Jesse M. Bering

Response to commentaries on BBS article "The Folk Psychology of Souls"

Main Text Word Count: 6,037

Title

The cognitive science of souls: clarifications and extensions of the evolutionary model

Abstract (77 words)

The commentaries are a promising sign that a research programme on the cognitive science of souls will continue to move toward empirical and theoretical rigor. Most of the commentators agree that beliefs in personal immortality, in the intelligent design of souls, and in the symbolic meaning of natural events can provide new insight into human social evolution. Here I clarify and extend the evolutionary model, further emphasizing the adaptiveness of the cognitive system that underlies these beliefs.

R1. Introduction

Only in the past few years has the cognitive substratum of religious and supernatural experience been penetrated by the precision tools of experimental science (for reviews, see Atran & Norenzayan 2004; Barrett 2000; Boyer 2001; Whitehouse 2004). Shaky hands are to be expected in these first attempts at scientific exploration, and some trembling is also evident at places in the target article. But, as **Bainbridge** so elegantly discusses in his commentary, picking up these tools to begin with is the hard part, and I am very grateful to the commentators for offering their expert hands to steady my own.

Since the focus of most of the commentaries concerned the evolutionary model and because this issue is central to the theoretical framing of the target article, most of my response will be devoted to that topic, highlighting theoretical points in the other commentaries whenever they have some bearing on the evolutionary history of a folk psychology of souls.

R2. Clarifying the evolutionary model

By far the most frequent criticism in the commentaries questioned my evolutionary model, seeking greater clarity and further attention to deviant cross-cultural examples of religious belief systems that seemingly violate the model's core assumptions. For example, **Johnson and Nyhof** protest that "the Darwinian mechanisms are left completely unspecified." **Pyysiäinen** finds similar fault with the paper's evolutionary cast ("The weakest part of Bering's contribution is precisely his evolutionary speculations"), a concern echoed in the vituperative commentary by **Hegdé and Johnson**, among others. According to the majority of commentators who addressed the evolutionary issues, the explanatory framework into which I have placed my research programme is flawed on several counts.

To recount the original argument, I claimed that three basic cognitive mechanisms—those that produce illusions of personal immortality, of teleological authorship in the design of individual souls, and of natural events as having symbolic meaning—formed an organized "system" at some point in recent human evolution as a result of the unique selective pressures operating in our social environment. I placed the word system in scare quotes because I see these illusions as being connected through a sort of abstract, conjunctive tissue that biases reasoning about personal existence, not as a modular, task-specific system yielding static behaviours independent of cultural variation (**Livingston**; **Pyysiäinen**). I have never claimed that "religion is innate" (**Bloch**) or even that there exists "an evolved system dedicated to afterlife beliefs" (**Harris & Astuti**; also Pyysiäinen). These are oversimplifications.

My evolutionary model identifies a suite of very basic cognitive building blocks, often but not always associated with religious beliefs. If this suite indeed comprises a true psychological adaptation that motivated adaptive responses under recurrent challenging conditions, then it should be *canalized* in modern humans, present in anyone with a normal cognitive profile (**Flusberg & Tager-Flusberg**) who does not develop under extreme and species-atypical conditions (**Evans & Wellman**). **Hegdé & Johnson** conflate adaptations, which have a heritability of zero, with heritable individual differences associated with adaptations (see Tooby & Cosmides 1992 pp. 122-131), presumably those that would contribute to varying degrees of religiosity in modern humans.

R2.1. The fact that there may be reputation maintenance mechanisms other than those entailed by the folk psychology of souls presents no difficulties for my evolutionary model

Several commentaries suggested that my evolutionary model was flawed because there are other psychological mechanisms, none of which involve souls or supernatural agents or seeing signs in natural events, which serve a reputation maintenance function (**Beit-Hallahmi; Boyer; Ferrari; Gjersoe & Hood; Greenberg et al**). This is a point elaborated by Boyer, who, after summarizing several alternative mechanisms that promote pro-social behaviours in humans, tells us that, "All these dispositions and processes evolved independently of supernatural and religious beliefs, operate in similar ways in people with or without such beliefs and regardless of differences in these beliefs, and recruit different neuro-cognitive machinery from the supernatural imagination."

I do not see a conflict here. As far as I am aware there is no law in natural selection theory stating that there cannot be distinctly evolved mechanisms serving the same adaptive purpose. On the contrary, if these mechanisms delivered a cumulative, buffering effect in solving a shared adaptive problem or—at the very least—did not impede one another's functioning, evolution should favour the selection of multiple adaptive designs. Moreover, contrary to the criticisms raised by **Hedgé & Johnson**, an evolved folk psychology of souls meets the important criterion of Darwinian conservativeness because the types of existential illusions generated would have emerged through a set of biases produced by pre-existing structures. The system that I have outlined therefore would not have required any substantive neuro-cognitive reorganization.

