damage. (We do not comment on the interesting debate whether delusions are beliefs. For one account, see Mishara, in press a).

M&D argue that brain dysfunction underlying delusions involves "breakage" in the belief evaluation system which is "adventitious, not designed" (sect. 10, para. 4). They advocate a two-factor model: (1) a perceptual insult which engenders odd experiences; (2) a deficit in belief evaluation which enables the entertainment and maintenance of bizarre and unlikely explanations for the experience (Davies & Coltheart 2000). Distinguishing deficit (organic-neural) versus motivational (psychological/psychodynamic-defense) approaches to delusions (McKay et al. 2007a), M&D conclude that "motivation" plays a psychological but not a biological role in the two-factor model. In contrast, we propose a single impairment in prediction-error–driven (i.e., motivational) learning in three stages: (1) delusional mood; (2) delusion as Aha-Erlebnis; (3) reconsolidation. Our model indicates how delusions may be adaptive as a shear pin function by enabling the patient to remain in *vital* connection with his/her environment:

1. Prior to delusions, a prodromal *delusional mood* may last for days, months, or even years (Conrad 1958; Jaspers 1946/1963). The patient experiences increasingly oppressive tension, *a feeling of non-finality* or expectation. Conrad (1958) calls this *Trema* (stage-fright) as the patient has the feeling that something very important is about to happen. Attention is drawn toward irrelevant stimuli, thoughts, and associative connections which are distressing and unpredictable (Kapur 2003; McGhie & Chapman 1961; Uhlhaas & Mishara 2007). This reflects an impairment in the brain's predictive learning mechanisms, such that unexpected events, prediction errors, are registered inappropriately (Corlett et al. 2007).

2. The delusions appear as an *Aha-Erlebnis*, or "revelation" (Conrad 1958), concerning what had been perplexing during delusional mood. In delusions of reference, harmless or accidental occurrences in the environment are taken as referring to the self. Conrad (1958) calls this a reflexive turning back on the self in which the universe is experienced as "revolving" around the self as middle-point. The delusions are not primarily a defensive reaction to protect the self, but involve a "reorganization" of the patient's experience to maintain behavioral interaction with the environment despite the underlying disruption to perceptual binding processes (Conrad 1958; Mishara 2010). At the Aha-moment, the "shear pin" breaks, or as Conrad puts it, the patient is unable to shift "reference-frame" to consider the experience from another perspective. The delusion disables flexible, controlled conscious processing from continuing to monitor the mounting distress of the wanton prediction error during delusional mood and thus deters cascading toxicity. At the same time, automatic habitual responses are preserved, possibly even enhanced (Corlett et al. 2009b).

3. *Reconsolidation*. Forming the delusion is associated with insight relief which stamps the delusion into memory (Miller 2008; Tsuang et al. 1988). Each time delusions are deployed, they are reinforced further, through a process of recall, reactivation, and reconsolidation, which strengthens them, conferring resistance to contradiction rather like the formation of motor-habits with overtraining (Adams & Dickinson 1981). When subsequent prediction errors occur, they are explicable in terms of the delusion and serve to reinforce it (Corlett et al. 2009b; Eisenhardt & Menzel 2007). Hence the paradoxical observation that challenging subjects delusions can actually strengthen their conviction (Milton et al. 1978). In each rehearsal of the delusion in the present instance, there is a "monotonous" spreading of the delusion to new experience (Binswanger 1965; Conrad 1958; Mishara, in press b) and, as such, it is both fixed and elastic (Corlett et al. 2009b). For example, we interviewed a middle-aged schizophrenia patient with the intractable erotomanic delusion that a college acquaintance had fallen in love with her and now controls parts of her life. Whenever she thinks of him, she hears a "car beep" or "trips while walking," i.e., signals intended to inform her that he knows she is thinking about him.

Neurobiologically, this reconsolidation-based-strengthening shifts control of behavior toward the striatal habit system. However, the ceding of behavior from effortful, conscious control is associated with a "mechanization" of experience. Schizophrenia patients delusionally refer to themselves in inhuman terms, for example as "machine," "computer," or "registering apparatus" (Binswanger 1965; Kraus 1997; Mishara 2007a), as if the delusion reflects its own disabling function of flexible conscious processing. Losing the experience as consistent intentional agent (Wegner 2004), the patients nevertheless continue to respond reflexively to the environmental cues incumbent upon them, necessary for continued survival. As complement to such delusions of alien control, however, the healthy individual has the converse "everyday delusion": She thinks that it is "I" who moves her own limbs. She calls the movement *mine* although it has its own momentum, automaticity, and finds its own way. That is, the healthy individual "overlooks" the impersonal-mechanical side of her movements in a "counter-delusion" to the patient who is unable to access the personal contribution (von Weizsäcker 1956). We are no more free from the necessity of "delusions" in our everyday functioning and its intermittent ceding to automatic processes than is the patient with schizophrenia.

