HHS Public Access

Author manuscript

Cogn Neuropsychiatry. Author manuscript; available in PMC 2018 January 30.

Published in final edited form as:

Cogn Neuropsychiatry. 2016 ; 21(1): 73–89. doi:10.1080/13546805.2015.1136206.

The doxastic shear pin: delusions as errors of learning and memory

S.K. Fineberg[a] and **P.R. Corlett**[a,*]

[a]Yale University, Department of Psychiatry, Ribicoff Research Facility. 34 Park Street, New Haven, CT, USA 06519

Abstract

We reconsider delusions in terms of a "doxastic shear pin", a (cognitive) mechanism that errs so as to prevent the destruction of the (brain) machine and to permit continued function at an albeit attenuated capacity. Delusions may disable flexible (but energetically expensive) inference, leaving the increasingly habitual delusion to flourish. With each recall, delusions are reinforced further, strengthened, and made more resistant to contradiction. Psychodynamic theories posit benefit from delusions in the form of defensive function. We aim to respond to deficit accounts of delusions – that delusions would be only a problem without any benefit – by considering delusion formation and maintenance in terms of predictive coding.

We posit that brains conform to a simple computational principle: to minimize prediction error (the mismatch between prior top-down expectation and current bottom-up input) across hierarchies of brain regions and psychological representation. Recent data suggests that delusions may form in the absence of constraining top-down expectations. Then, once formed, they become new prior expectations (priors) that motivate other beliefs, perceptions, and actions by providing strong (sometimes overriding) top-down expectation.

We argue that delusions form when the shear-pin functions correctly – by breaking. This permits continued engagement with an overwhelming world, and ongoing function in the face of what was a paralyzing difficulty. This crucial role should not be ignored when we treat delusions: we need to consider how a person will function in the world without them.

Keywords

Belief; delusion; epistemic benefit; prediction error; schizophrenia

I. Introduction

Delusions are fixed false beliefs that are considered cardinal symptoms of schizophrenia (Garety, 1992). They can be very distressing for the people who espouse them. Therefore, it may seem mischievous to consider potential adaptive functions of delusions (Hingley, 1992;

*Corresponding Author: Connecticut Mental Health Center, Yale University School of Medicine, Department of Psychiatry, 34 Park Street, New Haven, CT 06519, Office: 01 203 974 7866, Fax: 01 203 974 7662, philip.corlett@yale.edu.

Disclosures: none

Roberts, 1991). Dennett and McKay have explored adaptive misbeliefs – incorrect beliefs that, though wrong, may be helpful (R. T. McKay & Dennett, 2009). They argued that only positive illusions are adaptive misbeliefs. Psychodynamic theory holds that some psychotic beliefs serve a defensive function – for example, grandiose beliefs protect against low self-esteem (Neale, 1988). Evolutionary theorists have made similar claims: if a person convinces himself that he has important insight, when he shares that insight, he may gain social status (Hagen, 2008).

Dennett and McKay introduced a helpful metaphor – the doxastic shear-pin – a mechanism by which a misbelief serves to protect its author (R. T. McKay & Dennett, 2009).

We extend the doxastic shear-pin to delusion formation and maintenance, describing a mechanism in terms of aberrant predictive coding. These processes are themselves adaptive, allowing exploitation of environmental contingencies and flexible responding when those contingencies change.

We discuss the biological underpinnings of a shear-pin model, conceiving of beliefs as akin to stimulus-response habits that become resistant to contradictory evidence through overtraining.

We relate our approach to the philosophical concept of epistemic innocence – delusions provide an explanation for ineffable experiences, and are the best conclusion given the available data. Finally, an adaptive function of delusions would imply that with treatment, the adaptation the delusions provide could be lost. We will discuss the relevance of these ideas for how we treat people with delusions.

II. An example: the Capgras delusion

People with the Capgras delusion report that previously familiar people have been replaced by imposters (Capgras, 1923). Although these imposters look familiar, they do not evoke the feelings of recognition that characterize familiar people (M. Coltheart, Langdon, & McKay, 2007). We hold in mind this example as we consider the potential for adaptive function of delusions.

III. The doxastic shear pin

For McKay and Dennett, delusions might be adaptive, but in a psychological (i.e. wishful thinking) not biological sense. So for McKay and Dennett, delusions are *not* adaptive misbeliefs (R. T. McKay & Dennett, 2009). In response, we suggested that delusions might be *biologically* adaptive (A. L. Mishara, Corlett, P.R., 2009).

McKay and Dennett introduced the doxastic shear-pin (R. T. McKay & Dennett, 2009). In engineering, shear pins are built into systems to disable machines in trouble so that continued functioning does not destroy them. A broken shear-pin allows continued function, albeit at an attenuated level. We argue that delusions form when the doxastic shear-pin breaks. This approach is similar to psychological description of defenses and biases (R. T. McKay & Dennett, 2009). Sometimes wrong beliefs (like overconfidence in one's abilities)

can confer an adaptive advantage, e.g. when the benefits of winning contested resources outweigh the costs of competition (D. D. Johnson & Fowler, 2011). We argue that a broken doxastic shear pin (i.e. delusion formation) allows some continued engagement with the world, rather than no action at all (A. L. Mishara, Corlett, P.R., 2009).

IV. Learning and delusions

We have previously explained delusions as erroneous causal inferences (P. R. Corlett, Honey, & Fletcher, 2007; Hemsley, 1994; Miller, 1976). We focus on prediction errors (PEs): teaching signals that mark discrepancies between expected and actual events. In the 1960s, engineers working in neural networks (Widrow, 1960) and experimentalists working in animal models (R. A. Rescorla, Wagner, A.R., 1972) similarly accounted for new learning with PEs. As contingencies change, PEs update associations between cause and effect. Updating can occur directly by changing association strength, or indirectly by allocation of attention (Pearce & Hall, 1980). Indeed, stimuli that engender PEs garner more subsequent attention (Hogarth, Dickinson, Austin, Brown, & Duka, 2008).

The brain works to minimize uncertainty. It maintains a set of predictive associations (based on past experience) that is flexible enough to adapt, yet robust enough to avoid superstitions and instabilities (Friston, 2005b, 2009). PE minimization occurs at all levels from the single neuron (Fiorillo, 2008) up through the hierarchical neuroanatomy (Friston, 2005b, 2009). Expectations based on established associations are communicated from areas with more abstract representations downwards through the hierarchy (Mesulam, 2008). PEs are either cancelled by top-down expectancy (i.e. something unexpected is ignored), or propagated and used to update associations (i.e. new learning) (Friston, 2005b, 2009). Whether a PE is discarded or incorporated depends on precision – precise (consistent) errors drive new learning, and imprecise errors are less likely to garner updates. Precision is signaled by specific slow neuromodulators dedicated to each inference (e.g., acetylcholine for perceptual inference, dopamine for motor inference). These neuromodulators are implicated in the pathophysiology of psychosis (Adams, Stephan, Brown, Frith, & Friston, 2013; Friston, 2005a).

This model of mind/brain function and dysfunction, which is committed to veracity, may seem at odds with the generation of psychotic symptoms like hallucinations and delusions (P. R. Corlett, Taylor, Wang, Fletcher, & Krystal, 2010a). How can the complex and strongly held misbeliefs that characterize psychotic illness arise from a truth-seeking system? We know from behavioral economists that humans can depart from responses that minimize punishment and maximize reward (Kahneman, 1982). Can predictive coding depart likewise? We think so. For example, Bayesian models of message-passing in crowds can recapitulate the rumors and panic that arise after a salient event, such as a major disaster (Butts, 1998). We (and others) posit that delusions may, similarly, be explained within the predictive learning framework (P. R. Corlett et al., 2010a). Aberrant PE brain signals and attention to irrelevant stimuli have been associated with delusions in patients (Corlett et al, 2007) and delusion-like beliefs in controls (R. Morris, Griffiths, Le Pelley, & Weickert, 2013) (Le Pelley, Schmidt-Hansen, Harris, Lunter, & Morris, 2010).

