belief updates, but not surprise, and the link to paranoia. Proc Natl Acad Sci U S A 115:E10167–E10176.
45. Sterzer P, Adams RA, Fletcher P, Frith C, Lawrie SM, Muckli L, *et al.* (2018): The predictive coding account of psychosis. Biol Psychiatry 84 634–643.
46. Haarsma J, Knolle F, Griffin JD, Taverne H, Mada M, Goodyer IM, *et al.* (2020): Influence of prior beliefs on perception in early psychosis: Effects of illness stage and hierarchical level of belief. J Abnorm Psychol 129:581–598.
47. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ (2013): The computational anatomy of psychosis. Front Psychiatry 4:47.
48. Howes OD, Montgomery AJ, Asselin M-C, Murray RM, Valli I, Tabraham P, *et al.* (2009): Elevated striatal dopamine function linked to prodromal signs of schizophrenia. Arch Gen Psychiatry 66:13–20.
49. Fusar-Poli P, Meyer-Lindenberg A (2013): Striatal presynaptic dopamine in schizophrenia, part II: Meta-analysis of [$^{18}F/^{11}C$]-DOPA PET studies. Schizophr Bull 39:33–42.
50. Avram M, Brandl F, Cabello J, Leucht C, Scherr M, Mustafa M, *et al.* (2019): Reduced striatal dopamine synthesis capacity in patients with schizophrenia during remission of positive symptoms. Brain 142:1813–1826.
51. Ermakova AO, Knolle F, Justicia A, Bullmore ET, Jones PB, Robbins TW, *et al.* (2018): Abnormal reward prediction-error signalling in antipsychotic naive individuals with first-episode psychosis or clinical risk for psychosis. Neuropsychopharmacology 43:1691–1699.
52. Knolle F, Ermakova AO, Justicia A, Fletcher PC, Bunzeck N, Düzel E, Murray GK (2018): Brain responses to different types of salience in antipsychotic naïve first episode psychosis: An fMRI study. Transl Psychiatry 8:196.
53. Sheffield JM, Smith R, Suthaharan P, Leptourgos P, Corlett PR (2023): Relationships between cognitive biases, decision-making, and delusions. Sci Rep 13:9485.
54. Hauke DJ, Wobmann M, Andreou C, Mackintosh AJ, de Bock R, Karvelis P, *et al.* (2024): Altered perception of environmental volatility during social learning in emerging psychosis. Comp Psychiatry 8:1–22.
55. Reed EJ, Uddenberg S, Suthaharan P, Mathys CD, Taylor JR, Groman SM, Corlett PR (2020): Paranoia as a deficit in non-social belief updating. eLife 9:1–55.
56. Suthaharan P, Reed EJ, Leptourgos P, Kenney JG, Uddenberg S, Mathys CD, *et al.* (2021): Paranoia and belief updating during the COVID-19 crisis. Nat Hum Behav 5:1190–1202.
57. Sheffield JM, Suthaharan P, Leptourgos P, Corlett PR (2022): Belief updating and paranoia in individuals with schizophrenia. Biol Psychiatry Cogn Neurosci Neuroimaging 7:1149–1157.

58. Rossi-Goldthorpe R, Silverstein SM, Gold JM, Schiffman J, Waltz JA, Williams TF, *et al.* (2024): Different learning aberrations relate to delusion-like beliefs with different contents. [published online Apr 18]. Brain.
59. Murray GK, Cheng F, Clark L, Barnett JH, Blackwell AD, Fletcher PC, *et al.* (2008): Reinforcement and reversal learning in first-episode psychosis. Schizophr Bull 34:848–855.
60. Waltz JA, Gold JM (2007): Probabilistic reversal learning impairments in schizophrenia: Further evidence of orbitofrontal dysfunction. Schizophr Res 93:296–303.
61. Culbreth AJ, Gold JM, Cools R, Barch DM (2016): Impaired activation in cognitive control regions predicts reversal learning in schizophrenia. Schizophr Bull 42:484–493.
62. Deserno L, Boehme R, Mathys C, Katthagen T, Kaminski J, Stephan KE, *et al.* (2020): Volatility estimates increase choice switching and relate to prefrontal activity in schizophrenia. Biol Psychiatry Cogn Neurosci Neuroimaging 5:173–183.
63. Moutoussis M, Bentall RP, El-Deredy W, Dayan P (2011): Bayesian modelling of Jumping-to-Conclusions bias in delusional patients. Cogn Neuropsychiatry 16:422–447.