Several commentaries discussed the role of shame in inhibiting normatively deviant or antisocial behaviours (**Beit-Hallahmi; Ferrari; Gjersoe & Hood**), particularly how parents instil these feelings, and implied that this obviates the adaptive utility of belief in supernatural observation or punishment. Shame, however, is usually experienced after a social transgression has already occurred; it is the emotional aftermath of transgression (Tangney 2003). Although the negative affect associated with this experience may serve to discourage similar actions in the

future (see Fessler & Haley 2003; Gilbert & McGuire 1998), and may attenuate severity of punishment for an offence (Gold & Weiner 2000), shame may not be very effective at preventing the occurrence of a proscribed behaviour in the first place. Shame and observability, of course, are not mutually exclusive—in fact they are sister constructs. But belief that one is under surveillance by supernatural agents, and that there are consequences for misdeeds even when they occur in private, may effectively deter socially proscribed behaviours even in the absence of shame.

R2.2. The importance of avoiding solitary incidents of serious transgressions must be emphasized

Although reputations are mostly cumulative and can perhaps be formulated as an image score that people use to guide their interactions with social others (e.g. Nowak & Sigmund, 1998), a single black mark can erode an otherwise unbroken record of altruistic tendencies. "Words are wolves," according to Jean Genet. Language would have enabled our ancestors to essentialize others into social category memberships through the heuristics of emotionally loaded words. One need only consider how socially powerful are terms such as "rapist," "paedophile," "thief," "murderer," "slut," "racist," "child abuser," or, recently, "terrorist," to see the hazards of a publicly revealed, solitary moral breach from the gene's point of view.[1]

Nemoroff and Rozin's (1994) findings of moral contagion (i.e. emotional aversion to physical objects such as clothing that have been in contact with representatives of such derogated social categories) may be seen as evidence of this type of negative essentialism (**Gjersoe & Hood**). Any public distancing from socially repudiated others would serve to advertise a personal commitment to in-group norms—that one is not like the derogated individual. The sociologist Erving Goffman (1963) noted that people who are wanted on criminal warrants were once

referred to as "having smallpox" and their criminal disease was said to be catching. Merely being seen with them could lead to arrest on suspicion.

The folk psychology of souls (which at its core constitutes a social relationship between the self and supernatural agents) would have helped our ancestors to censor selfish decisions associated with others judging them as being essentially bad and/or morally undesirable (and therefore to avoid the genetic consequences of this labelling). This was particularly the case in situations where people were strongly tempted by selfish desires and underestimated the likelihood of detection by other in-group members. Miscalculating the odds of social exposure for certain behaviours would have had calamitous effects on reputation and therefore genetic fitness. The folk psychology of souls provided adaptive illusions of watchful supernatural agents that helped to counteract these dangerous miscalculations. These illusions involved seeing supernatural agent(s) as being emotionally invested in the self's existence, as sharing (or at least understanding) the in-group's moral values, and as communicating their attitudes and opinions *about* the self through the occurrence of natural events and biographical experiences (Bering 2002a; Bering & Johnson 2005; Bering et al. 2005; Johnson & Bering in press). Natural events and biographical experiences were perceived as the "evidence" that such supernatural agents were real (cf. **Bullot**) and were capable of punishing and rewarding social behaviours, either in this life or in the hereafter.

R2.3. Experimental findings demonstrate social sensitivity to being observed

Beliefs in watchful supernatural agents appear to militate against the psychological state of *deindividuation*, which occurs whenever "individuals are not seen or paid attention to as individuals" (Festinger et al 1952, p. 382). Festinger and his colleagues described how deindividuation is strongly associated with social disinhibition and loss of inner restraints.

Numerous laboratory experiments have in fact shown that participants who believe that they are making decisions under anonymous conditions tend to be less altruistic, more aggressive, and more punitive than those who believe that their identities are known (e.g., Diener et al 1976; Ellison et al 1995; Rehm et al 1987; Zimbardo 1969).

Building on experimental economic games, a flurry of recent studies have also provided evidence that ambient gaze, even when artificial, unconsciously primes pro-social behaviours in human participants (Bateson et al 2006; Burnham & Hare in press; Haley & Fessler 2005; Milinski et al 2002; Wedekind & Braithwaite 2002). Burnham and Hare (in press), for example, found that people made more altruistic decisions in a task involving allocation of scarce resources even when the "witness" was simply an image of a robot with large human-like eyes. Similar results were reported by Haley & Fessler (2005), where participants behaved more generously on a computerized task when stylized eyespots were present on the screen, which the authors interpreted as evidence that subtle cues concerning observability factor prominently in reputation management. (**Gjersoe & Hood's** discussion of Titchener's classic unseen gaze findings, where people believe they can "feel" when others are looking at them behind their backs, may be interpretable within this evolutionary framework as well; see also Colwell, Schröeder, & Sladen 2000). Finally, as discussed in the target article, Bering et al (2005) found that, when left alone in a room, participants who were led to believe that a ghost may be observing them cheated less on a competitive task compared to those who did not receive this supernatural prime.

Real-world findings provide complementary evidence that perceptions of anonymity are positively correlated with antisocial behaviours. In a cross-cultural analysis of warfare practices, for example, Watson (1973) discovered that warriors who hid their identities before going into battle were more likely to kill, mutilate, and torture than those who did not. More recently, Silke (2003) found that, of all sectarian violence incidents reported in Northern Ireland over a two-year

period (1994-1996), paramilitary members who wore masks during their offences attacked more people, inflicted more serious injuries, committed more acts of vandalism, and were more likely to threaten their victims after attacking them than paramilitary members who were implicated in sectarian violence but who did not hide their faces.