Learning about environmental reward contingencies allows organisms to explore and exploit their environments (A. Dickinson, Shanks, D.R., 1995). The neurobiological changes in early schizophrenia, however, disrupt this learning (Murray et al., 2008). How, then, might delusions contribute to survival?

The effects of motivation have been challenging for simple rules to minimize prediction error (Dayan & Balleine, 2002). To reconcile the effects of satiation on subsequent responding, researchers have posited multiple instrumental controllers that compete to guide action choices. These are thought to work through a balance between goal-directed and habitual learning (Daw, Niv, & Dayan, 2005) (Hitchcott, Quinn, & Taylor, 2007). Goal-directed learning involves flexible action-outcome relationships based on computations in prefrontal cortex (Daw et al., 2005). This learning is sensitive to rapid changes in outcome value (Dayan & Balleine, 2002). Habits, more rigid representations of stimulus-response relations, are encoded in striatum (Daw et al., 2005). A recent account of reinforcement learning actually suggests that the goal-directed system is associated with processing higher up the Bayesian hierarchy. This depends on bootstrapping from simpler habitual reflexes lower down (Pezzulo, Rigoli, & Friston, 2015), and as mentioned above, depends on precision of expectations (Daw et al., 2005) – control is ceded to the most precise level (Pezzulo et al., 2015).

Delusions may arise because goal-directed learning (at the highest levels of the hierarchy) is impaired (Gold et al., 2013; Gold et al., 2012; Gold, Waltz, Prentice, Morris, & Heerey, 2008). Simpler associative mechanisms (lower in the hierarchy) might drive ongoing but less flexible instrumental engagement (A. L. Mishara, Corlett, P.R., 2009).

We propose a single impairment in PE (across the hierarchy) in three stages:

1. Delusional mood. In the prodrome, attention is drawn to irrelevant stimuli: people report feeling uncomfortable and confused (Kapur, 2003a; McGhie & Chapman, 1961). This may reflect inappropriate PEs. Functional neuroimaging studies of drug-induced and endogenous early psychosis reveal PEs in frontal cortex in response to unsurprising events, and PE intensity correlates with delusion severity (P. R. Corlett et al., 2006; P. R. Corlett, Murray, et al., 2007).

   During the prodrome, the stress-mediator cortisol increases by up to 500% (Sachar, Mason, Kolmer, & Artiss, 1963). Heightened stress impairs goal-directed learning and promotes habit formation (Schwabe & Wolf, 2009).

2. Delusion formation. In response to prodromal confusion and stress, the “doxastic shear-pin” breaks. Delusions appear in an *aha-moment*, when explanatory insight occurs and flexible processing is disabled. Habitual responses are preserved and possibly even enhanced (P. R. Corlett, Krystal, et al., 2009; P. R. Corlett et al., 2010b). Cortisol falls as delusions crystalize (Sachar et al., 1963). Forming the delusion is associated with ‘insight relief’ that helps consolidate it in memory (Miller 2008; Tsuang et al. 1988). Cortisol rises once more as delusions conflict with reality (Sachar et al., 1963). As people recover and relinquish their delusions, cortisol responses normalize (Sachar et al., 1963).

3. Explaining things with the delusion. As in overtraining during instrumental learning, the delusion becomes increasingly habitual as it is used (P. R. Corlett et al., 2010b)). In the shear-pin metaphor, delusions are an adaptive product of the shear-pin break. They enable patients to stay engaged with the environment and to exploit its regularities, though the patient may be inflexible and unresponsive to corrective feedback (P. R. Corlett et al., 2010b)).

Delusion formation, associations, and psychodynamic motivation

In a psychodynamic frame, some symptoms arise from conflict between conscious and unconscious motivation. For example, someone with hidden feelings of social failure may develop paranoia, sensing that many people are so interested in him that they constantly intrude and observe him (Lyon, Kaney, & Bentall, 1994). A man with conflicted feelings about his wife may suffer the Capgras delusion (R. McKay, Langdon, & Coltheart, 2005).

Some have considered psychodynamic processes in information processing terms (Bowlby, 1980; Pally, 2005, 2007). Cognitive 2-factor explanations posit that both a perceptual dysfunction (Factor 1) and a belief evaluation deficit (Factor 2) are necessary for delusions to form. McKay and colleagues suggested that motivational processes could influence Factor 2 (R. McKay, Langdon, & Coltheart, 2007), i.e. that wishful thinking changes belief evaluation. On the other hand, factor 1 is a further possibility, as people may actually sense things differently due to motivated biases (R. T. McKay & Dennett, 2009).

In our model (and those before us (Helmholtz, 1878/1971)), the two factors are strongly inter-related (P. R. Corlett & Fletcher, 2015). Differentiating top-down (belief) from bottom up (sensation) effects is a challenge, since, in a generative system, top-down and bottom-up processes sculpt one another. Learned biases can alter perception. We see illusory stimuli that conform to our expectations rather than the sensory data incident on the retina (Pearson & Westbrook, 2015). Regina Pally also implicates top-down effects in her analysis of the relationship between PE and psychodynamics (Pally, 2005, 2007). She posits that predictions are responsible for the recapitulation of harmful early-life relationships during adulthood (Pally, 2005, 2007).

Computational psychiatry and predictive coding have engaged directly with psychoanalysis (Carhart-Harris & Friston, 2010). In reviewing data about confabulation (memory errors that fill in gaps in memory), Aikaterini Fotopoulou notes that, consistent with ideas of motivated self-deception (Hagen, 2008), many confabulations are positively valenced, i.e. aligned with the patient's wants, and involving limbic reward processing regions (Fotopoulou, 2010).

Self-deception is relevant to delusions. It entails simultaneously believing some proposition (p) and its antithesis (not-p). Subjects may be psychologically motivated to state one belief but act according to another (Harold A Sackeim & Ruben C Gur, 1979). This is relevant to the double-bookkeeping in which some delusional patients engage (see below (L.A. Sass, 1994)).

In one laboratory based self-deception task, subjects first predict an uncertain outcome, then describe the outcome when they see it. Some subjects (self-deceivers) stick with their initial

prediction even when presented a contrary outcome (as if they can't see what's right in front of them). They are more likely to engage in this deception when incentivized for correct prediction (Mijovic-Prelec & Prelec, 2010). To explain this finding, the Prelecs call on an actor-critic model, such as those proposed to explain instrumental learning with PE (Sutton, 1998). By this account, the mind is organized into multiple interacting agents, each operating on different information and each revealing different outputs to its co-agents. The actor chooses an action and the critic gives that action a score. The critic tries to learn the actor's policy and the actor tries to get the best possible score (perhaps even better than it deserves). This architecture portends self-deception – the actor tries to fool the critic (Mijovic-Prelec & Prelec, 2010). In reinforcement learning applications of actor-critic, the critic learns the environmental states and the actor learns an action policy given those states. Prediction errors update the actor and the critic. Actor and critic have been localized to different striatal sub-regions (actor – dorsal striatum, critic – ventral (O'Doherty et al., 2004)). If actor and critic are disconnected in computational models, the critic no longer trains the actor, sensitized reward responses ensue, and behavior becomes inflexible (Takahashi, Schoenbaum, & Niv, 2008). Though previously focused on cocaine addiction, this model might explain the genesis and maintenance of self-deception and delusions. However, we advise caution. The actor-critic architecture of reinforcement learning may not map neatly onto the Prelecs' model. Whether it does is an empirical matter.

Gur and Sackeim examined self-deception using galvanic skin response (GSR) to mark the progress of conditioning. GSR is a metric of salience. The skin sweats more in response to, or in anticipation of, salient events. In their examination of self-deception, Gur and Sackeim played recordings to their subjects of the subjects' own voice and others' voices. Subjects were asked if the voice was their own or another person's. People show increased GSR to their own voice. Here, self-deception is defined as saying the voice belongs to someone else, despite increased GSR signaling which would suggest that the subject actually finds it familiar. People are more likely to self-deceive in the lab if they also endorse self-deceptive statements like "I have never lied" or "I have never stolen," which are unlikely to be true (H. A. Sackeim & R. C. Gur, 1979).

