Theory of Mind, Religiosity, and Autistic Spectrum Disorder: A Review of Empirical Evidence Bearing on Three Hypotheses

Robert N. McCauley

Center for Mind, Brain, and Culture

Emory University

philrnm@emory.edu

and

George Graham

(formerly Professor of Philosophy at

University of Alabama at Birmingham,

Georgia State University,

and

A.C. Reid Professor at Wake Forest University)

grahamg@copper.net

Key Words: By-Product Theory, theory of mind, content bias, Autistic Spectrum Disorder, religiosity, mind-blind atheism

250 Word Abstract

The cognitive science of religions' By-Product Theory contends that much religious thought and behavior can be explained in terms of the cultural activation of maturationally natural cognitive systems. Those systems address fundamental problems of human survival, encompassing such capacities as hazard precautions, agency detection, language processing, and theory of mind. Across cultures they typically arise effortlessly and unconsciously during early childhood. They are not taught and appear independent of general intelligence. Theory of mind (mentalizing) undergirds an instantaneous and automatic *intuitive understanding* of minds, mental representations, and their implications for agents' actions. By-Product theorists hypothesize about a social cognition content bias, holding that mentalizing capacities inform participants' implicit understanding of religious representations of agents with counter-intuitive properties. That hypothesis, in combination with Baron-Cohen's account of Autistic Spectrum Disorder (ASD) in terms of diminished theory of mind capacities (what he calls "mind-blindness"), suggests an impaired religious understanding hypothesis. It proposes that people with ASD have substantial limitations in intuitive understanding of and creative inferences from such representations. Norenzayan argues for a mind-blind atheism hypothesis, which asserts that the truth of these first two hypotheses suggests that people with ASD have an increased probability, compared to the general population, of being atheists. Numerous empirical studies have explored these three hypotheses' merits. After carefully pondering distinctions between intuitive versus reflective mentalizing and between explicit versus implicit measures and

affective versus cognitive measures of mentalizing, the available empirical evidence provides substantial support for the first two hypotheses and non-trivial support for the third.

150 Word Abstract

By-Product Theory contends that much about religions can be explained via cultural activation of maturationally natural cognitive systems, which address problems of survival. Theory of mind (mentalizing) supports an intuitive understanding of minds, mental representations, and their implications for action. By-Product theorists defend a social cognition content bias hypothesis, holding that mentalizing capacities inform implicit understanding of agents. That hypothesis, in combination with Baron-Cohen's account of Autistic Spectrum Disorder (ASD) in terms of diminished theory of mind capacities, suggests an impaired religious understanding hypothesis, proposing that people with ASD have limitations in intuitive understanding and creative inference regarding such representations. Norenzayan's mind-blind atheism hypothesis holds that these two hypotheses' truth suggests that people with ASD are more likely to be atheists. After pondering intuitive versus reflective mentalizing, explicit versus implicit measures, and affective versus cognitive measures, the evidence supports the first two hypotheses and provides non-trivial support for the third.

Current Word Counts:
overall file: 8898 words
text of article and references minus upfront material: 8397
text of article minus references and upfront material: 7311

Introduction

Contrary to a popular English idiom, exceptions do not prove rules, unless the exception is explicable on the basis of the rules, i.e., unless the principles the rules embody *explain why the exception is exceptional.* In this case the rules in question arise from the theoretical assumptions of the By-Product Theory in the cognitive science of religions (CSR) (Boyer, 2001; Barrett, 2004; McCauley, 2011), which hold that successful religious representations and, correspondingly, substantial swaths of religious thought and behavior are grounded in humans' maturationally natural cognitive capacities. The maturationally natural cognitive systems in question are concerned with a host of tasks that are critical to survival and that have nothing to do with religions. Thus the theory's name – it frames much about religions in terms of cognitive *by-products*, elicited by cultural representations that cue these systems' operations. These systems include such things as hazard precautions, agency detection, language processing, and theory of mind.