R2.4. Cross-cultural variability, supernatural beliefs, and evolutionary dynamics

Although belief in supernatural observability has not yet been targeted as a key research question in evolutionary models of religion, the ethnographic literature does suggest that such beliefs feature prominently in most religious systems. In Pettazonni's (1955) cross-cultural analysis of the types of attributes that are most frequently attributed to the gods, one recurrent and defining characteristic is the gods' deep *knowing* of people as unique individuals (i.e. their "hearts and souls"). In Borneo, the Iban believe that "anyone who successfully cheats another, or escapes punishment for his crimes, even though he may appear to profit temporarily, ultimately suffers supernatural retribution" (Sandin & Sather 1980, p. xxviii). And Malinowski (1935, viii) wrote that "from the study of past religions, primitive and developed, we shall gain the conviction…that every religion implies some reward of virtue and the punishment of sin." Implicit here is the assumption that supernatural agents who dole out moralistic consequences are believed also to survey and observe private behaviours, keeping their thumbs on individuals within the group.

Cohen et al ask whether "variations in beliefs in afterlife or observant spirits are linked to recurrent variations in social or physical ecology" (also **Whitehouse**; see Reynolds & Tanner 1995). Although we do not yet have the data to answer this important question, structuring the present evolutionary model under these (ecologically dynamic) terms may put into context the striking cultural diversity associated with the moral dimension of supernatural beliefs. For example, answering this question would potentially be capable of addressing the sceptical query

posed by **Greenberg et al,** who ask, "If immortality beliefs were a simple default by-product of cognition, why would these beliefs be so varied across cultures and so complex?" At the moment, I agree that such variation is difficult to understand, but this is due to the embryonic stage of data gathering in this area, not to any serious limitations of the simulation constraint hypothesis or the general evolutionary theory I have offered.

The commentaries reveal a wide variety of religious beliefs that appear uniquely tied to specific cultures, geographic areas, and historical settings. Why these adaptive climates give rise to particular beliefs and not others is a question for evolutionary analysis, just as **Cohen et al** reason. For example, collective symbolic interpretations of disease and misfortune may serve to enculturate children into specific moral environments (environments that are themselves products of specific ecological and social factors). Such symbolic interpretations offer children a very clear picture of what it is that their society does not condone. In their fascinating description of Cotard's syndrome, **Cohen & Conseli** write that "collective and cultural significance dominates biographical experiences…first syphilis then AIDS symbolized the amalgam of flesh, punishment, sin, guilt, sexuality and the devil."

We need not look at exotic cultures to see how collective symbolic interpretations of natural events can influence moral development. As an eight-year-old, I once became panic-stricken that an upcoming doctor's visit, which I knew would involve a routine drawing of blood, would publicly identify me as a homosexual. I was naïve to the medical facts about how people contracted the HIV-virus, but I knew that AIDS could be detected in blood. I also understood that many saw AIDS as a moral condemnation of gay men, specifically as God's culling of homosexuals. Whether I personally saw such a moralistic message in AIDS was inconsequential for this cultural illusion to impact on my decision not to divulge my sexual orientation, a decision that can be understood within fitness-related terms. Other peoples' symbolic interpretation of this disease was enough to teach me that something in my blood would expose me as being *essentially* bad, worthy of being shunned—and in fact I had such anticipatory anxiety about the social consequences of being labelled a "homosexual" that I collapsed in the waiting room.

Although I agree with **Pyysiäinen** that we are not in a position to advance a detailed evolutionary argument until a "more rigorous methodology" is developed—a task that will require massive interdisciplinary collaboration—it is unclear to me how one could ever begin to construct such a methodology without first having a general evolutionary theory capable of generating hypotheses and offering an interpretive lens through which to view the findings. I have posited a general evolutionary theory that can act as such as a crucible for weighing competing, non-adaptationist hypotheses, something recognized by several commentators (**Cohen et al; Evans & Wellman; Hughes**; **Whitehouse**).

R2.6. Absent third party punishment is a uniquely human adaptive problem

Gjersoe & Hood comment that "many social animals also show behavioural inhibition and pro-social behaviour without necessitating a specialized cognitive mechanism for belief in souls." This is not in debate. But what these commentators overlook is the fact that theory of mind, and the concomitant emergence of declarative language, introduced a genuinely novel adaptive problem in human sociality—that of *absent third party punishment*. In short, absent third party punishment is any punishment that is administered by a person (or persons) who were not present at the time of the offence, but who learned about the offence through a second-hand source (Fehr & Fischbacher 2004). Human beings are able to mentally represent an absent third party's state of ignorance about the unobserved event and are strategically motivated—and emotionally driven—to disclose their victimization to these naïve third parties through declarative language. This is in fact the very basis of all criminal justice systems, no matter how informal.