Learning theorists often use GSR to assay predictive learning in human subjects. As conditioning progresses, GSR tracks conditioned cues rather than the outcomes they predict. However, GSR changes in just one trial with reassuring instructions (e.g. "*That cue is now safe*"); what was learned over multiple trials is immediately extinguished (Lovibond, 2003, 2004). Perhaps for GSR conditioning, conscious expectation of the outcome must develop to the cue (Dawson & Furedy, 1976) although there are also contradictory data (Schell, Dawson, & Marinkovic, 1991).

Consider also the Perruchet effect (Perruchet, 1985). This is the observation that while conscious predictions demonstrate a *gambler's fallacy* (treating independent pairings as non-independent – "*If I haven't had a salient event following the cue in a while, then one must be coming*"), GSR responses do not (McAndrew, Jones, McLaren, & McLaren, 2012). There is a dissociation between expressed belief and skin conductance, which is perhaps a metric of unconscious belief. This implies that there are multiple learning systems representing

conflicting beliefs. Those systems may be responsible for the self-deception examined by Gur, Sackeim, and the Prelecs.

Clearly, the number and representational nature of learning systems in people and other animals is still a topic of active inquiry (Mitchell, De Houwer, & Lovibond, 2009). Furthermore, although we point out the methodological and inferential overlap between self-deception studies and studies of conditioning and expectation, this overlap may only be superficial. Psychodynamic notions of self-deception may not align with associative learning theory (although they may (Bowlby, 1980)). Whether they do will be best settled with new data rather than argument from the extant literature.

A. Working around delusions—Our analysis centers on the role of delusions in sustaining action. However, not all delusions motivate action. Some people double-bookkeep, claiming a delusional belief and yet acting otherwise (Bleuler, 1908; L. A. Sass, 1994). Someone might claim their food is being poisoned yet continue to eat (L.A. Sass, 1994). Bortolotti and Broome appeal to biological motivation here. Patients more afflicted by negative symptoms, who lack motivation and whose prospective cognition is impaired, may be less likely to act on their delusions (L. Bortolotti, Broome, M., 2012).

On the other hand, acting *in spite of* one's delusions could be biologically adaptive (maintaining nutrition, housing, etc.). Dickinson and Balliene (1993) accounted for the impact of incentive motivation on instrumental responding in animal learning in their associative cybernetic model. These authors sought to explain how motivational states invigorate action in some situations but not others. According to their account, instrumental behavior is invigorated by a motivational system that codes the current value of an outcome (e.g. a reward) – thus allowing for the effects of changes in outcome value by satiation. This system may be instantiated in the orbitofrontal cortex (OFC) and ventral striatum (Daw et al., 2005), both structures associated with amotivation (Lebreton et al., 2009). The interaction of this motivational system with learning systems may dictate the degree to which delusions are acted upon (L. Bortolotti, Broome, M., 2012). For example, delusion severity is associated with disrupted OFC responses during Pavlovian-to-Instrumental transfer (R. W. Morris, Quail, Griffiths, Green, & Balleine, 2015). Normally, stimuli with learned Pavlovian incentive value invigorate instrumental responding (R. A. Rescorla & Solomon, 1967). However, in patients with delusions, who attribute salience inappropriately to irrelevant cues (P. R. Corlett, Murray, et al., 2007; Kapur, 2003a), this transfer seemed to occur even with irrelevant events, such that they too drove instrumental action (R. W. Morris et al., 2015). Future computational modeling work may well discern the causal models being inferred during Pavlovian to Instrumental Transfer (Cartoni, Puglisi-Allegra, & Baldassarre, 2013), and how those models may be perturbed in patients with delusional beliefs.

B. Delusion maintenance—While many delusions have upsetting content, they may nonetheless ease the overwhelming chaos of the prodrome (Kapur, 2003b). They serve to infer the best explanation for that chaos (M. Coltheart, Menzies, & Sutton, 2010). Delusions are also remarkably elastic: they expand and morph around contradictory data (Garety, Hemsley, & Wessely, 1991; Milton, Patwa, & Hafner, 1978; Simpson & Done, 2002). Of note, patients can learn about other new things (they don't have an all-encompassing

learning deficit) and can even critically evaluate others' delusions (Rokeach, 1964). However, once a delusion is formed, subsequent PEs are explicable in the context of the delusion and serve to reinforce it (P. R. Corlett, Krystal, et al., 2009; P. R. Corlett et al., 2010b). Hence the seemingly-paradoxical observation that challenging subjects' delusions can actually strengthen their conviction (Milton et al., 1978; Simpson & Done, 2002).

The illusory truth effect is relevant here – merely having considered a proposition enhances judgment of its veracity in the future (Begg, Anas, & Farinacci, 1992). Patients with delusions are particularly susceptible to this effect (Moritz, 2012). Similar effects have been observed with conditioned memory reactivation in rodents. Merely reactivating a fear-conditioned context (by reminding animals of prior fear memories) can strengthen future responding (Lee, 2008). We recently showed that reactivating fear memories in rodents and humans on ketamine enhanced subsequent memory strength (Corlett et al, 2013; Honsberger et al, 2015). Similarly, *backfire effects a*re observed in politics (Bullock, 2009) and science (McRaney, 2013), whereby data that clearly contradict a cherished belief strengthen rather than weaken it. When a belief does crucial explanatory work, contradictory data may strengthen the belief by engaging it and reconsolidating it more strongly, much like the habitization associated with instrumental overtraining (P. R. Corlett, Krystal, et al., 2009; P. R. Corlett et al., 2010b).

V. Capgras and PE

Let's return to the Capgras delusion. Cognitive neuropsychiatric (Halligan & David, 2001) explanations of Capgras (and other delusions) range from single factor (B. A. Maher, 1974), to two-factor (Max Coltheart & Davies, 2000) to interactionist (Young, 2008). The single factor account appeals to a deficit in perception of familiarity; the delusion formation process being a reasonable or sensible consequence of such an unsettling experience (B. A. Maher, 1974; B.A. Maher, 1988b). Two-factor theorists appeal to deficits in both familiarity processing and belief evaluation such that the unlikely explanation ("My loved one has been replaced") is favored over the actual explanation ("Something is wrong with my brain"). We attempt some rapprochement between prediction error, 2-factor theory, and psychodynamic motivational accounts of delusions. Under motivational explanations, people concocted imposter beliefs like Capgras (for example) to cover up new (and guilty) lack of affection while maintaining a sense of (good) self (R. McKay et al., 2007). We suggest that PE theory can also incorporate such functionality.

Recall that several features are part of the PE account: 1) through experience we learn to expect a certain set of circumstances, 2) the consequences when these expectancies are violated, which are either to discard the dissonant experience or to incorporate it by updating the set of expectancies. Absent but expected events are crucial in the phenomenology of Capgras (M. Coltheart et al., 2007) and perhaps delusions more broadly (L. Sass & Byrom, 2015). When confronted with someone who resembles a loved-one, we 1) expect to feel familiarity, and 2) generate a PE when that feeling is absent (P. R. Corlett & Fletcher, 2015; P. R. Corlett et al., 2010b). We suggest that the continued aberrant PE leads to Capgras (P. R. Corlett et al., 2010b), since sustained PEs call for a new explanatory belief (P. R. Corlett, 2015).