Finally, the authors outline Bayesian mechanisms of rational belief formation. We propose that delusions form via the same Bayesian learning mechanisms but we challenge the strict separation between perception and belief upon which two-factor accounts are predicated (Corlett et al. 2009a; Fletcher & Frith 2009; Hemsley & Garety 1986; Uhlhaas & Mishara 2007). In our account, delusions also depend on aberrations of perception which occur when neuronal noise induces mismatches between expectancy (Bayesian priors) and experience (sensory inputs/evidence), but in terms of the single factor, prediction error.

ACKNOWLEDGMENTS

Both authors are recipients of the NARSAD Young Investigator Award. Phil Corlett is supported by the University of Cambridge Parke-Davis Exchange Fellowship in Biomedical Sciences.

The evolution of religious misbelief

doi:10.1017/S0140525X09991312

Ara Norenzayan, Azim F. Shariff, and Will M. Gervais

Department of Psychology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

ara@psych.ubc.ca **www.psych.ubc.ca/~ara**
azim@psych.ubc.ca **www.psych.ubc.ca/~azim**
will.gervais@gmail.com

Abstract: Inducing religious thoughts increases prosocial behavior among strangers in anonymous contexts. These effects can be explained both by behavioral priming processes as well as by reputational mechanisms. We examine whether belief in moralizing supernatural agents supplies a case for what McKay & Dennett (M&D) call evolved misbelief, concluding that they might be more persuasively seen as an example of *culturally* evolved misbelief.

Is belief in supernatural agency an example of evolved "misbelief"? McKay & Dennett (M&D) consider recent psychological experiments that have investigated whether religious beliefs cause prosocial behavior such as generosity and honesty (for reviews, see Norenzayan & Shariff 2008; Shariff et al. 2010). In M&D's philosophical analysis, whether or not religion supplies a case of evolved misbelief turns out to depend on the psychological mechanism that best accounts for these effects. We therefore revisit the experimental evidence and discuss in some depth the ideomotor and supernatural watcher accounts for these effects.

M&D cite Randolph-Seng and Nielsen (2008), who critiqued Shariff and Norenzayan (2007), questioning the plausibility of the supernatural watcher hypothesis because the data could not conclusively distinguish between the ideomotor and supernatural watcher explanations. These two mechanisms gain plausibility given two distinct but well-supported empirical literatures. There is considerable evidence showing that prosocial behavior can be facilitated both by activating nonconscious altruistic thoughts (e.g., Bargh et al. 2001), and by heightened reputational concerns (e.g., Fehr & Fischbacher 2003). These two mechanisms are not mutually exclusive, however, and may even reinforce each other in everyday life.

The interesting question therefore is: What kind of laboratory evidence can provide support for the supernatural watcher account above and beyond behavioral-priming processes? First, if the priming effects of God concepts are weaker or nonexistent for non-believers, then the effect could not be solely due to ideomotor processes, which are typically impervious to prior explicit beliefs or attitudes. Second, if God primes make religious participants attribute actions to an external source of agency, these effects could not be explained by ideomotor processes, as such manipulations disambiguate the felt presence of supernatural watchers from their alleged prosocial consequences. Finally, if the supernatural watcher explanation is at play, religious primes should arouse social evaluation of the self. Moreover, such reputational awareness should moderate the magnitude of the prime's effect on prosocial behavior.

As M&D note, evidence on the first point is currently mixed. However, close examination of the findings betrays a revealing pattern. All but one of these priming studies recruited student samples, which can be problematic since beliefs, attitudes, and social identity among students can be unstable, raising questions about the reliability of chronic individual difference measures of religious belief and identity measures for students who are still in transition to adulthood (Sears 1986; Henrich et al., in press). Thus, student atheists might be at best "soft atheists." In the only religious priming experiment we are aware of that recruited a non-student adult sample (Shariff & Norenzayan 2007, Study 2), the effect of the prime emerged again for theists, but disappeared for these "hard" atheists (see Fig. 1). In addition, Henrich et al. (2009) found that across 14 small-scale societies of varying group size, where there is variability in whether supernatural agents are morally concerned, belief in the moralizing Abrahamic God (along with degree of market integration) predicted larger offers in the dictator and ultimatum games. These initial findings speak against an exclusively ideomotor account of the results, and suggest that belief – not just alief – is involved in religious prosociality.