What is unique about human sociality is that anybody who witnesses a social event is a *carrier* of strategic information who can then transmit that information to other minds, over great

spans of time and geographical distances. "Seeing" therefore took on new meaning for human beings, the only species for which, given these social cognitive verities, short-term selfish gains were traded in for long-term reputation gains. According to Johnson (2005, p. 414), "Information about person A could propagate via person B to person C, D, E, and so on…even if person B and C do not care, it may not be until person Z hears the news, or until *enough* people hear the news, or until some authority hears the news, perhaps weeks later, that punishment will come." Given the calamitous effects a mired reputation could have on the actor's genetic fitness (through punitive tactics such as castigation, ostracism, exclusion, group expulsion, or even execution), the presence of nearly any watchful agent, human or supernatural, became capable of influencing behavioural decision-making.[2]

R2.7. Selective pressures for solutions to the adaptive problem were intense

It is impossible to overstate how strongly the third party punishment problem would have influenced the course of human social evolution. This is especially evident when one considers the relatively low degree of privacy afforded to our ancestors, who lived in small-scale gossipy societies of only 120-150 individuals (Dunbar & Spoors 1995). In the environment of evolutionary adaptedness, individuals would have been unable to easily emigrate to new social groups and to "start over" if they spoilt their reputation in their natal group (a strategy of sociopaths in modern societies, Mealey 1995).[3]

Notice that inclusive fitness is also likely to be negatively impacted by a spoiled public identity because of a *sanguineous bias*, stigma attached to the biological kin of the individual whose reputation is impugned through transgression (e.g. see May 2000 for stigma effects on murderers' relatives). This means that third party punishment does not necessarily end at death. These inclusive fitness issues concerning the effects of reputation on biological kin also mean, in

principle, that effectively managing reputation is a more pressing evolutionary problem than mitigating existential anxieties through symbolic immortality (**Greenberg et al**). Indeed, many of the extensive findings from the Terror Management Theory literature can be understood in these terms. If one is reminded of his own inevitable death, better for his family members' genes that he go out as a staunch, reliable defender of his community's values than as apathetic or as a social dissident.

R3. Propositional beliefs about the supernatural do not always cause behaviour (and sometimes they are in opposition to behaviour)

It is important to understand that the three existential illusions identified in the target article (immortality, teleological authorship of the soul, and symbolic natural events) may not be as salient in industrialized societies today as they were in the environment of evolutionary adaptedness, where they were unlikely to be punctured by scientific knowledge or discouraged through cultural secularization. Even in modern scientific nations, however, among well-educated and scientifically literate people, the biases identified in the target article are not recognized as illusions and continue to have deep emotional resonance. **Sandelands**, for example, concludes his theologically inspired commentary by stating that, "a full and true study of man must begin in God." (In some sense this is true: Our species like any other must be understood within the parameters of the modern synthesis and God is just another slave to human genes.) And, moreover, even in recognizing them as illusions we fail to sever their emotional underlay, which may still pump-prime behaviour—the level at which natural selection operates. I do not *believe* in the afterlife, but as a potential homeowner I certainly would *feel* uncomfortable living in a house where a stranger has recently died. In this case it is my eerie feelings and not my belief or disbelief in the afterlife that would be a better predictor of whether I make an offer on the house. This is not to say that propositional beliefs about religion and the supernatural are frequently

epiphenomenal, but rather they are more properly viewed as rough indices of unconscious

reasoning (and perhaps phenomenal states) than as accurate predictors of behaviour.

R3.1. Global secularization cannot extirpate a true psychological adaptation

      Beit-Hallahmi writes that, "The global secularization process means that we no longer

interpret misfortune as caused by supernatural agents." But this "god-of-the-gaps" hypothesis has

now been disconfirmed in social psychology experiments (see Weeks & Lupfer 2000 for an

account of distal-proximal attributions to God). Moreover, the argument that scientific or secular

explanations "replace" more naïve or irrational supernatural explanations is intuitively

unpersuasive; obviously they can occur alongside one another (e.g. Subbotsky 2001).

Theologians who saw the recent tsunamis of East Asia as an angry, moralistic message from God

were probably not naïve to the fact that they were caused by earthquakes on the Indian Ocean

floor. No matter how culturally secularized we become, God pokes through, whispering in the

most godless of scientists' ears. At the end of their commentary, **Gjersoe & Hood** correctly point

out that in order to understand supernatural beliefs from a scientific perspective we must first

acknowledge and recognize our own supernatural dispositions.

      The best research designs in the cognitive science of religion are those that are able to pry

apart unconscious reasoning from explicit or "theologically correct" religious beliefs (see Barrett,

2000). Socrates' "idea of immortality" as described by **Ferrari** is therefore of questionable

countenance, since this "reasoned conclusion" would be heavily influenced by the same

underlying cognitive constraints that motivate others to think in this fashion. Innate psychological

biases with regard to the supernatural (and the behaviours they generate) reveal themselves most

clearly when they directly contradict stated beliefs. For example, those who believe God can do

everything at once actually reason as if God were constrained by a human attention span (Barrett

& Keil 1996); some people who believe that the mind stops at death nevertheless reason about a dead person as if he still has thoughts (Bering 2002b); scientific theorizers are wary of magical incantations (Subbotsky 2001); others who consider themselves to be materialists refuse to sign a contract relinquishing their souls at death to an experimenter (Haidt et al 2004); and children who say they don't believe in monsters shy away from a box they are told contains a monster (Harris et al 1991).