PE guides the Capgras sufferer toward a particular conclusion: that a replica has replaced their loved-one. Two-factor theorists argue that this explanation is so irrational that it must require a deficit in belief evaluation. They underline reports of patients who lack familiarity responses for their loved-ones, but who never endorse delusional explanations (Tranel & Damasio, 1985). It may be that the imposter belief is not particularly unlikely (R. McKay, 2012). Furthermore, PEs have been invoked to explain selection between beliefs (Waldmann, 1998) (FitzGerald, Dolan, & Friston, 2014), so we suggest that PE dysfunction could lead to deficits in both factors 1 and 2 (P. R. Corlett, Fletcher, P.C., 2014). Predictive coding theory (and its application to psychosis) does not draw as strong a distinction between perception and belief as do 2-factor theorists (P. R. Corlett, Fletcher, P.C., 2014). Top-down expectations may sculpt perception to conform with priors (as is observed with perceptual illusions – (Gregory, 1996)). Physiological motivation can also alter top-down perception – hungry subjects perceive food images in noise (Sanford, 1937). Psychological motivation can do likewise – poorer children perceive coins as physically larger than do their wealthier counterparts (Bruner & Goodman, 1947) – although these particular studies did not replicate and the New Look approach to perception fell out of favor (Erdelyi, 1974). There are top-down attentional effects from prefrontal and parietal regions onto sensory cortices that may mediate some of these effects (Firestone & Scholl, 2015) and have relevance to delusions (Dima et al., 2009; Schmidt et al., 2012). Psychodynamic defenses might likewise alter the influence of top-down priors (R. McKay et al., 2007) so that believing changes seeing through effects on attentional allocation. Furthermore, knowledge (based on prior experience) may penetrate perception more so in people with psychosis than those without (Teufel et al., 2015), adding credence to our proposal that top-down priors may be overly engaged and sculpting perception inappropriately in psychosis. But, again, the penetrability of perception by belief is still a subject of ongoing debate that has yet to be resolved empirically (Firestone & Scholl, 2015).

VI. Why this particular belief?

Delusions are often socially relevant: they are ideas about oneself in relation to others. Their content is crucially related to an individual's specific concerns (Reed, 1972). Recently, social learning has been analyzed in terms of predictive coding (Behrens, Hunt, Woolrich, & Rushworth, 2008). We form beliefs about others and make predictions about their future actions using PE (Behrens et al., 2008; King-Casas et al., 2008). Phenomena from simple associative learning, like Kamin blocking (Kamin, 1969), have also been demonstrated in social learning –learning that one worker is productive blocks the attribution of productivity to a new worker (Cramer et al., 2002). Key questions for future research include whether there is neural circuitry dedicated to social learning or whether social learning draws more extensively upon canonical predictive circuitry because social inference is difficult and computationally intensive (Heyes & Pearce, 2015). We note with interest that voltage gated calcium channels have been implicated in the genetic risk for psychotic illnesses in genome wide association studies (Jiang et al., 2015) and may contribute to prediction error signaling (Liu et al., 2014) and to social learning in rodent models (Jeon et al., 2010).

Also important are the particulars of past and present experience, including personal and cultural context. Cold war era persecutory delusions commonly involved KGB agents

(Kihlstrom, 1988). Delusions of reference have evolved from worries about radio or satellite monitoring to concerns about the internet, from The Matrix to Transcendence (Stompe, Ortwein-Swoboda, Ritter, & Schanda, 2003) (Gold et al., 2012).

VII. Delusions and Epistemic Innocence

Epistemically innocent (L. Bortolotti, 2015) beliefs are defined as those that confer adaptive advantage by increasing knowledge, and that are based on the available data. Delusions decrease uncertainty, providing a new explanatory framework for knowing. Importantly, in doing so, they allow the person with a delusion to re-engage on some level with the otherwise too chaotic world. Also, others might have access to contradictory evidence, but the authors of delusions often do not.

What of epistemic innocence and associationism? Unexplained PE or uncertainty is stress inducing. We do not like being surprised. Explaining surprise so that events can be better predicted in the future drives belief formation. If the belief is wrong, or even delusional, it still explains away the uncertainty.

Computational modeling of learning and perception allows us to test the consequences of specific changes in a model learner (Stephan & Mathys, 2014). These experiments may shed further light on the requirements for epistemic innocence in a model where fine manipulations are possible. For example, some models produce biases, e.g. spreading of erroneous rumors in a social network (Butts, 1998), tendency to ignore base rates when making probabilistic decisions (Soltani & Wang, 2010), and habit formation (FitzGerald et al., 2014).

One relevant example is the confirmation bias (Lord et al, 1979; Nickerson, 1998), in which prior beliefs bias current decision making, specifically, contradictory data are ignored if they violate a cherished hypothesis. The confirmation bias has been tied to striatal PE learning through theoretical (Grossberg, 2000) and quantitative computational models (Doll, Jacobs, Sanfey, & Frank, 2009) as well as genetics (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007; Heyser, Fienberg, Greengard, & Gold, 2000)) (Doll et al 2009). Of interest to this discussion, confirmation bias is increased in individuals with delusions (Balzan, Delfabbro, Galletly, & Woodward, 2013). However, patients with chronic schizophrenia do not show an enhanced fronto-striatal confirmation bias (Doll et al., 2014) – the relationship with delusions in particular was not examined. It is possible that confirmation biases are specific to delusion contents (encapsulated) rather than a general deficit. Woodward and colleagues showed delusion-related confirmation biases (Balzan et al., 2013). At first, it is hard to think that maintaining a belief in the face of contradiction could be adaptive. However, Boorstin (1958) has argued that confirmation bias permitted the 17th-century New England Puritans to prosper: they had no doubts and allowed no dissent, so were freed from religious argument, and more able to focus on practical matters. Their doctrine was so clear and strongly held that they had an all-encompassing explanation. As in this example, confirmation bias may save energy and allow work on more pressing tasks. Confirmation bias also protects ones' sense of self as a person with a consistent and coherent web of beliefs living in a predictable world.

VIII. Implications for treatment

In the normative approach to defining delusions that dominates most cognitive neuroscience approaches, delusions are conceived of as symptoms to be eradicated. The recovery movement has campaigned for people who experience mental health symptoms to be able to advocate for themselves, to improve their care, and to have more influence on clinicians and researchers alike. In describing her own experience of delusions, Amy Johnson has invoked WB Yeats – suggesting that clinicians and scientists ought to tread lightly in their work, as they tread on patients' delusions (A. Johnson, Davidson, L., 2013). This is perhaps the most resounding endorsement of the epistemic innocence of delusions – Amy's delusions matter to her.

The non-clinical situations in which people with radically different beliefs clash may be instructive. For example, confronting anti-vaccination believers with contrary evidence can backfire, strengthening their conviction that vaccines are harmful (Nyhan & Reifler, 2015; Nyhan, Reifler, & Ubel, 2013). We, and others, have argued that delusions are often grounded in personal experiences; "I know it sounds crazy, but I saw it with my own eyes, Doctor"). Likelihood of relinquishing beliefs on the basis of others' testimony is strongly correlated to the credibility of the source (Nyhan et al., 2013). Do the people trying to change one's mind have a vested reason to disagree? Perhaps large-scale anti-stigma educational activities in mental health would benefit from including more people with lived experience to spread the word about mental illness (Corrigan, 2012).

In summary, delusions can engender significant suffering and distress. However, in addition to the problems they can bring, delusions form through neurobiological and psychological efforts to adapt (to learn from prediction errors, to use defenses to maintain a functioning self). Delusions are epistemically innocent: new knowledge that may be flawed, but nonetheless useful, and based on the data available to the author. We have emphasized that delusions may not be so different from non-delusional beliefs, including religious ideas, political affiliations, and scientific theories. Delusions formed by adaptive reinforcement learning processes can become cherished and difficult to replace (B. A. Maher, 1974). Clinicians would do well to consider the utility of delusions both in terms of effective approaches to change and compassionate work with the people who have made these beliefs.

Acknowledgments

We are very grateful to Adina Bianchi and Jacob Leavitt who read and commented on drafts of this paper.

Funding: This work was supported by the Connecticut State Department of Mental Health and Addiction Services. P.R.C. was funded by an IMHRO/Janssen Rising Star Translational Research Award and CTSA Grant Number UL1 TR000142 from the National Center for Research Resources (NCRR) and the National Center for Advancing Translational Science (NCATS), components of the National Institutes of Health (NIH), and NIH roadmap for Medical Research. S.K.F. was supported by NIMH Grant #-5T32MH019961, 'Clinical Neuroscience Research Training in Psychiatry'. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the views of the NIH.