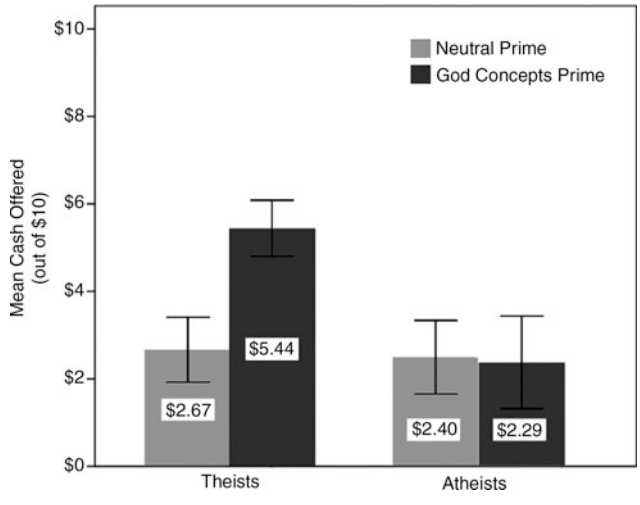

Figure 1 (Norenzayan et al.). Results from the dictator game in Shariff and Norenzayan (2007, Study 2) indicate that priming God concepts increased generosity for religious believers but not for atheists. Error bars represent standard error of the mean.

Regarding the second question, one experiment clearly separates the felt presence of a supernatural agent from prosocial outcomes. Dijksterhuis et al. (2008) found that after being subliminally primed with the word "God," believers (but not atheists) were more likely to ascribe an outcome to an external source of agency, rather than their own actions. In addition, religious belief positively correlates with greater concern with social evaluation of the self (Trimble 1997), and recent experimental evidence points to this being a causal relationship. Gervais and Norenzayan (2009) found that priming God concepts (using the same sentence unscrambling task of Shariff and Norenzayan [2007]) increased public self-awareness (Govern & Marsch 2001) – a measure that taps into feelings of being the target of social evaluation. In contrast, and as predicted, the prime had no effect on private self-awareness. Ongoing research is examining whether prosocial effects of religious primes are moderated by measures of evaluative concern, a key prediction of the supernatural watcher hypothesis, which would be incompatible with a purely ideomotor account. Thus, although M&D are right that more research is needed to reach firm conclusions, the evidence regarding the supernatural watcher hypothesis is more compelling than M&D's cautious approach suggests. But does that mean that belief in supernatural agents is an example of adaptive misbelief?

M&D briefly mention both by-product theories of religion and cultural evolutionary explanations for cooperation. We have argued elsewhere (Norenzayan & Shariff 2008; Norenzayan, in press; Shariff et al. 2010) that integrating these two frameworks yields a more cogent explanation for the rise and persistence of religious beliefs than theories which invoke a more direct genetic evolutionary argument (e.g., Bering et al. 2005; Johnson & Bering 2006). Once belief in supernatural agency emerged as a by-product of mundane cognitive processes, cultural evolution favored the spread of a special type of supernatural agent – moralizing high Gods. Growing evidence is converging on the conclusion that sincere belief in these omniscient supernatural watchers facilitated cooperation and trust among strangers (Norenzayan & Shariff 2008). Not surprisingly, this cultural spread coincided with the expansion of human cooperation into ever larger groups over the last 15 millenia (Cauvin 2000). This evolutionary scenario has the virtue of explaining an otherwise puzzling feature of religious prosociality – namely, the systematic cultural variability in the prevalence of moralizing Gods across societies that correlates with group size (e.g., Roes & Raymond 2003). Contrary to a genetic adaptation account, the deities of most small-scale societies, which more closely approximate ancestral conditions, are neither fully omniscient nor morally concerned. It is the evolutionarily recent anonymous social groups, facing the breakdown of reputational and kin selection mechanisms for cooperation, which most strongly espouse belief in such Gods. Thus, beliefs in moralizing supernatural agents may not qualify as genetically evolved misbeliefs. But they could instead be seen as examples of culturally evolved ones that played a key historical (although not irreplaceable) role in the rise and stability of large cooperative communities.

The (mis)management of agency: Conscious belief and nonconscious self-control

doi:10.1017/S0140525X09991336

Brandon Randolph-Seng

Texas Tech University, Rawls College of Business, Area of Management, Lubbock, TX 79409-2101.

b.randolph-seng@ttu.edu www.webpages.ttu.edu/brandolp

Abstract: McKay & Dennett (M&D) identify positive illusions as fulfilling the criteria for an adaptive misbelief, but could there be other