Similarly, if the folk psychology of souls is a true psychological adaptation, then it should be empirically detectable even in atheists. For example, McAdam's (2001) findings from narrative psychology suggest that people tend to fall into one of two categories: those who view personal misfortunes as *contaminative episodes* in their life stories (where the event permanently disrupted an otherwise positive life course and cast a dark shadow over their biographies), and those who view such events as *redemptive episodes* (where the event, although difficult at the time, was responsible for a positive redirection of their life course). It may be possible to detect intentionality themes in atheists' self-narratives through the use of such paradigms (e.g. "it was a 'life lesson,' "it wasn't supposed to happen," and so on).

R4. Developmental considerations

Evans & Wellman argue that, "if Bering's selectionist explanation was on target then one might predict a unique and relatively robust developmental trajectory, regardless of input." This is certainly true, and I believe that this trajectory will be borne out. To test Evans & Wellman's prediction, we need first to have an accurate developmental model that delineates the ages at which the three existential illusions (immortality, teleological authorship of the soul, and symbolic natural events) appear in childhood. We do not yet have enough data to construct such a model and therefore developmental research in this area is urgently needed. Although **Ferrari**

and **Estes** are right to point out that cognitive developmentalists have for decades been exploring related questions about children's distinction between the mind and body, particularly in the area of theory of mind, this "abundant research" (Estes) hardly constitutes a targeted attempt at systematically revealing the social cognitive factors that lend themselves so seamlessly to the existential illusions highlighted in the article. On the contrary, such a targeted research programme is strikingly absent, not only in developmental psychology, but in all the subdisciplines of experimental psychology.

R4.1. Contradictory findings on the development of children's afterlife beliefs

It appears that the little we do know about the development of a folk psychology of souls is contradictory, as discussed in the commentaries by **Evans & Wellman** and **Harris & Astuti**. These commentators tell us that recent findings on children's afterlife beliefs have failed to replicate the pattern reported by Bering & Bjorklund (2004; also Bering, Hernandez-Blasí, and Bjorklund 2005). For example, Harris & Gimenez's (2005) findings suggest that afterlife beliefs increase with age rather than decrease and are moderated by the religious context of the experimenters' questions (with children being more likely to endorse psychological functioning after death when information about the dead character includes words like "priest" and "God"). Therefore, Harris & Astuti question whether belief in the afterlife is in fact a default cognitive stance.

R4.2. Some "contradictory" findings may not be contradictory

This conflicting pattern of developmental findings, however, is difficult to interpret at present. To begin with, the central research questions motivating these other studies on children's concepts of death are very different from my own (as well as from each other) and the

methodologies vary accordingly. **Evans & Wellman** cite work by Barrett & Behne (2005) as evidence that, in contrast to my findings, 4- and 5-year-olds in this study did not attribute psychological states to dead agents. Barrett & Behne's study, however, did not investigate children's afterlife beliefs, but instead concerned children's ability to differentiate between dead and sleeping animals in the physical environment. The investigators reasoned that this is an adaptive function in that it prevents unnecessary vigilance toward the bodies of dead animals through the cue-driven activation of an innate "living/dead remapping mechanism."

In the study by Barrett & Behne (2005), children were asked five questions about the dead versus sleeping animal: *Can it move? Know you were there? Move if touched? Can it be afraid? Can it hurt you?* The fact that the youngest children answered "no" in reference to the dead animal, but "yes" in reference to the sleeping animal, is hardly *prima facie* evidence against my argument that belief in the afterlife is a cognitive default. In fact, if belief in the afterlife is a cognitive default, then we would actually predict the pattern of findings reported by Barrett & Behne (2005). That is, three-year-olds should answer "no" to questions about the *bodies* of dead animals (notice the key word "it" in the questions posed to children) if indeed they view the mind as being liberated from the body at death.

R4.3. Methodological concerns presently limit theoretical inferences

Like **Evans & Wellman**, **Harris & Astuti** state that their own research programme on the development of afterlife beliefs reveals a set of findings that in many ways contradict the developmental trajectory reported by Bering & Bjorklund (2004), or at least tells a more complicated story with religious testimony and cultural exposure encouraging such beliefs. Again, however, it is difficult to compare findings across these studies. We deliberately avoided eschatological language in our research design because we were wary of biasing children's

answers through the experimenters' language and behaviours, and in fact our empirical reports list many of the safeguards we used to protect against such biases (**Hughes**). In contrast, such language was an important manipulated variable for both Harris & Giménez (2005) and Astuti & Harris (2006).