References

Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. Frontiers in psychiatry. 2013; 4:47.doi: 10.3389/fpsyt.2013.00047 [PubMed: 23750138]

Balzan R, Delfabbro P, Galletly C, Woodward T. Confirmation biases across the psychosis continuum: the contribution of hypersalient evidence-hypothesis matches. The British journal of clinical psychology/the British Psychological Society. 2013; 52(1):53–69. DOI: 10.1111/bjc.12000

Begg IM, Anas A, Farinacci S. Dissociation of Processes in Belief – Source Recollection, Statement Familiarity, and the Illusion of Truth. Journal of Experimental Psychology-General. 1992; 121(4): 446–458.

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF. Associative learning of social value. Nature. 2008; 456(7219):245–249. [PubMed: 19005555]

Bentall RP, Kinderman P, Kaney S. The self, attributional processes and abnormal beliefs: towards a model of persecutory delusions. Behav Res Ther. 1994; 32(3):331–341. doi: 0005-7967(94)90131-7 [pii]. [PubMed: 8192633]

Bleuler E. Die Prognose der Dementia praecox (Schizophreniegruppe). Allgemeine Zeitschrift für Psychiatrie und psychischgerichtliche Medizin. 1908; 65:436–464.

Bortolotti L. The epistemic innocence of motivated delusions. Conscious Cogn. 2015; 33:490–499. DOI: 10.1016/j.concog.2014.10.005 [PubMed: 25459652]

Bortolotti L, Broome M. Affective Dimensions of the Phenomenon of Double Bookkeeping in Delusions. Emotion Review. 2012; 4(2):187–191.

Bowlby J. Attachment and loss. 1980; 3

Bruner JS, Goodman CC. Value and need as organizing factors in perception. J Abnorm Psychol. 1947; 42(1):33–44. [PubMed: 20285707]

Bullock JG. Partisan Bias and the Bayesian Ideal in the Study of Public Opinion. Journal of Politics. 2009; 71(3):1109–1124.

Butts C. A Bayesian model of panic in belief. Computational & Mathematical Organization Theory. 1998; 4(4):373–404.

Capgras J, Reboul-Lachaux J. L'illusion des "soises" dans un delire systematise. Bulletin de Society Clinique de Medicine Mentale. 1923; 11:6–16.

Carhart-Harris RL, Friston KJ. The default-mode, ego-functions and free-energy: a neurobiological account of Freudian ideas. Brain : a journal of neurology. 2010; 133(Pt 4):1265–1283. DOI: 10.1093/brain/awq010 [PubMed: 20194141]

Cartoni E, Puglisi-Allegra S, Baldassarre G. The three principles of action: a Pavlovian-instrumental transfer hypothesis. Front Behav Neurosci. 2013; 7:153.doi: 10.3389/fnbeh.2013.00153 [PubMed: 24312025]

Coid JW, Ullrich S, Kallis C, Keers R, Barker D, Cowden F, Stamps R. The relationship between delusions and violence: findings from the East london first episode psychosis study. JAMA psychiatry. 2013; 70(5):465–471. DOI: 10.1001/jamapsychiatry.2013.12 [PubMed: 23467760]

Coltheart, M., Davies, M. Pathologies of belief. Oxford: Blackwell; 2000.

Coltheart M, Langdon R, McKay R. Schizophrenia and monothematic delusions. Schizophr Bull. 2007; 33(3):642–647. [PubMed: 17372282]

Coltheart M, Menzies P, Sutton J. Abductive inference and delusional belief. Cogn Neuropsychiatry. 2010; 15(1):261–287. doi: 917849504 [pii]10.1080/13546800903439120. [PubMed: 20017038]

Conrad, K. Die BeginnendeSchizophrenie. Stuttgart: G. Thieme; 1958.

Corlett PR. Answering some phenomenal challenges to the prediction error model of delusions. World Psychiatry. 2015; 14(2):181–183. DOI: 10.1002/wps.20211 [PubMed: 26043333]

Corlett PR, Aitken MR, Dickinson A, Shanks DR, Honey GD, Honey RA, Fletcher PC. Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. Neuron. 2004; 44(5):877–888. DOI: 10.1016/j.neuron.2004.11.022 [PubMed: 15572117]

Corlett PR, Cambridge V, Gardner JM, Piggot JS, Turner DC, Everitt JC, Fletcher PC. Ketamine effects on memory reconsolidation favor a learning model of delusions. PLoS One. 2013; 8(6):e65088.doi: 10.1371/journal.pone.0065088 [PubMed: 23776445]

Corlett PR, Fletcher PC. Delusions and prediction error: clarifying the roles of behavioural and brain responses. Cogn Neuropsychiatry. 2015; :1–11. DOI: 10.1080/13546805.2014.990625 [PubMed: 25078663]

Corlett PR, Fletcher PC. Computational Psychiatry: A Rosetta Stone linking the brain to mental illness. Lancet Psychiatry. 2014

Corlett PR, Frith CD, Fletcher PC. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. Psychopharmacology (Berl). 2009; 206(4):515–530. DOI: 10.1007/s00213-009-1561-0 [PubMed: 19475401]

Corlett PR, Honey GD, Aitken MR, Dickinson A, Shanks DR, Absalom AR, Fletcher PC. Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: linking cognition, brain activity, and psychosis. Arch Gen Psychiatry. 2006; 63(6):611–621. DOI: 10.1001/archpsyc.63.6.611 [PubMed: 16754834]

Corlett PR, Honey GD, Fletcher PC. From prediction error to psychosis: ketamine as a pharmacological model of delusions. J Psychopharmacol. 2007; 21(3):238–252. DOI: 10.1177/0269881107077716 [PubMed: 17591652]

Corlett PR, Krystal JH, Taylor JR, Fletcher PC. Why do delusions persist? Front Hum Neurosci. 2009; 3:12.doi: 10.3389/neuro.09.012.2009 [PubMed: 19636384]

Corlett PR, Murray GK, Honey GD, Aitken MR, Shanks DR, Robbins TW, Fletcher PC. Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. Brain. 2007; 130(Pt 9):2387–2400. DOI: 10.1093/brain/awm173 [PubMed: 17690132]

Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH. Toward a neurobiology of delusions. Prog Neurobiol. 2010a doi: S0301-0082(10)00119-X [pii]10.1016/j.pneurobio.2010.06.007.

Corlett PR, Taylor JR, Wang XJ, Fletcher PC, Krystal JH. Toward a neurobiology of delusions. Prog Neurobiol. 2010b; 92(3):345–369. DOI: 10.1016/j.pneurobio.2010.06.007 [PubMed: 20558235]

Corrigan PW. Research and the elimination of the stigma of mental illness. Br J Psychiatry. 2012; 201(1):7–8. DOI: 10.1192/bjp.bp.111.103382 [PubMed: 22753850]

Cramer RE, Weiss RF, William R, Reid S, Nieri L, Manning-Ryan B. Human agency and associative learning: Pavlovian principles govern social process in causal relationship detection. Q J Exp Psychol B. 2002; 55(3):241–266. [PubMed: 12188526]

Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 2005; 8(12):1704–1711. [PubMed: 16286932]

Dawson ME, Furedy JJ. The role of awareness in human differential autonomic classical conditioning: the necessary-gate hypothesis. Psychophysiology. 1976; 13(1):50–53. [PubMed: 1244630]

Dayan P, Balleine BW. Reward, motivation, and reinforcement learning. Neuron. 2002; 36(2):285–298. [PubMed: 12383782]

Dickinson A, Burke J. Within-compound associations mediate the retrospective revaluation of causality judgements. Q J Exp Psychol B. 1996; 49(1):60–80. [PubMed: 8901386]

Dickinson, A., Shanks, DR. Instrumental action and causal representation. In: Sperber, D.Premack, D., Premack, AJ., editors. Causal Cognition. Oxford: Clarendon Press; 1995. p. 5-25.