64. Adams RA, Napier G, Roiser JP, Mathys C, Gilleen J (2018): Attractor-like dynamics in belief updating in schizophrenia. J Neurosci 38:9471–9485.
65. Reed EJ, Uddenberg S, Mathys CD, Taylor JR, Groman SM, Corlett PR (2020): Expecting the unexpected: The paranoid style of belief updating across species. bioRxiv. https://doi.org/10.1101/2020.02.24.963298.
66. Schmack K, Gòmez-Carrillo de Castro AGC, Rothkirch M, Sekutowicz M, Rössler H, Haynes JD, *et al.* (2013): Delusions and the role of beliefs in perceptual inference. J Neurosci 33:13701–13712.
67. Geana A, Barch DM, Gold JM, Carter CS, MacDonald AW, Ragland JD, *et al.* (2022): Using computational modeling to capture schizophrenia-specific reinforcement learning differences and their implications on patient classification. Biol Psychiatry Cogn Neurosci Neuroimaging 7:1035–1046.
68. Romaniuk L, Honey GD, King JR, Whalley HC, McIntosh AM, Levita L, *et al.* (2010): Midbrain activation during Pavlovian conditioning and delusional symptoms in schizophrenia. Arch Gen Psychiatry 67:1246–1254.
69. Schlagenhauf F, Sterzer P, Schmack K, Ballmaier M, Rapp M, Wrase J, *et al.* (2009): Reward feedback alterations in unmedicated schizophrenia patients: Relevance for delusions. Biol Psychiatry 65:1032–1039.
70. Sinclair AH, Barense MD (2019): Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. Trends Neurosci 42:727–739.
71. Beevers CG, Mullarkey MC, Dainer-Best J, Stewart RA, Labrada J, Allen JJB, *et al.* (2019): Association between negative cognitive bias and depression: A symptom-level approach. J Abnorm Psychol 128:212–227.
72. Hermans K, van der Steen Y, Kasanova Z, van Winkel R, Reininghaus U, Lataster T, *et al.* (2020): Temporal dynamics of suspiciousness and hallucinations in clinical high risk and first episode psychosis. Psychiatry Res 290:113039.
73. Hacker D, Birchwood M, Tudway J, Meaden A, Amphlett C (2008): Acting on voices: Omnipotence, sources of threat, and safety-seeking behaviours. Br J Clin Psychol 47:201–213.
74. Moutoussis M, Shahar N, Hauser TU, Dolan RJ (2018): Computation in psychotherapy, or how computational psychiatry can aid learning-based psychological therapies. Comput Psychiatr 2:50–73.
75. Kumari V, Fannon D, Peters ER, Ffytche DH, Sumich AL, Premkumar P, *et al.* (2011): Neural changes following cognitive behaviour therapy for psychosis: A longitudinal study. Brain 134:2396–2407.
76. Gee DG, Cohodes EM (2021): Influences of caregiving on development: A Sensitive period for biological embedding of predictability and safety cues. Current directions in psychological science 30:376–383.
77. Mason L, Peters ER, Dima D, Williams SC, Kumari V (2016): Cognitive behavioral therapy normalizes functional connectivity for social threat in psychosis. Schizophr Bull 42:684–692.
78. Isham L, Griffith L, Boylan AM, Hicks A, Wilson N, Byrne R, *et al.* (2021): Understanding, treating, and renaming grandiose delusions: A qualitative study. Psychol Psychother 94:119–140.
79. Mason L, Peters E, Williams SC, Kumari V (2017): Brain connectivity changes occurring following cognitive behavioural therapy for psychosis predict long-term recovery. [published correction appears in Transl Psychiatry 2017; 7:e1209] [published correction appears in Transl Psychiatry 2017; 7:e1210].Transl Psychiatry 7:e1001.
80. Quarmley M, Gur RC, Turetsky BI, Watters AJ, Bilker WB, Elliott MA, *et al.* (2019): Reduced safety processing during aversive social conditioning in psychosis and clinical risk. Neuropsychopharmacology 44:2247–2253.
81. Holt DJ, Coombs G, Zeidan MA, Goff DC, Milad MR (2012): Failure of neural responses to safety cues in schizophrenia. Arch Gen Psychiatry 69:893–903.
82. Goldsmith LP, Lewis SW, Dunn G, Bentall RP (2015): Psychological treatments for early psychosis can be beneficial or harmful, depending on the therapeutic alliance: An instrumental variable analysis. Psychol Med 45:2365–2373.