Furthermore, the coding procedures used to determine whether children attributed continued psychological functioning to a dead agent meaningfully differed between our studies and those described by **Harris & Astuti**. Our data were coded on the basis of children's follow-up answers to the questions rather than their initial yes or no response. We reasoned that a "no" response is inherently ambiguous and should not be seen as clear evidence for non-continuity judgements after death. Young children in our study often answered "no" to the initial questions about the dead agent's continued capacities ("Can Brown Mouse still see?"), but upon further questioning it became clear that they were nevertheless reasoning in terms of an afterlife (e.g. "…because it's too dark in the alligator's belly"). Harris and Giménez (as well as Astuti & Harris, 2006 and Barrett & Behne, 2005) failed to operationalize children's "no" answers in this way, instead taking them at face value as evidence of an understanding of the non-functionality of the capacity in question.

It is therefore impossible to know whether the findings these authors report is a product of the religious context of the story, as they argue, or is in fact an artefact of their coding procedure. In addition, Harris & Giménez (2005) treated their religious/secular variable as a within-subjects factor, so that all children heard the two death narratives in the same order, first the religious narrative ("Now that Sarah's grandmother is with God, *can she still…*") and then the secular narrative ("Now that Bill's grandfather is dead and buried, *can he still…*"). This potential confound of an order effect, where the demand characteristics of the study are so transparent (especially to older children), again makes it difficult to make theoretical inferences based on

these data. Finally, the youngest children in the Harris and Giménez study were seven-year-olds, whereas our most robust findings for afterlife beliefs came from the three- and four-year-olds we tested, providing the basis of our nativist claims.

R5. Cognitive processes underlying the folk psychology of souls

Several of the commentaries focused on the precise mechanisms by which existential illusions are generated. **Bullot**, for example, provides a distinction between two types of agent tracking mechanisms that he believes weighs heavily on the theoretical integrity of an evolved folk psychology of souls. In *perceptual tracking*, Bullot reasons, "a target individual is directly tracked by a sensory-motor system." In contrast, *epistemic tracking* occurs when an agent "is spatio-temporally pursued by indirect epistemic means such as communication and reasoning."

According to Bullot, because both types of tracking require empirical or material traces of the agents' ontological existence, such as behavioural residue (e.g., fingerprints), the present case of souls, gods, and ghosts poses an important problem for the evolutionary model. Bullot reasons that ghosts and gods should therefore be characterized as *fictionally-grounded* referents, "which rest on descriptive resources and individual/collective imagination," rather than *empirically-grounded* referents. Presumably this would be evidence of the cultural origins of supernatural agents. I have argued, however, that our species has an innate predisposition to see natural events as *actual* empirical traces of *real* supernatural agents. It follows from this that natural events serve the same function as tractable social behaviours, activating similar epistemic tracking mechanisms. An infinite array of life's vicissitudes strewn throughout personal biographies is represented as God's "behaviours." It is unclear to me how Sherlock Holmes, a fictional character that Bullot compares to gods and ghosts, could leave similar empirical and

perceptible traces that are capable of confirming peoples' intuitive hypotheses of his ontological existence.

R5.1. Individuation=Ensoulment

In a related commentary, **Newman at el** describe their fascinating research programme on identity tracking, concluding that my argument for a specialized cognitive system dedicated to reasoning about souls lacks parsimony. This conclusion hinges on a series of studies revealing that participants believe in the continued psychological existence of not only individual people after dramatic transformations (such as someone who has died or whose memories are placed in a robotic body), but also the continued identity of individual objects that have undergone similarly dramatic transformations (e.g. a specific car, "Rustbucket" is still identified as "Rustbucket" after going through a "particle pipeline" and reconfigured into a boat). "Thus," argue Newman et al, "belief in the persistence of individuals through radical changes in kind is not restricted to persons and need not include the notion of a soul."

I interpret these findings very differently from these commentators, however, and see them as generally supportive of my evolutionary model rather than as falsifying my hypotheses about the folk psychology of souls (see also **Nichols**). Newman et al miss the critical fact that by individuating the target objects in this manner (e.g., through proper naming) they may be doing something akin to ensouling inanimate objects. Through the experimental individuation of objects, participants may be reasoning about such objects through an animistic lens. As a consequence of this, they are likely to tacitly endow these objects with psychological states, in effect viewing them as possessing souls. **Bloch's** sardonic comments about the ecological validity of the mouse puppet show paradigm (i.e. Bering & Bjorklund 2004; Bering, Hernandez-Blasí, &

Bjorklund 2005) similarly dismisses the animating effects of individuating target "characters" in an experimental context.

R5.2. The simulation constraint hypothesis

Several commentaries focussed on the simulation constraint hypothesis (**Antony**; **Cohen & Consoli**; **Jack & Robbins; Kemmerer & Gupta; Preston et al**). To revisit the central thesis of this hypothesis, I claimed that a delimiting phenomenological boundary prevents people from experiencing the absence of certain categories of mental states, such as emotions, desires, and various episteme (the most "ethereal" qualia). Because we can never know what it feels like to be without such states, these natural representational borders encourage afterlife beliefs. When we attempt to reason about what it will be "like" after death—and what it is "like" for those who have already died—we inevitably get ensnared by simulation constraints and reason in terms of a continued consciousness.

