experience lets them down (e.g., Baker et al. 2009; Massey & Gelman 1988).

Last, it is necessary that a proposed adaptive misbelief should convey adaptive advantage to the individual. Overconfidence in one's theories conveys adaptive advantage insofar as it enables them to creatively simplify a problem by ignoring some of the complicating factors. "[I]t seems possible for the child to experience surprise and question his theory only if the prediction he makes emanates from an already powerful theory expressed in action" (Karmiloff-Smith & Inhelder 1974, p. 209). Thus, Phase 2 enables children to unify representations into coherent (but overgeneralised) theories that in turn lead to new, broader theories and greater behavioural mastery. Second, overconfidence in one's theories sustains and enhances health in an everyday sense by decreasing exposure to cognitive dissonance, which has been shown to lead to feelings of anxiety and stress (Aronson 1969), which in turn result in negative physiological effects. Consequently, overconfidence in one's theories may also result in exaggerated feeling of control, a positive illusion that M&D list as adaptive in its own right.

Thus, overconfidence in the veracity and generalisability of one's theory fits the criteria laid out by M&D as necessary to be considered as an adaptive misbelief. Children certainly *believe* that they are right; this belief is systematic and misinforms the organism as a whole, occurs for all children across a range of microdomains, and persists into adulthood. Therefore, it can be considered a naturally occurring feature of a properly functioning doxastic system. It can also be construed as adaptive in leading the individual to undertake adaptive actions and by enhancing health and fitness. In children, this tendency is evident not only in subjective self-evaluation, but also in objective theories about how the world works that, in turn, guide their behaviour. A phase in which this is especially prominent occurs across a variety of microdomains and may be a fundamental and important feature of properly functioning theory-building doxastic systems.

Error management theory and the evolution of misbeliefs

Martie G. Haselton[a] and David M. Buss[b]

[a]*Departments of Communication and Psychology, University of California, Los Angeles, Los Angeles, CA 90095;* [b]*Department of Psychology, University of Texas, Austin, TX 78712.*
haselton@ucla.edu   dbuss@psy.utexas.edu
http://www.sscnet.ucla.edu/comm/haselton/home.html
http://www.davidbuss.com

Abstract: We argue that many evolved biases produced through selective forces described by error management theory are likely to entail misbeliefs. We illustrate our argument with the male sexual overperception bias. A misbelief could create motivational impetus for courtship, overcome the inhibiting effects of anxiety about rejection, and in some cases transform an initially sexually uninterested woman into an interested one.

McKay & Dennett (M&D) provide a useful analysis of the evolution of misbelief, making a number of important distinctions, including one between misbeliefs that are tolerable byproducts of evolved psychological adaptations and those that would have been adaptive in and of themselves. A reasonable primary hypothesis is that selection has shaped the human mind to form true beliefs about the world. The ultimate criterion of evolutionary selection, as M&D rightly point out, is reproductive success, not the accurate detection or preservation of truth. We, and others, have argued that selection has favored

psychological adaptations that do not always maximize truthful beliefs; these adaptations instead can result in *misbeliefs* (e.g., Haselton & Buss 2000; Haselton & Nettle 2006).

Humans appear to possess cognitive biases which lead to systematic misbeliefs and require scientific explanation. These include the positive illusions that compel us to have a rosy outlook on the future (Taylor & Brown 1988), sex-linked biases such as men's tendency to overestimate women's sexual interest (e.g., Abbey 1982), and perceptual biases such as auditory looming, the tendency to overestimate the proximity to self of approaching objects compared to receding objects that are in fact equally distant (Neuhoff 2001). We articulated error management theory (EMT; Haselton & Buss 2000) as a theory to explain how evolution could lead to adaptive biases, some of which entail misbeliefs. Many problems of judgment under conditions of uncertainty can be framed as having two possible errors – false positive and false negative errors. According to EMT, in forming judgments under uncertainty, if there were recurrent asymmetries in the costs of these errors over evolutionary history, selection should produce a system that errs in the less costly direction. For example, for men estimating a woman's sexual interest, we hypothesized that the reproductively more costly error would have been to underestimate her interest and miss a reproductive opportunity. Thus, EMT predicts that men possess an adaptive bias toward overestimating women's sexual interest.