Theory of mind, as a distinct object of study and field of research in psychology, concerns the development, structure, and functioning of human cognitive capacities for attributing, recognizing, and drawing inferences about minds, their contents, and those contents' impact on behavior in others (and in the self), which is sometimes referred to as "mentalizing." The exception of interest to the By-Product Theory's account of much of religious cognition concerns the deficits in theory of mind abilities that are characteristic of Autistic Spectrum Disorder (ASD). During the preschool years, most children with ASD do not even detect, let alone read, others' minds. That is not at all startling to learn about a substantial fraction of that population, who never learn to talk. It is not startling, because many

elementary *social* accomplishments in the first year of life, such as infants' abilities to engage in *joint attention* with caregivers, appear to be critical to the acquisition of language (Tomasello, 1999).

Various theorists have argued that cognitive impairments in theory of mind, associated with ASD, should lead to some abnormalities in religiosity among this population (McCauley, 2000; Atran & Norenzayan, 2004; Bloom, 2007; McCauley & Graham 2020). They are suggesting that, at the very least, we should expect the population of people with ASD to prove exceptional on some fronts with regard to some salient dimensions of religious cognition. If that is, in fact, true, then this *is* the sort of exception that does "prove" the rule. (We employ scare quotes around the idiom here, because, technically, *proof* is not available in empirical matters.) The point is that negative findings about the religiosity of people with ASD *would* provide support for the By-Product Theory.

Three Hypotheses: From Barriers to Intuitive Understanding and Inference to Mind-Blind Atheism

Different theorists have proposed impairments among the ASD population that stretch from conjectures about cognitive barriers to intuitive understanding and inference (McCauley, 2011) all the way to mind-blind atheism (Norenzayan, 2013). Space limitations preclude the consideration of every possibility here. In what follows, we shall confine ourselves to the examination of the evidence for these two positions.

Ara Norenzayan, Will Gervais, and Kali Trzesniewski (2012; Norenzayan, 2013) have proposed that the impediments with regard to theory of mind will *reduce religious belief* among people with ASD. Norenzayan and colleagues argue that such deficits in theory of mind preclude understanding the meaning and implications of claims about the gods, about what they think, and about what they do, and that that tends to lessen the probabilities that people with ASD will be religious. Their argument is straightforward: "if mentalizing supports the mental representation of supernatural agents, then

mentalizing deficits associated with the autistic spectrum . . . may undermine intuitive support for supernatural agent concepts and reduce belief in God . . . " The authors conjecture that "mentalizing might be a necessary component of belief in God" (Norenzayan et al., 2012, p. 1). Subsequently, in his book, *Big Gods,* Norenzayan argues that at least four different combinations of psychological, intellectual, and social dynamics can lead to atheism. The first of the four he characterizes as "mind-blind atheism" (Norenzayan, 2013, pp. 177-180). Mind-blind atheism holds that some atheists, viz., many people with ASD as well as many people among the general population, reject belief in God, at least in part, because of their substantially or more moderately diminished theory of mind capacities (respectively). Crucially, one of the consequences of the conception of autism as a *spectrum* is that at its high functioning end it fades into a region of the larger spectrum of the general population with regard to theory of mind capacities that includes people with sub-clinical theory of mind limitations.

Norenzayan cites additional forms of evidence for mind-blind atheism, including empirical studies of his own to which we shall turn in the next section. For now, though, we want to illuminate the logical and evidential relations between three hypotheses in play. CSR's By-Product Theory appropriates theory of mind as follows:

(1) *social cognition content bias hypothesis:* theory of mind capacities and operations play a central role in characterizing: (a) religious representations of intentional agents with counter-intuitive properties, (b) those representations' cognitive appeal, and (c) their (intuitively available) inferential potential.

(2) *impaired religious understanding hypothesis*: deficits in theory of mind, especially those of people with ASD, impose substantial limitations on their intuitive understanding of and creative inferences from religious representations about intentional agents with counter-intuitive properties.

(3) *mind-blind atheism hypothesis*: the substantial limitations on their intuitive religious understanding and inference of people with ASD decrease the probabilities that they will be religious and increase the probabilities that they will be atheists (relative to the general population).