Dima D, Roiser JP, Dietrich DE, Bonnemann C, Lanfermann H, Emrich HM, Dillo W. Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modelling. Neuroimage. 2009; 46(4):1180–1186. [PubMed: 19327402]

Doll BB, Jacobs WJ, Sanfey AG, Frank MJ. Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. Brain research. 2009; 1299:74–94. DOI: 10.1016/j.brainres.2009.07.007 [PubMed: 19595993]

Doll BB, Waltz JA, Cockburn J, Brown JK, Frank MJ, Gold JM. Reduced susceptibility to confirmation bias in schizophrenia. Cognitive, affective & behavioral neuroscience. 2014; 14(2): 715–728. DOI: 10.3758/s13415-014-0250-6

Erdelyi MH. A new look at the new look: Perceptual defense and vigilance. Psychological Review. 1974; (81):1–25. [PubMed: 4812878]

Fineberg SK, Steinfeld M, Brewer JA, Corlett PR. A Computational Account of Borderline Personality Disorder: Impaired Predictive Learning about Self and Others Through Bodily Simulation. Front Psychiatry. 2014; 5:111.doi: 10.3389/fpsyt.2014.00111 [PubMed: 25221523]

Fiorillo CD. Towards a general theory of neural computation based on prediction by single neurons. PLoS ONE. 2008; 3(10):e3298. [PubMed: 18827880]

Firestone C, Scholl BJ. Cognition does not affect perception: Evaluating the evidence for ‘top-down’ effects. Behav Brain Sci. 2015; :1–77. DOI: 10.1017/S0140525X15000965

FitzGerald TH, Dolan RJ, Friston KJ. Model averaging, optimal inference, and habit formation. Frontiers in human neuroscience. 2014; 8:457.doi: 10.3389/fnhum.2014.00457 [PubMed: 25018724]

Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci. 2009; 10(1):48–58. [PubMed: 19050712]

Fotopoulou A. The affective neuropsychology of confabulation and delusion. Cognitive neuropsychiatry. 2010; 15(1):38–63. DOI: 10.1080/13546800903250949 [PubMed: 19823958]

Fotopoulou A, Conway M, Griffiths P, Birchall D, Tyrer S. Self-enhancing confabulation: revisiting the motivational hypothesis. Neurocase. 2007; 13(1):6–15. DOI: 10.1080/13554790601160566 [PubMed: 17454684]

Fotopoulou A, Conway MA, Solms M. Confabulation: motivated reality monitoring. Neuropsychologia. 2007; 45(10):2180–2190. DOI: 10.1016/j.neuropsychologia.2007.03.003 [PubMed: 17428509]

Fotopoulou A, Solms M, Turnbull O. Wishful reality distortions in confabulation: a case report. Neuropsychologia. 2004; 42(6):727–744. DOI: 10.1016/j.neuropsychologia.2003.11.008 [PubMed: 15037052]

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A. 2007; 104(41): 16311–16316. [PubMed: 17913879]

Friston K. Hallucinations and perceptual inferences. Behav Brain Sci. 2005a; 28(6):764–766.

Friston K. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci. 2005b; 360(1456):815–836. [PubMed: 15937014]

Friston K. The free-energy principle: a rough guide to the brain? Trends Cogn Sci. 2009; 13(7):293–301. [PubMed: 19559644]

Garety PA. Making Sense of Delusions. Psychiatry-Interpersonal and Biological Processes. 1992; 55(3):282–291.

Garety PA, Hemsley DR, Wessely S. Reasoning in Deluded Schizophrenic and Paranoid Patients – Biases in Performance on a Probabilistic Inference Task. Journal of Nervous and Mental Disease. 1991; 179(4):194–201. DOI: 10.1097/00005053-199104000-00003 [PubMed: 2007889]

Gold JM, Strauss GP, Waltz JA, Robinson BM, Brown JK, Frank MJ. Negative Symptoms of Schizophrenia Are Associated with Abnormal Effort-Cost Computations. Biological psychiatry. 2013; doi: 10.1016/j.biopsych.2012.12.022

Gold JM, Waltz JA, Matveeva TM, Kasanova Z, Strauss GP, Herbener ES, Frank MJ. Negative symptoms and the failure to represent the expected reward value of actions: behavioral and computational modeling evidence. Archives of General Psychiatry. 2012; 69(2):129–138. DOI: 10.1001/archgenpsychiatry.2011.1269 [PubMed: 22310503]

Gold JM, Waltz JA, Prentice KJ, Morris SE, Heerey EA. Reward processing in schizophrenia: a deficit in the representation of value. Schizophrenia Bulletin. 2008; 34(5):835–847. DOI: 10.1093/schbul/sbn068 [PubMed: 18591195]

Gottlieb JD, Cather C, Shanahan M, Creedon T, Macklin EA, Goff DC. D-cycloserine facilitation of cognitive behavioral therapy for delusions in schizophrenia. Schizophrenia research. 2011; 131(1–3):69–74. DOI: 10.1016/j.schres.2011.05.029 [PubMed: 21723096]

Gregory R. What are illusions? Perception. 1996; 25(5):503–504. [PubMed: 8865293]

Hagen, E. Non-bizarre delusions as strategic deception. In: Elton, S., O'Higgins, P., editors. Medicine and Evolution: Current Applications, Future prospect. Taylor & Francis; 2008.

Halligan PW, David AS. Cognitive neuropsychiatry: towards a scientific psychopathology. Nat Rev Neurosci. 2001; 2(3):209–215. [PubMed: 11256082]

Helmholtz, H von. The Facts of Perception. In: Kahl, R., editor. Selected Writings of Herman von Helmholtz. Weslyan University Press; 1878/1971.

Hemsley, DR. Perceptual and cognitive abnormalities as the basis for schizophrenic symptoms. In: David, AS., Cutting, JC., editors. The Neuropsychology of Schizophrenia. Hove, UK: Laurence Erlbaum Associates; 1994. p. 97-118.

Heyes C, Pearce JM. Not-so-social learning strategies. Proc Biol Sci. 2015; 282:1802.doi: 10.1098/rspb.2014.1709

Heyser CJ, Fienberg AA, Greengard P, Gold LH. DARPP-32 knockout mice exhibit impaired reversal learning in a discriminated operant task. Brain Res. 2000; 867(1–2):122–130. [PubMed: 10837805]

Hingley SM. Psychological theories of delusional thinking: in search of integration. The British journal of medical psychology. 1992; 65(Pt 4):347–356. [PubMed: 1486056]

Hitchcott PK, Quinn JJ, Taylor JR. Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. Cereb Cortex. 2007; 17(12):2820–2827. [PubMed: 17322558]

Hogarth L, Dickinson A, Austin A, Brown C, Duka T. Attention and expectation in human predictive learning: the role of uncertainty. Q J Exp Psychol (Hove). 2008; 61(11):1658–1668. DOI: 10.1080/17470210701643439 [PubMed: 18942033]

Hohwy J. Delusions, Illusions and Inference under Uncertainty. Mind & Language. 2013; 28(1):57–71. DOI: 10.1111/Mila.12008

Hohwy J, Rajan V. Delusions as Forensically Disturbing Perceptual Inferences. Neuroethics. 2012; 5(1):5–11. DOI: 10.1007/S12152-011-9124-6

Jaspers, K. General Psychopathology. Manchester University Press; 1963.

Jeon D, Kim S, Chetana M, Jo D, Ruley HE, Lin SY, Shin HS. Observational fear learning involves affective pain system and Cav1.2 Ca2+ channels in ACC. Nat Neurosci. 2010; 13(4):482–488. DOI: 10.1038/nn.2504 [PubMed: 20190743]

Jiang H, Qiao F, Li Z, Zhang Y, Cheng Y, Xu X, Yu L. Evaluating the association between CACNA1C rs1006737 and schizophrenia risk: A meta-analysis. Asia Pac Psychiatry. 2015; doi: 10.1111/appy.12173

Johnson, A., Davidson, L. Recovery to Practice: Dear Amy & Larry. 2013. Retrieved from http://www.dsgonline.com/rtp/special.feature/2013/2013_05_21/WH_2013_05_21_fullstory.html

Johnson DD, Fowler JH. The evolution of overconfidence. Nature. 2011; 477(7364):317–320. DOI: 10.1038/nature10384 [PubMed: 21921915]

Kahneman, D., Slovic, P., Tversky, A. Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press; 1982.