M&D affirm the logic of EMT, but argue that selection can solve adaptive problems of the sort explained by EMT in ways other than creating misbeliefs. They argue that humans do not need to possess *biased beliefs* if *biased actions* can accomplish the same ends while preserving true beliefs. We agree entirely with this point. It is possible, for example, that selection could design an adaptation in which men *acted* as if a larger number of women were sexually interested in them than actually were, in order for them not to miss a potential sexual opportunity, while not truly believing that those women are sexually interested. Similarly, it might be possible for selection to fashion an adaptation in which people act as though more people harbor homicidal intent than they actually do, in order to avoid the costly cases in which people actually do harbor such thoughts, without actually believing that those individual do harbor homicidal intent.

Just because selection *can* solve these adaptive problems without misbelief does not mean that selection *has* solved these problems without misbelief. The argument that selection could craft an adaptation for thermoregulation other than sweat glands (e.g., dogs thermoregulate through evaporation from a protruding tongue) is not an argument that selection has not fashioned sweat glands in humans.

Ultimately, the question of whether misbeliefs are part of the design of EMT biases is an open issue that must be decided on a case-by-case basis with empirical research. However, we suggest that there are no compelling reasons to discount the possibility that misbeliefs, including functional misbeliefs, are part of the evolved design of EMT biases. Consider the male sexual overperception bias. A misbelief that a woman is sexually interested could facilitate access to sexual opportunities in at least three ways. First, it could provide the motivational impetus for courtship behavior. Second, it could allay a man's anxiety about being rejected, eliminating a common cognitive barrier to initiating courtship (Kugeares 2002). If it turns out that his belief was indeed incorrect, it is not terribly costly for him to revise his beliefs about a particular woman after being rebuffed (e.g., "I thought she was sending me sexual signals, but it turns out I was wrong"). Third, a man's misbelief, by motivating attraction tactics or elevating confidence, could transform a woman who is initially sexually uninterested in him into one who is sexually interested – an outcome showing that the initial misbelief itself can sometimes provide functional benefits. Hence, the EMT-generated misbelief can, in principle, solve the adaptive

problem of maximizing sexual opportunities more effectively than an adaptation lacking the misbelief design feature.

Although we advanced the theory to explain cognitive biases, the core logic of EMT is neutral in predicting where in the perception-belief-action chain selection will shape a bias. All that is required is that, ultimately, humans behave so that they minimize the more costly of the two errors in question, even if this cost minimization ends up producing a larger number of overall errors. To discover where in this chain a bias exists must be empirically adjudicated. On the basis of the existing empirical evidence, however, we suggest that biasing action is unlikely to be the sole outcome of selection in which there has been recurrent cost asymmetries associated with errors.

M&D's analysis will stimulate empirical research about particular EMT biases. Some biases may be instances of biased action without involving misbelief. Others may entail misbeliefs. A subset of these may be cases in which the misbelief is not simply a tolerable byproduct of an adaptively biased cognitive system but is itself adaptive. M&D make a compelling argument that positive illusions qualify as adaptive misbeliefs because they positively affect an individual's fitness by motivating striving for favorable outcomes. We suggest that that some EMT biases, such as the male sexual overperception bias, also can motivate adaptive action through misbeliefs by providing motivational impetus for action, overcoming inhibitions associated with action, and transforming the psychological states of others in ways beneficial to the holder of misbeliefs.

God would be a costly accident: Supernatural beliefs as adaptive

Dominic D. P. Johnson

Department of Politics and International Relations, University of Edinburgh, Edinburgh, EH8 9LD Scotland, United Kingdom.
dominic.johnson@ed.ac.uk   http://dominicdpjohnson.com/

Abstract: I take up the challenge of why *false* beliefs are better than "cautious action *policies*" (target article, sect. 9) in navigating adaptive problems with asymmetric errors. I then suggest that there are *interactions* between supernatural beliefs, self-deception, and positive illusions, rendering elements of all such misbeliefs adaptive. Finally, I argue that supernatural beliefs cannot be rejected as adaptive simply because recent experiments are inconclusive. The great costs of religion betray its even greater adaptive benefits – we just have not yet nailed down exactly what they are.

The greatest challenge to McKay & Dennett's (M&D's) argument is why *false* beliefs are necessary to achieve adaptive behavior – why not (as M&D note in sect. 9, para. 2) just have "cautious action *policies*" instead? I don't believe this problem was completely resolved in the target article, so I tackle it with reference to the "supernatural punishment hypothesis" (Johnson 2009; Johnson & Bering 2006; Johnson & Krüger 2004), since the same problem haunts that hypothesis as well.