Preston et al reason that belief in the soul stems in large part from the illusion of conscious will, the feeling that the self is a sort of abstract homunculus that consciously wills the body to act (when in fact this feeling of authorship of our own actions is epiphenomenal). I agree that this is an important component of the folk psychology of souls, but I see it as a refinement to the present model rather than an alternative account. **Cohen & Consoli's** description of Cotard's syndrome as being characterized by the delusion that one is already dead, as well as **Kemmerer & Gupta's** discussion of the neurobiological basis of out-of-body-experiences, do seem to provide at least indirect support for the simulation constraint hypothesis.

In his thoughtful commentary, **Antony** reasons that in order to "run a simulation" of a dead person's mind, one must already have a belief in the afterlife, which runs contrary to this

nativist position on the illusion of immortality. "Prior to simulating a dead agent's mind," he argues, "it must be assumed there is a mind to simulate. But that *already* is to assume an afterlife." He then writes, "It follows that nothing about a simulation itself can explain our belief in an afterlife, since some such a belief or assumption is a precondition for the planning and running of any such simulation." There is probably some truth to Antony's chicken-and-egg reasoning, but it is unclear to me why he sees this as a problem for my nativist arguments concerning the origins of afterlife beliefs. If children are confronted with someone who has died, they need not "assume an afterlife"—at least in any meaningful conceptual sense, and certainly not in terms of a propositional belief about an afterlife—to attempt to reason about the dead person's current state of mind through appeal to their own mental states.

R8. Concluding remarks

The commentaries in response to the target article are all that a *Behavioural and Brain Sciences* author could wish for. They are filled with incisive criticisms, counterarguments, and references to important work of which I was unaware. All of these undoubtedly will add to a more informed cognitive analysis of the subtle strands that bind together morality, souls and meaning.

NOTES

1.   This is the literary device behind Nathaniel Hawthorne's *The Scarlet Letter*. Hester Prynne, an otherwise virtuous woman, is publicly labelled an "adulteress" (literally, with a capital "A") and shunned by her small Puritan community.

2.   In Sartre's famous play "No Exit" (1946/1989), in which three strangers find themselves uncomfortably together in a drawing-room of Hell, there are no mirrors or windows in the room, sleep is not permitted, and the light is always on. The characters' eyelids are paralyzed, disallowing them even the luxury of blinking. One of the characters, Garcin, reacts with muted horror to the prospect of being constantly observed by the others for all eternity. He is also convinced that he is under surveillance by demons, "…*all those eyes intent on me. Devouring me*." It is not hard to see why this would be such an exquisite torture.

3.   Interestingly, after a long historical period where people may have been able to emigrate to new social groups and to "start over" if they spoilt their reputation, the present media age in some ways more accurately reflects the conditions faced by our ancestors. With newspapers, telephones, cameras, television, and the Internet at our disposal, personal details about medical problems, spending activities, criminal and financial history, and divorce records (to name just a few potentially sundry tidbits), are not only permanently archived, but can be distributed in microseconds to, literally, billions of other people. The old adage "wherever you go, there you are" takes on new meaning in light of the evolution of information technology. The Internet, in particular, is an active microcosm of human sociality that has not yet been properly analyzed in Darwinian terms. From background checks to matchmaking services, from anonymous web site browsing to piracy and identity theft, from "Googling" ourselves and peers to flaming bad professors (e.g. www.ratemyprofessor.com) and stingy customers (e.g. www.bitterwaitress.com), the Internet is ancient social psychology meeting new information technology.

Acknowledgements

References

Astuti, R. & Harris, P. L. (2006) Understanding mortality and the life of the ancestors in rural Madagascar. Paper submitted for publication.

Atran, S. & Norenzayan, A. (2002). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and Brain Sciences* 27:713-770.

Barrett, H. C. & Behne, T. (2005). Children's understanding of death as the cessation of agency: A test using sleep versus death. *Cognition* 96:93-108.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences* 4: 29-34.

Barrett, J. L. & Keil, F. (1996). Conceptualizing a nonnatural entity: Anthropomorphism in God concepts. *Cognitive Psychology* 31:219-247.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters* (doi:10.1098/rsbl.2006.0509).

Bering, J. M. (2002a). The existential theory of mind. *Review of General Psychology* 6:3-24.

Bering, J. M. (2002b). Intuitive conceptions of dead agents' minds: The natural foundations of afterlife beliefs as phenomenological boundary. *Journal of Cognition and Culture* 24:263-308.

Bering, J. M. & Bjorklund, D.F. (2004). The natural emergence of reasoning about the afterlife as a developmental regularity. *Developmental Psychology* 40:217-233.

Bering, J. M. & Johnson, D. D. P. (2005). "O Lord . . . you perceive my thoughts from afar": Recursiveness and the evolution of supernatural agency. *Journal of Cognition and Culture* 5:118-142.

Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought*. Basic Books.

Burnham, T. & Hare, B. (in press). Engineering human cooperation: Does involuntary neural activation increase public goods contributions in human adults? *Human Nature*.

Colwell, J., Schröder, & Sladen, D. (2000). The ability to detect unseen staring: A literature review and empirical tests. *British Journal of Psychology* 91:71-85.

Diener, E., Fraser, S., Beaman, A., & Kelman, R. (1976). Effects of deindividuation variables on stealing among Halloween trick-or-treaters. *Journal of Personality & Social Psychology* 33:178-183.