The impaired religious understanding hypothesis is virtually a deductively valid consequence of the combination of Simon Baron-Cohen's (1995) mindblindness hypothesis about ASD and the social cognition content bias hypothesis above. Baron-Cohen accords theory of mind a pivotal role in making sense of many features of autism, proposing that people with autism are "mindblind," i.e., that for them theory of mind abilities, if they emerge at all, do not emerge unconsciously and comparatively effortlessly in the typical developmental timeframe, but, instead, are assiduously constructed later in life as the result of an inductive exercise applying their systemizing abilities to the limited social transactions they have experienced personally, resulting in an ersatz theory of mind.

The social cognition content bias hypothesis (1) and the impaired religious understanding hypothesis (2) provide theoretical support for the mind-blind atheism hypothesis (3) as follows. Even if people with autism find religious representations of agents with counter-intuitive properties attention grabbing and memorable (which is not obvious), they, unlike the general population, are substantially restricted in the inferences they are able to readily draw from such representations, at least until they have spent years constructing an ersatz theory of mind on their own. Without access to religious representations' associated collection of intuitive default inferences (about minds), which others possess, people with autism are less likely, ultimately, to find such religious representations cognitively appealing and, thus, are less likely to find them attractive targets for their affiliation. The mind-blind atheism hypothesis is a thoroughly plausible consequence of the combination of (1) and (2), but their truth only makes the truth of (3) probable. Those two hypotheses do not guarantee (3)'s truth for three reasons. The first two reasons are closely related and have to do with *motivational* matters.

The first reason concerns limits on the impact of the relevant content bias on *positive* motivation regarding religious commitment. The mind-blind atheism hypothesis assumes (from (1)) that religious representations' cognitive appeal tends to increase those representations' attractiveness as targets for commitment and affiliation. Even if that assumption is true, though, Gervais and Joseph Henrich (2010) argue correctly that this cognitive appeal does not suffice to explain people's *particular* religious commitments, since it is unable to explain why religious people only find some subset of all of the available religious representations appealing enough to merit their commitment. It is insufficient by itself to explain, for example, why contemporary religious people do not (also) believe in Zeus. Gervais and Henrich argue that cultural support and influence, rooted in *context biases* in human minds that make people especially sensitive to the attitudes, conduct, and properties of *prestigious* people in their group, also play a role in explaining commitment and affiliation (Henrich & Gil-White, 2001).

The second reason concerns limits on the impact of the relevant obstacles to understanding and inference on *negative* motivation regarding religious commitment. The mind-blind atheism hypothesis assumes (from (2)) that the obstacles to intuitive understanding and inference tend to decrease religious representations' attractiveness as targets of commitment and affiliation -- even, possibly, to the point of atheism. These obstacles, however, do not *suffice* to explain either diminished religious commitment or outright atheism on the part of some people with ASD, since they are also present in the cases of religious involvement of some people with ASD (Brezis, 2012). These obstacles are insufficient, by themselves, to explain either the presence or absence of religious involvements or the atheism of people with ASD without a parallel appeal, again, to cultural considerations.

Both reasons highlight (i) that religious representations' cognitive appeal or lack thereof affect religious commitment and affiliation (yea or nay), but (ii) that by themselves they are insufficient to explain those verdicts without invoking cultural and, possibly, intellectual influences as well. The third reason, by contrast, has to do with the relationship of implicit and explicit cognition.

Hypotheses (1) and (2) principally address matters of *implicit* cognition. Hypothesis (3), by contrast, is primarily concerned with *explicit* cognition. The reasons and causes for why people explicitly affirm particular claims are many and varied. That is probably as or more true about religious claims than any other class of claims. That impaired implicit religious understanding is likely to increase the probability that someone will claim to be an atheist or offer what are, in effect, atheistic responses to items aimed at eliciting their views seems plausible enough. But, again, the truth of hypothesis (2) does not guarantee the truth of hypothesis (3). The various findings on theological incorrectness, for example, suggest that this inference is, by no means, automatic (Barrett & Keil, 1996).