The argument is that the costs of selfishness increased when humans evolved language and Theory of Mind (ToM), because social transgressions became much more likely to be detected and punished. Supernatural punishment offered a cautionary mind-guard to reduce selfishness and avoid real-world costs. But why bring God into it? A Darwinian perspective suggests that atheists could simply develop a "cautious action policy" – becoming more prudent about when to be selfish. A first line of defense comes from M&D's categories of evolutionary limitations: (1) economics – a fear of supernatural agency may have been biologically cheaper or more efficient; (2) history – a capacity for supernatural beliefs may have been more readily

available, given the prior evolution of ToM; (3) adaptive landscape – fear of detection and punishment by *supernatural* agents may have been a small step up the local fitness peak from fear of detection and punishment by *human* agents.

A stronger line of defense is that, while a cautious action policy might work in principle, the whole point of error management theory is that it pays to *overestimate* the probability of detection, not to get it right or to weigh up the costs and benefits "rationally" (Haselton & Buss 2000; Haselton & Nettle 2006; Nettle 2004). Believing (irrationally) that *supernatural* agents are watching is a good way to ensure systematic overestimation of the *actual* risk of detection and punishment (by other human beings; Johnson 2009). The power of religion appears to stem precisely from its irrational and non-falsifiable features (Rappaport 1999), and empirical data suggest that religious beliefs are more effective at promoting group survival than similar but *non*-religious beliefs (Sosis & Bressler 2003). Cautious action policies might work in reducing selfishness, but they may not be as effective as God.

My next concern is that supernatural agency, self-deception, and positive illusions are treated as independent phenomena, with only positive illusions making the cut for an adaptive misbelief. However, there are important *interactions* between these three phenomena that make elements of all of them adaptive.

First, self-deception is essential to many supernatural beliefs. If supernatural punishment is to affect people's behavior, they must believe in it – despite lacking any direct evidence whatsoever and despite having to ignore counter-evidence. This is classic self-deception (Trivers 2000). Interestingly, this self-deception can be reinforced by the belief itself – in many religions, it is common for someone's misfortune to be treated as *evidence* of wrongdoing, since gods or spirits "evidently" punished the victim (Bering & Johnson 2005).

Second, self-deception is essential to many positive illusions. For example, positive illusions have been suggested to be adaptive in conflict, bluffing superior power or skill to deter opponents (Johnson 2004; Trivers 2000; Wrangham 1999). Self-deception is essential here to avoid "behavioural leakage" that would otherwise give the game away (nervous Nellies are less convincing bluffers than cool-hand Lukes). This may be why, as Daniel Kahneman notes, "all the biases in judgment that have been identified in the last 15 years tend to bias decision-making toward the hawkish side" (quoted in Shea 2004). Positive illusions appear to be advantageous enough that numerous psychological biases converge to promote them despite the evidence.

Third, supernatural beliefs may be an *example* of positive illusions. As M&D note, people often cite God as giving them "the strength to go on." If health or fitness advantages derive from such beliefs, then religious beliefs are adaptive according to M&D's own criteria. Religious beliefs may involve all three types of positive illusions: positive self-evaluations (God chose me/us), illusions of control (God will help me/us in difficult times), and optimism about the future (God has a plan; Heaven awaits). Similar beliefs are common among the world's numerous religions.

My final concern is M&D's rejection of supernatural beliefs as adaptive, which hinges on a perceived lack of empirical evidence. This is problematic for three reasons. First, in the literature M&D focus on, researchers tend to use religious primes derived from Western Judeo-Christian traditions (e.g., "divine," "God," and "prophet" in Shariff & Norenzayan 2007), whereas the relevant supernatural concepts in our evolutionary history could be anything from dead ancestors, spirits, ghosts, witches, inanimate objects, and so forth. Similarly, modern *religious* agents are only one possible type of supernatural agency, whereas subjects' behavior may also be influenced by other *sources* such as superstition, folklore, karma, Just World beliefs (the belief that victims of tragedy somehow deserved it), or everyday "comeuppence" and "just deserts." Given this diversity of possible supernatural agents and sources, personal religious