substance, only the expectation (perhaps never actually realized) that there *ultimately is* a physical source of intentional action (Csibra et al. 1999). Ontological violations block such expectations being realized *even in principle* (e.g., invisible agents versus heard but unseen beings). They countermand rules for eventual processing, not actual perception. Second, not all mental states are equally bound to ordinary intuitions about bodies. Recent studies indicate that children from 5 years on up more readily attribute epistemic mental states (see, think, know) to beings in the afterlife than psychobiological mental states (hunger, thirst, sleepiness) (Bering & Bjorklund 2002). Ordinary distinctions between mind and body (e.g., dreaming) thus seem to provide at least *some* intuitive support for extraordinary beings with disembodied minds (Hobbes 1651/1901).

14. Barrett and Nyhof (2001, p. 79) list as common items: "a being that can see or hear things that are not too far away"; "a species that will die if it doesn't get enough nourishment or if it is severely damaged"; "an object that is easy to see under normal lighting conditions." Such items fall so far below ordinary expectations that communication should carry some new or salient information that Barrett and Nyhof (2001, pp. 82–83) report: "common items were remembered so poorly relative to other items. . . . In some instances of retelling these items, participants tried to make the common property sound exciting or unusual." In other words, some subjects tried to meet minimum conditions of relevance (Sperber & Wilson 1986). For the most part, common items failed these minimum standards for successful communication.

15. Highest degradation was observed in the mostly MCI and all INT conditions, conforming to an inverse quadratic function, $F(3, 89) = 4.49$, $p < .05$. Memory degraded least in the Mostly INT condition, and increased as the proportion of MCI beliefs increased, resulting in a linear trend, $F(2, 65) = 3.53$, $p = .06$.

16. Only additional evidence could show whether children "continue" to think of God in the same way after they become aware of false beliefs (as Barrett et al. 2001 intimate), or (as seems more likely) come to have different reasons for thinking that God would not be deceived.

17. To deal with deficits in counterfactual thinking, St. Paul's Church in Alabama (Trenton Diocese) has a special program for autistics: "The church requires that children who receive Holy Communion be able to recognize the difference between ordinary bread and the Eucharist. . . . The St. Paul's program was designed to teach the difference" (Rev. Sam Sirianni, cited in Raboteau 2000).

Open Peer Commentary

Gods are more flexible than resolutions

George Ainslie

Veterans Affairs Medical Center, Coatesville, PA 19320.
George.Ainslie@med.va.gov **http://www.Picoeconomics.com**

Abstract: The target article proposes that "counterintuitive beliefs in supernatural agents" are shaped by cognitive factors and survive because they foster empathic concern and counteract existential dread. I argue that they are shaped by motivational forces similar to those that shape our beliefs about other people; that empathic concern is rewarded in a more elementary fashion; and that a major function of these supernatural beliefs may be to provide a more flexible alternative to autonomous willpower in controlling not only dread but also many other unwelcome urges.

The useful hypotheses in this article include: (1) that religion is a form of motivated belief, that is, that religious beliefs and their attendant practices survive insofar as they serve a purpose; (2) that a principal purpose of religion is to deter "social deception and defection in the pursuit of self-preservation"; (3) that another principal purpose of religion is to control "emotionally eruptive existential anxieties" (sect. 1, para. 7); and (4) that human experience, and religious experience in particular, converges "on more or less the same life paths – much as rain that falls anywhere in a mountain-valley landscape, drains into a limited set of lakes or rivers" (sect. 8, para. 3). The authors present a case for how humans may be innately prepared to construct the supernatural beings that populate most religions, because of people's "hair-triggered" attribution of agency to ambiguous percepts, the increased memorability of "minimally counterintuitive" ideas, and people's ability to imagine counterfactual omniscient personae. However, this article presents little about what incentive people have to construct these beings – only some unsurprising data that subjects value religious ideas more in fear-provoking situations.