Dunbar, R. I. M. & Spoors, M. (1995). Social networks, support cliques, and kinship. *Human Nature* 6:273-290.

Ellison, P., Govern, J., Petri, H., & Figler, M. (1995). Anonymity and aggressive driving behaviour: A field study. *Journal of Social Behavior & Personality* 10:265-272.

Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Simon & Schuster, Inc.

McAdams, D. P. (2002). The psychology of life stories. *Review of General Psychology 5*: 120-122.

Poling, D. A. & Evans, E. M. (2004). Are dinosaurs the rule or the exception? Developing concepts of death and extinction. *Cognitive Development* 19: 363-383.

Fehr, E. & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior* 25:63-87.

Fessler, D. M. T. & Haley, K. J. (2003). The strategy of affect: Emotions in human cooperation. In: *The genetic and cultural evolution of cooperation*, ed. P. Hammerstein, MIT Press.

Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of deindividuation in a group. *Journal of Abnormal & Social Psychology* 47:382-389.

Gilbert, P. & McGuire, M. T. (1998). Shame, status, and social roles: Psychobiology and evolution. In: *Shame, interpersonal behavior, psychopathology, and culture*, ed. P.

Gilbert and B. Andrews. Oxford University Press.

Gold, G. J. & Weiner, B. (2000). Remorse, confession, group identity, and expectancies about repeating a transgression. *Basic and Applied Social Psychology,* 22: 291-300.

Haidt, J., Björklund, F., & Murphy, S. (2004). Moral dumbfounding: When intuition finds no reason. (Unpublished manuscript, University of Virginia).

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior* 26:245-256.

Harris, P. L., Brown, E., Marriott, C., Whittall, S., & Harmer, S. (1991). Monsters, ghosts and witches: Testing the limits of the fantasy-reality distinction in young children. *British Journal of Developmental Psychology* 9:105-123.

Harris, P.L. & Giménez, M. (2005). Children's acceptance of conflicting testimony: The case of death. *Journal of Cognition and Culture* 5:143-164.

Johnson, D. D. P. (2005). God's punishment and public goods: A test of the supernatural punishment hypothesis in 186 world cultures. *Human Nature* 16:410-446.

Johnson, D. D. P. & Bering, J. M. (in press). Hand of God, mind of man: Punishment and cognition in the evolution of cooperation. *Evolutionary Psychology*.

May, H. (2000). "Murderers' relatives": Managing stigma, negotiating identity. *Journal of Contemporary Ethnography* 29:198-221.

Mealey, L. (1995). The sociobiology of sociopathology: An integrated evolutionary model. *Behavioral and Brain Sciences* 20:523-599.

Milinski, M., Semmann, D., & Krambeck, H. J. (2002). Donors to charity gain in both indirect reciprocity and political reputation. *Proc. R. Soc. B.* 269: 881-883.

Nemeroff, C. & Rozin, P. (1994). The contagion concept in adult thinking in the United States: Transmission of germs and interpersonal influence. *Ethos* 22:158-186.

Nowak, M.A. and K. Sigmund. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393:573-577.

Reynolds, V., & Tanner, R. (1995). *The social ecology of religion*. Oxford University Press.

Rehm, J., Steinleitner, M., & Lilli, W. (1987). Wearing uniforms and aggression: A field

      experiment. *European Journal of Social Psychology* 17:357-360.

Sartre, J-P. (1946/1989). *No exit and other plays*. Vintage International.

Silke, A. (2003). Deindividuation, anonymity and violence: Findings from Northern Ireland. The

      Journal of Social Psychology 143:493-499.

Subbotsky, E. (2001). Causal explanations of events by children and adults: Can alternative

      causal modes coexist in one mind? *British Journal of Developmental Psychology*

      19:23-45.Tomasello M. & Call, J. (1997). *Primate cognition*. Oxford University Press.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states—

      the question is which ones and to what extent? *Trends in Cognitive Sciences* 7:153-

      156.

Tooby, J. & Cosmides, L. (1992). The psychological foundations of culture. In: *The adapted*

      mind: Evolutionary psychology and the generation of culture, ed. L. Cosmides, J.

      Tooby, & J. H. Barkow. Oxford University Press.

Watson, R. I. (1973). Investigation into deindividuation using a cross-cultural survey technique.

      Journal of Personality & Social Psychology 25:342-345.

Wedekind, C. & Braithwaite, V. A. (2002). The long-term benefits of human generosity in

      indirect reciprocity. *Current Biology* 12: 1012-1015.

Weeks, M. & Lupfer, M. B. (2000). Religious attributions and proximity of influence: An

      investigation of direct interventions and distal explanations. *Journal for the Scientific*

      Study of Religion 39:348-362.

Whitehouse, H. (2004). *Modes of religiosity: A cognitive theory of religious transmission*.

      Altamira Press.

Zimbardo, P. G. (1969). The human choice: Individuation, reason and order, versus

      deindividuation, impulse and chaos. In: *Nebraska Symposium on Motivation: Vol 17*, ed.

W. J. Arnold & D. Levine. University of Nebraska Press.