Kamin, L. Predictability, surprise, attention, and conditioning. In: Campbell, BA., Church, RM., editors. Punishment and Aversive Behavior. New York: Appleton-Century-Crofts; 1969.

Kapur S. Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. American Journal of Psychiatry. 2003a; 160(1):13–23. DOI: 10.1176/appi.ajp.160.1.13 [PubMed: 12505794]

Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry. 2003b; 160(1):13–23. [PubMed: 12505794]

Kihlstrom, JF., Hoyt, IP. Hypnosis and the psychology of delusions. In: Oltmanns, TF., Maher, BA., editors. Delusional Beliefs. New York: John Wiley and Sons; 1988.

Kinderman P. Attentional bias, persecutory delusions and the self-concept. Br J Med Psychol. 1994; 67(Pt 1):53–66. [PubMed: 8204542]

Kinderman P, Bentall RP. Self-discrepancies and persecutory delusions: evidence for a model of paranoid ideation. J Abnorm Psychol. 1996; 105(1):106–113. [PubMed: 8666699]

King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR. The rupture and repair of cooperation in borderline personality disorder. Science. 2008; 321(5890):806–810. DOI: 10.1126/science.1156902 [PubMed: 18687957]

Le Pelley ME, Schmidt-Hansen M, Harris NJ, Lunter CM, Morris CS. Disentangling the attentional deficit in schizophrenia: pointers from schizotypy. Psychiatry Research. 2010; 176(2–3):143–149. DOI: 10.1016/j.psychres.2009.03.027 [PubMed: 20138371]

Lebreton M, Barnes A, Miettunen J, Peltonen L, Ridler K, Veijola J, Murray GK. The brain structural disposition to social interaction. The European journal of neuroscience. 2009; 29(11):2247–2252. DOI: 10.1111/j.1460-9568.2009.06782.x [PubMed: 19490022]

Lee JL. Memory reconsolidation mediates the strengthening of memories by additional learning. Nat Neurosci. 2008; 11(11):1264–1266. [PubMed: 18849987]

Liu Y, Harding M, Pittman A, Dore J, Striessnig J, Rajadhyaksha A, Chen X. Cav1.2 and Cav1.3 L-type calcium channels regulate dopaminergic firing activity in the mouse ventral tegmental area. J Neurophysiol. 2014; 112(5):1119–1130. DOI: 10.1152/jn.00757.2013 [PubMed: 24848473]

Lovibond PF. Causal beliefs and conditioned responses: retrospective revaluation induced by experience and by instruction. J Exp Psychol Learn Mem Cogn. 2003; 29(1):97–106. [PubMed: 12549586]

Lovibond PF. Cognitive processes in extinction. Learn Mem. 2004; 11(5):495–500. doi: 11/5/495 [pii]10.1101/lm.79604. [PubMed: 15466299]

Lyon HM, Kaney S, Bentall RP. The defensive function of persecutory delusions. Evidence from attribution tasks. Br J Psychiatry. 1994; 164(5):637–646. [PubMed: 7921714]

Maher BA. Delusional thinking and perceptual disorder. J Individ Psychol. 1974; 30(1):98–113. [PubMed: 4857199]

Maher, BA. Anomalous experience and delusional thinking: The logic of explanations. In: Oltmanns, TF., Maher, BA., editors. Delusional Beliefs. New York: John Wiley and Sons; 1988a. p. 15-33.

Maher, BA. Delusions as normal theories. New York: Wiley; 1988b.

McAndrew A, Jones FW, McLaren RP, McLaren IP. Dissociating expectancy of shock and changes in skin conductance: an investigation of the Perruchet effect using an electrodermal paradigm. J Exp Psychol Anim Behav Process. 2012; 38(2):203–208. DOI: 10.1037/a0026718 [PubMed: 22250788]

McGhie A, Chapman J. Disorders of attention and perception in early schizophrenia. Br J Med Psychol. 1961; 34:103–116. [PubMed: 13773940]

McKay R. Delusional Inference. Mind & Language. 2012; 27(3):330–355. DOI: 10.1111/J.1468-0017.2012.01447.X

McKay R, Langdon R, Coltheart M. “Sleights of mind”: delusions, defences, and self-deception. Cogn Neuropsychiatry. 2005; 10(4):305–326. [PubMed: 16571464]

McKay R, Langdon R, Coltheart M. Models of misbelief: Integrating motivational and deficit theories of delusions. Conscious Cogn. 2007; 16(4):932–941. [PubMed: 17331741]

McKay RT, Dennett DC. The evolution of misbelief. The Behavioral and brain sciences. 2009; 32(6):493–510. discussion 510–461. DOI: 10.1017/S0140525X09990975 [PubMed: 20105353]

McRaney D. You Are Now Less Dumb. 2013 Gotham.

Mesulam M. Representation, inference, and transcendent encoding in neurocognitive networks of the human brain. Ann Neurol. 2008; 64(4):367–378. [PubMed: 18991346]

Mijovic-Prelec D, Prelec D. Self-deception as self-signalling: a model and experimental evidence. Philos Trans R Soc Lond B Biol Sci. 2010; 365(1538):227–240. DOI: 10.1098/rstb.2009.0218 [PubMed: 20026461]

Miller R. Schizophrenic psychology, associative learning and the role of forebrain dopamine. Med Hypotheses. 1976; 2(5):203–211. [PubMed: 9558]

Milton F, Patwa VK, Hafner RJ. Confrontation vs. belief modification in persistently deluded patients. Br J Med Psychol. 1978; 51(2):127–130. [PubMed: 646958]

Mishara AL. Klaus Conrad (1905–1961): delusional mood, psychosis, and beginning schizophrenia. Schizophrenia Bulletin. 2010; 36(1):9–13. DOI: 10.1093/schbul/sbp144 [PubMed: 19965934]

Mishara AL, Corlett PR. Are delusions biologically adaptive? Salvaging the doxastic shear pin. Behav Brain Sci. 2009; 32:530–531.

Mitchell CJ, De Houwer J, Lovibond PF. The propositional nature of human associative learning. The Behavioral and brain sciences. 2009; 32(2):183–198. discussion 198–246. DOI: 10.1017/S0140525X09000855 [PubMed: 19386174]

Moritz S, Kother U, Woodward TS, Veckenstedt Dechene A, Stahl C. Repetition is good? An internet trial on the illusory truth effect in schizophrenia and nonclinical participants. Journal of Behavior Therapy and Experimental Psychiatry. 2012 E-pub.

Morris R, Griffiths O, Le Pelley ME, Weickert TW. Attention to irrelevant cues is related to positive symptoms in schizophrenia. Schizophrenia Bulletin. 2013; 39(3):575–582. DOI: 10.1093/schbul/sbr192 [PubMed: 22267535]

Morris RW, Quail S, Griffiths KR, Green MJ, Balleine BW. Corticostriatal control of goal-directed action is impaired in schizophrenia. Biol Psychiatry. 2015; 77(2):187–195. DOI: 10.1016/j.biopsych.2014.06.005 [PubMed: 25062683]

Murray GK, Corlett PR, Clark L, Pessiglione M, Blackwell AD, Honey G, Fletcher PC. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Mol Psychiatry. 2008; 13239(3):267–276. DOI: 10.1038/sj.mp.4002058

Neale, JM. Defensive Function of Manic Episodes. In: Oltmanns, TF., Maher, B., editors. Delusional Beliefs. New York: Wiley; 1988. p. 138-156.