I agree that supernatural religion is probably an extension of "emotional mechanisms that evolved for mundane adaptive tasks" (sect. 1, para. 2), and that part of its usefulness is sometimes to control selfishness and emotional eruptions. However, I do not think the authors have specified adequate motivational mechanisms to account for these effects. Part of this problem comes from the inadequacy of how behavioral science has come to imagine self-interest and altruism. Rational self-interest is identified with beating out competitors for resources, and rational altruism merely with taking the long view of this competition so as to identify situations where cooperation will be more profitable, hedonically or genetically, than competition (Dawkins 1989; Frank et al. 1993). Given the human openness to seduction by short-term prospects, altruism is sometimes suggested to require self-control (Rachlin 2002), but the point is still to maximize your own survival resources. The authors are right to reject this "'mind-blind' functionalism" (sect. 1.5); but the role they give to religious belief remains one of controlling an innate tendency toward selfishness, through belief in vigilant gods.

An adequate theory of altruism needs to explain why people start out as highly empathic children (Harris 1987; Zahn-Waxler et al. 1992), who then learn to a variable extent to control empathy as an impulse. That is, why is there a basic self-interest in cultivating vicarious emotional experience, which is then partially displaced by the more "objective" self-interest of (say) economic man? This area is largely terra incognita. Motivational theory has not examined even nonvicarious emotions as rewards until recently (Lewis & Haviland-Jones 2000); they are awkward targets for controlled research, and it is hard even to theorize about rewards that require no specific stimulus and have many of the characteristics of behaviors. However, mounting evidence that all reward-responsive organisms discount delayed rewards proportionally to this delay (hyperbolically) rather than at fixed rates (exponentially; Kirby 1997) suggests one mechanism for vicarious emotional reward, based on the innate impulsiveness that such discounting predicts (Ainslie 1995; 2001, pp. 161–86). I can only summarize it here: Emotions are reward-dependent behaviors that have their own appetites and lead to their own innate rewards, rather than being elicited reflexes. Because of a hyperbolic impatience for their rewards, these behaviors are limited by premature satiation, which causes extinction of deliberately emitted emotions; to stay fresh they must be occasioned by uncontrollable events. Such a contingency makes external occasions for emotion valuable, and these occasions seem especially well paced by the apparent experience of other people. Thus, vicarious reward creates an incentive to help the people whose experiences you choose as occasions for emotion, and to resist temptations to exploit them. The recent discovery of "mirror neurons" that initiate copies of other people's behaviors (Iacoboni et al. 1999) suggests a reason why vicarious experience may stand out from other available occasions for emotion. Whatever the mechanism, empathic engagement with its sometime result of altruism is apparently a primary motivated process.

We perceive other people's experiences not piecemeal but through mental models, and we construct models of gods in the same way that we construct models of each other. These models reflect our take on what others are going through, modified by projection, transference, and other distortions. Ordinarily we "believe in" other people (as opposed to how we experience fictional characters) only when we can test our models against observations of them. However, when the models are especially evocative, we may lower our threshold for belief and experience a dead relative, or Elvis, or a god as present. Such extra occasions for emotion are valuable in their own right – as valuable as the emotions are – but insofar as they can remain robust without confirmatory evidence from actual people, they may also improve our self-control.

Selfishness that gets too much in the way of vicarious reward is an impulse that needs to be controlled, as are not only "emotionally eruptive existential anxieties" and other corrosive emotions but also the self-destructive urges that get called sins. Most of these cannot be subsumed under selfishness. Of the seven deadly sins of Christianity, for instance (gluttony, lust, wrath, pride, envy, avarice, and sloth), only wrath and avarice could be argued to be as harmful to others as they are to the sinners themselves. Self-control is a broad task, and it is central to religion.

Self-control is usually regarded as the function of willpower; but I have argued elsewhere that willpower is nothing more than the fruit of recognizing a limited-warfare relationship among successive selves – another product of hyperbolic discounting – and that it suffers from the same limitations as other solutions to limited warfare (Ainslie 2001, pp. 90–104, 143–60). Specifically, willpower is the technique of regarding choices as test cases for how you will decide in similar future cases; great reliance on this technique leads to rigidity and the risk of permanent damage to willpower in cases where the will fails. That is, autonomous self-control can lead to the kind of lawyerliness that psychologists call compulsiveness and theologians call scrupulosity. But the obvious alternative commitment method, openness to the influence of actual other people, is fallible – this influence is itself impulsive at times, evadable, and sometimes self-serving.