The aim of these analytical comments is to clarify that in the studies that we review in the following sections the evidence *for* the mind-blind atheism hypothesis, since it bears on a stronger claim about both motivation and explicit cognition (compared to hypotheses (1) and (2)), furnishes, at the very least, indirect evidence for those first two hypotheses as well**.** By contrast, studies that have yielded evidence that is *contrary* to the mind-blind atheism hypothesis (3) do *not* necessarily count against either hypothesis (1) or (2), since, as we noted above, (3) makes different claims (about *explicit* cognition) and stronger claims (about the provocation of outright atheism) than (1) and (2).

Empirical Evidence Corroborating the Three Hypotheses

Catherine Caldwell-Harris and her colleagues (2011) examined these topics before Norenzayan explicitly advanced the mind-blind atheism hypothesis (3). Although their studies, in fact, pertain to our three hypotheses, the Caldwell-Harris group described their research as "exploratory, rather than hypothesis-driven" in recognition of the conceptual and causal complexities that stand behind the notions of religiosity, high functioning individuals with autism (HFA), and their interactions.

The Caldwell-Harris group carried out two studies, the first of which used independent coders to analyze nearly two hundred consecutive posts concerned with religion (per a set of coding protocols) from each of two web-based discussion groups, wrongplanet.net and golivewire.com/teen, which had forums entitled "Religion/Philosophy/Politics" and "Religion and Philosophy," respectively. The first is a site for HFA whereas the participants with the second are an overwhelmingly neuro-typical population. Coders were blind to the origins of the posts that they coded and exhibited inter-rater reliability concerning statements about religious beliefs of 93% (Caldwell-Harris et al., 2011, p. 3363). The two groups exhibited significant differences with regard to religious beliefs with the HFA participants less likely to identify as members of organized religions and more likely to be categorized by the coders as either atheists or agnostics (Caldwell-Harris et al., 2011, p. 3364).

In their second study the Caldwell-Harris team recruited two groups, one HFA and the other neuro-typical, provided them the opportunity to volunteer their diagnostic information, and had them complete a variety of standardized questionnaires that assess the respondents' probabilities for qualifying as autistic via the Autism Spectrum Quotient (ASQ), their mentalizing capacities via the Reading the Mind in the Eyes Test, their penchant for systemizing via the Systemizing Quotient (SQ) as well as a set of questions about their religious orientation and beliefs.

ASQ, Reading the Mind in the Eyes, and SQ as well as the Empathy Quotient (EQ), which is employed in many of the subsequent studies we discuss below, are all tools, designed by Baron-Cohen and his colleagues for use with both clinical populations and the general population. See Baron-Cohen et al, 2001a, Baron-Cohen et al., 2001b, and Wakabayashi et al., 2006, respectively. Unlike the others, which are essentially questionnaires, the Reading the Mind in the Eyes test makes participants choose from among four candidate mental states (e.g., hateful, jealous, arrogant, panicked) to describe a photograph of a thin horizontal slice of people's faces, basically showing their eyes only.

Although employing a decidedly different design, this second study yielded similar findings to the first. Members of the HFA group were significantly less likely than the members of the neuro-typical group "to belong to an organized religion" and "were more likely to be atheist . . . " (Caldwell-Harris et al., 2011, p. 3365). Interestingly, within *both* the HFA and neuro-typical groups, ASQ scores were significantly higher for the atheists than for those identified as Christians or Jews.

Studies by Norenzayan and his colleagues (Norenzayan et al., 2012) and other researchers (e.g., Lindeman et al., 2015; Wlodarski & Pearce, 2016) go further, employing statistical analyses (path analysis, in particular) that quantify the influence of the exercise of theory of mind capacities on the findings that they obtain. Norenzayan and his colleagues' findings suggest that the effect of theory of mind abilities on both intuitive religious understanding and belief in God consistently arises and is of moderate size.