Nettle D, Clegg H. Schizotypy, creativity and mating success in humans. Proceedings Biological sciences/The Royal Society. 2006; 273(1586):611–615. DOI: 10.1098/rspb.2005.3349

Nielssen O, Langdon R, Large M. Folie a deux homicide and the two-factor model of delusions. Cognitive neuropsychiatry. 2013; 18(5):390–408. DOI: 10.1080/13546805.2012.718246 [PubMed: 22974316]

Nyhan B, Reifler J. Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. Vaccine. 2015; 33(3):459–464. DOI: 10.1016/j.vaccine.2014.11.017 [PubMed: 25499651]

Nyhan B, Reifler J, Ubel PA. The hazards of correcting myths about health care reform. Med Care. 2013; 51(2):127–132. DOI: 10.1097/MLR.0b013e318279486b [PubMed: 23211778]

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science. 2004; 304(5669):452–454. DOI: 10.1126/science.1094285 [PubMed: 15087550]

Padoa-Schioppa C, Schoenbaum G. Dialogue on economic choice, learning theory, and neuronal representations. Current Opinion in Behavioral Sciences. 2015 in press.

Pally R. Non-conscious prediction and a role for consciousness in correcting prediction errors. Cortex. 2005; 41(5):643–662. discussion 731–644. [PubMed: 16209328]

Pally R. The predicting brain: unconscious repetition, conscious reflection and therapeutic change. Int J Psychoanal. 2007; 88(Pt 4):861–881. [PubMed: 17681897]

Pearce JM, Hall G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev. 1980; 87(6):532–552. [PubMed: 7443916]

Pearson J, Westbrook F. Phantom perception: voluntary and involuntary nonretinal vision. Trends Cogn Sci. 2015; 19(5):278–284. DOI: 10.1016/j.tics.2015.03.004 [PubMed: 25863415]

Perruchet P. A pitfall for the expectancy theory of human eyelid conditioning. Pavlov J Biol Sci. 1985; 20(4):163–170. [PubMed: 4069791]

Pezzulo G, Rigoli F, Friston K. Active Inference, homeostatic regulation and adaptive behavioural control. Prog Neurobiol. 2015; doi: 10.1016/j.pneurobio.2015.09.001

Reed, GF. The psychology of anomalous experience: a cognitive approach. London: Hutchinson; 1972.

Rescorla RA, Solomon RL. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychol Rev. 1967; 74(3):151–182. [PubMed: 5342881]

Rescorla, RA., Wagner, AR. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In: Black, AH., Prokasy, WF., editors. Classical conditioning II: Current research and theory. New York: Appleton-Century-Crofts; 1972. p. 64-99.

Roberts G. Delusional belief systems and meaning in life: a preferred reality? The British journal of psychiatry Supplement. 1991; (14):19–28. [PubMed: 1840775]

Rokeach, M. The three Christs of Ypsilanti. New York: Alfred Knopf; 1964.

Sachar EJ, Mason JW, Kolmer HS Jr, Artiss KL. Psychoendocrine Aspects of Acute Schizophrenic Reactions. Psychosomatic medicine. 1963; 25:510–537. [PubMed: 14080091]

Sackeim HA, Gur RC. Self-deception, other-deception, and self-reported psychopathology. J Consult Clin Psychol. 1979; 47(1):213–215. [PubMed: 429664]

Sackeim HA, Gur RC. Self-deception, other-deception, and self-reported psychopathology. Journal of Consulting and Clinical Psychology. 1979; 47(1):213. [PubMed: 429664]

Sackeim HA, Gur RC. Voice recognition and the ontological status of self-deception. J Pers Soc Psychol. 1985; 48(5):1365–1372. [PubMed: 3998994]

Sanford RN. The Effects of Abstinence from Food upon Imaginal Processes: A Further Experiment. The Journal of Psychology: Interdisciplinary and Applied. 1937; 3(1)

Sass L, Byrom G. Phenomenological and neurocognitive perspectives on delusions: A critical overview. World Psychiatry. 2015; 14(2):164–173. DOI: 10.1002/wps.20205 [PubMed: 26043327]

Sass LA. Civilized Madness – Schizophrenia, Self-Consciousness and the Modern Mind. History of the Human Sciences. 1994; 7(2):83–120. DOI: 10.1177/095269519400700206

Sass, LA. Paradoxes of delusion: Wittgenstein, Schreber, and the schizophrenic mind. Ithaca: Cornell University Press; 1994.

Sass LA. Some Reflections on the (Analytic) Philosophical Approach to Delusion. Philosophy, Psychiatry & Psychology. 2004; 11(1):71–80.

Schell AM, Dawson ME, Marinkovic K. Effects of potentially phobic conditioned stimuli on retention, reconditioning, and extinction of the conditioned skin conductance response. Psychophysiology. 1991; 28(2):140–153. [PubMed: 1946880]

Schmidt A, Bachmann R, Kometer M, Csomor PA, Stephan KE, Seifritz E, Vollenweider FX. Mismatch negativity encoding of prediction errors predicts S-ketamine-induced cognitive impairments. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology. 2012; 37(4):865–875. DOI: 10.1038/npp.2011.261 [PubMed: 22030715]

Schwabe L, Wolf OT. Stress prompts habit behavior in humans. The Journal of neuroscience : the official journal of the Society for Neuroscience. 2009; 29(22):7191–7198. DOI: 10.1523/JNEUROSCI.0979-09.2009 [PubMed: 19494141]

Simpson J, Done DJ. Elasticity and confabulation in schizophrenic delusions. Psychol Med. 2002; 32(3):451–458. [PubMed: 11989990]

Soltani A, Wang XJ. Synaptic computation underlying probabilistic inference. Nature neuroscience. 2010; 13(1):112–119. DOI: 10.1038/nn.2450 [PubMed: 20010823]

Stephan KE, Mathys C. Computational approaches to psychiatry. Current opinion in neurobiology. 2014; 25:85–92. DOI: 10.1016/j.conb.2013.12.007 [PubMed: 24709605]

Stewart J, Varela FJ, Coutinho A. The relationship between connectivity and tolerance as revealed by computer simulation of the immune network: some lessons for an understanding of autoimmunity. Journal of autoimmunity. 1989; 2(Suppl):15–23.

Stompe T, Ortwein-Swoboda G, Ritter K, Schanda H. Old wine in new bottles? Stability and plasticity of the contents of schizophrenic delusions. Psychopathology. 2003; 36(1):6–12. doi: 69658. [PubMed: 12679586]

Sutton, RS., Barto, AG. Reinforcement Learning: An Introduction. MIT Press; 1998.

Tabor, JD., Galagher, EV. Why Waco?. University of California Press; 1997.

Takahashi Y, Schoenbaum G, Niv Y. Silencing the critics: understanding the effects of cocaine sensitization on dorsolateral and ventral striatum in the context of an actor/critic model. Frontiers in neuroscience. 2008; 2(1):86–99. DOI: 10.3389/neuro.01.014.2008 [PubMed: 18982111]

Teufel C, Subramaniam N, Dobler V, Perez J, Finnemann J, Mehta PR, Fletcher PC. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. Proc Natl Acad Sci U S A. 2015; 112(43):13401–13406. DOI: 10.1073/pnas.1503916112 [PubMed: 26460044]

Tranel D, Damasio AR. Knowledge without awareness: an autonomic index of facial recognition by prosopagnosics. Science. 1985; 228(4706):1453–1454. [PubMed: 4012303]

Van Hamme LJ, Wasserman EA. Cue competition in causality judgments: The role of nonrepresentation of compound stimulus elements. Learning and Motivation. 1994; 25:127–151.

Waldmann, MR., Martignon, L. A Bayesian network model of causal learning. In: Gernsbacher, MA., Derry, SJ., editors. Proceedings of the Twentieth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Earlbaum; 1998. p. 1102-1107.

Widrow B, Hoff ME Jr. Adaptive switching circuits. IRE WESCON convention rec. 1960

Yon D, Press C. Back to the future: synesthesia could be due to assocuiative learning. Frontiers in psychology. 2014

Young G. Capgras delusion: an interactionist model. Conscious Cogn. 2008; 17(3):863–876. [PubMed: 18314350]