God would be a costly accident: Supernatural beliefs as adaptive

God would be a costly accident: Supernatural beliefs as adaptive

doi:10.1017/S0140525X09991245

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

affiliations and devoutness among experimental subjects may be somewhat independent of how supernatural beliefs – in general – influence people's behavior (M&D predict an interaction of personal religious devoutness and behavior). Current experiments may not, therefore, be able to differentiate the behavior of "believers" and "non-believers" – Joe Bloggs may be an avowed atheist who, on his way to Las Vegas, is nevertheless very concerned about seeing a black cat or wearing his lucky jacket or what his grandmother would have said.

Second, even if we had incontrovertible evidence that supernatural cues (e.g., via experimental primes) promoted higher donations in economic games, this is far from evidence that religious beliefs are biologically *adaptive.* On the contrary, it could be evidence that religious primes turn people into suckers who give away precious resources. Such behavior, on its own, would not survive natural selection – without additional field experiments measuring fitness consequences, evidence for altruism is hardly evidence of an adaptive trait. Therefore, the (excellent) current laboratory experiments that M&D focus on cannot yet be used as deal-breakers as to whether (mis)belief is adaptive or not.

Third, having rejected supernatural beliefs as adaptive, M&D's null hypothesis is that religious beliefs are a non-adaptive byproduct of cognitive mechanisms adapted for other purposes – evolutionary accidents, in other words. However, if religious beliefs are accidental byproducts, we might expect natural selection to have eradicated them because (as M&D note) they impose significant fitness costs in terms of time, effort, and resources (Sosis & Alcorta 2003). So why do they persist?

Even if some religious beliefs persist as "sticky" cultural parasites, it does not preclude them from also promoting individual or group fitness at certain times or contexts (in which case they may not be "parasites"). The universality and power of religious beliefs of some form or other – despite their costs – to billions of people around the world, every culture in history, and every hunter-gatherer society, strongly suggests that religion confers adaptive fitness benefits, for individuals and/or groups (at least in some contexts, for some people, and for some periods of human history). Of course, universality need not imply adaptation: other non-adaptive traits such as chins and male nipples are also globally and historically universal. However, they do not impose significant costs. Religion does.

The only theories that solve this paradox are religion-as-adaptive hypotheses that propose how costly (mis)beliefs beget even greater benefits for individuals and/or groups (Johnson 2008; Norenzayan & Shariff 2008; Sosis & Alcorta 2003; Wilson 2002), or are outweighed by the costs of non-belief (Cronk 1994; Johnson 2009; Johnson & Bering 2006). Byproduct theories of religion offer no solution to its greatest puzzle, for God would be a costly accident.