Here is where a felt relationship with a god or even a sentient ancestor (e.g., "I can just hear Mother") could be a solution. Your sense of being on good terms with this entity forms the stake that you bet against impulses; but the entity is not rigid as a resolution is. It is, rather, a mental model like your model of other people, and made of human expectations. The information that shapes this model into a felt presence comes indirectly, from the forms of communing and divination to which the authors refer, and is not normally controlled by any one individual. Furthermore, there can be ways that you can overcome your expectation that the entity is angry or disappointed ("atonement") – not surefire ways, which would undermine your experience that the entity is genuinely another agent, but ways that might be more effective than efforts to repair an autonomous but failed will.

In sum, the mundane transactions from which the supernatural is formed need to be more motivationally important than just hair-trigger attributions, mnemonic advantages, and a rich imagination, although all of these may have their role. What I have sketched is just one possibility, but it illustrates the potential for functional modeling when a mechanism for motivational conflict is added to the mixture.

NOTE

The author of this commentary is employed by a government agency and, as such, this commentary is considered a work of the U.S. government and not subject to copyright within the United States.

Counterfactuality in counterintuitive religious concepts

Justin L. Barrett

31 Harrow Lane, Lexington, VA 24450. **justinlb@umich.edu**

Abstract: In sketching a preliminary scientific theory of religion, Atran & Norenzayan (A&N) generally agree with cognitive scientists of religion in the factors that coalesce to form religion. At times they misrepresent, however, the notion of "counterintuitive" concepts as they apply to religious concepts, confusing counterintuitive with counterfactual, category mistakes, and logical contradiction.

Presenting again the theoretic core of Atran's recent book on the subject (Atran 2002a), Atran & Norenzayan (A&N) rightly highlight the central factors currently occupying comprehensive theories in the cognitive science of religion. As Boyer foreshadowed throughout the past decade (e.g., Boyer 1994; 1995; 1996; 1998b) and detailed more recently (Boyer 2001; 2003), a thorough-going theory of religion should account for the convergence of a number of recurrent features of religions: counterintuitive concepts centering on intentional agents, collective practices that result in enhanced group cohesion, and the connection of these concepts and practices to morality and existential concerns such as death. Similar to how Boyer (2001; 2003) and I have written about the convergence of these mutually reinforcing features, A&N see a "canalization" of factors due to evolutionary forces. For the sake of clarification, I will amplify the notion of "counterintuitive" concepts as characteristic of religious cognition.

A&N rightly note the recurrence of counterintuitive concepts as central components of religious traditions. Following Boyer (1994; 2001; Boyer & Ramble 2001), *counterintuitive* has acquired a peculiar meaning in the cognitive science of religion. A counterintuitive concept is one that violates intuitive assumptions about the properties of a particular thing. These intuitive properties derive from culturally independent implicit reasoning systems. To illustrate, as has been demonstrated by developmental psychologists, the understanding that physical objects will fall unless supported arises in infancy and thus becomes an intuitive assumption for physical objects. A solid, physical object that does not require support, but may remain hovering in mid-air would be counterintuitive in this technical sense.

A&N frequently use the terms *counterintuitive* and *counterfactual* together. Note, however, that counterintuitive and counterfactual are not the same thing. Though we typically trust our intuitions to give us truthful assumptions about the world, they only serve as best guesses and may be false. Likewise, counterintuitive but factual conditions and properties abound. For instance, Venus flytraps violate our intuitive assumptions regarding the nonpredatory and inanimate character of plants; that invisible microorganisms can kill large mammals is counterintuitive; and that the earth revolves around the sun violates our intuitive evaluation of visual information. Indeed, one of the striking (and valuable) features of science is its ability to demonstrate that the physical world sometimes does *not* match our intuitive assumptions. Science is frequently counterintuitive (McCauley 2000).

Apart from increasing precision, distinguishing concepts' factuality from concepts' intuitiveness pays critical theoretical dividends for a *scientific* treatment of religion. Most importantly, it liberates the scientist from having to play philosopher, theologian, or anti-theologian and having to decide whether particular metaphysical claims are true or false before being able to consider concepts as religious or not. Such evaluations lie outside the tools of science.

Counterintuitive concepts also must be distinguished from "category mistakes" and contradictions. A category mistake involves modifying a thing with a predicate that does not and may not meaningfully apply to its ontology. For example, a "god that happened yesterday" would be a category mistake but is not counterintuitive (in the technical sense Boyer has coined). Such a no-