The Norenzayan group's four studies (Norenzayan et al., 2012) probably offer the best empirical evidence that is currently available for the three hypotheses and their connection. Their Study 1 stands apart from the other three on two fronts. First, it compares a matched control sample with a test sample of adolescents, who have clinical diagnoses of autism (with no other diagnoses), but, second, (consequently) its total number of participants (twenty-five in all, including twelve diagnosed with autism) is much smaller than in the other three studies. Participants in Study 1 were tested on an Intuitive Belief in God scale, which was highly correlated in earlier research with other well-known scales measuring religious devotion. The researchers obtained, in addition, Intelligence Quotient (IQ), ASQ, and EQ scores for the participants and explored the influence of IQ and participants' proclivities for mentalizing as independent predictors of belief in God in a logistical regression model. In Study 1 IQ proved to be unrelated to religious belief and uncorrelated with participants' inclinations to mentalize. (This corroborates the By-Product theorists' contention that problems with theory of mind are domain-specific and have little or no connection with overall intelligence.) The neuro-typical participants were

more than nine times more likely than the autistic participants to express a strong belief in God. In the logistic regression only mentalizing was a significant predictor of belief in God, and "for each standard deviation decrease in mentalizing, participants were only 21% as likely to strongly endorse God" (Norenzayan et al., 2012, p. 2).

Their other three studies used substantially larger numbers of participants from three samples (numbering 327, 725, and 452, respectively) of the general population -- a Canadian student sample in Study 2 and two samples of American adults recruited via Amazon's Mechanical Turk in Study 3 and from an on-line survey company in Study 4. Although none of these samples is representative, the last two were diverse on multiple fronts (ethnicity, religious affiliation and attendance, educational attainment, etc.) in proportions that roughly approximated the American population at large. Unlike Study 1, in which theory of mind abilities were treated as a discrete variable (autism diagnosis or not), these three studies treated these abilities as a continuous variable, based on participants' performance on ASQ measures. The dependent measure in each case was participants' performance on either the same Intuitive Belief in God Scale employed in Study 1 (in Studies 2 and 4) or the Spiritual Well-Being Scale (in Study 3) (Paloutzian & Ellison, 1991). EQ, as a separate measure of mindreading, was examined as a possible mediating variable in all three studies. Additional potential mediating variables explored were participants' proclivities for systemizing, assessed via their SQ scores, in Study 2, participants' conscientiousness and agreeableness, which are the two personality traits, among the so-called big five, that have proved to be related to religiosity, in Study 3 (Saroglou, 2002 and 2010), and a further measure of their mindreading abilities, assessed via their Reading the Mind in the Eyes scores, in Study 4. Study 3 contained measures of a collection of control variables that included "age, educational attainment, income level, and frequency of religious attendance" (Norenzayan et al., 2012, p. 7), and Study 4 controlled for age, education, frequency of religious attendance, and interest in math, science, and engineering.

The results of the three studies reflect a consistent pattern. The findings of Study 2 are straightforward and corroborate the three hypotheses. Participants with higher ASQ scores were significantly less likely to believe in God. The effect was significantly mediated by their (diminished) theory of mind abilities, as measured by their EQ scores, while their inclinations to systemize, as measured by their SQ scores, were *not* a significant mediating variable. The latter finding would appear to count as evidence *against* any proposal that entertained the possibility that enhanced systemizing capacities might explain the ability of some people with ASD, across a couple decades of social experience, to diligently assemble an ersatz theory of mind that would serve from a religious standpoint. The question remains, especially in light of these findings, whether such an ersatz theory of mind would be *sufficient* to support some inferential proficiency with regard to religions' representations of the gods' minds.

Study 3 replicates and extends the findings of Study 2 with a larger and more diverse sample of American participants. In a logistic regression model with all of the aforementioned controls ASQ scores significantly predicted religious belief, and, as in Study 2, mentalizing significantly mediated this relationship, whereas again systemizing failed to do so. In a logistic regression model that included *all* of the variables under scrutiny in Study 3, mentalizing proved to be "a specific, independent, and robust predictor of belief . . . " along with older age and frequency of religious attendance (Norenzayan et al., 2012, p. 2).

That frequency of attendance has an impact on explicit avowals of religious belief is not surprising. Just how influential it is, compared with facility at mentalizing, inevitably complicates assessments of findings about people with ASD who regularly attend religious services and of how those findings bear on the three hypotheses. Consequently, Study 4 included controls on religious attendance, and it produced similar results. When controlling all four covariates (age, education, religious attendance, and interest in math, science, and engineering) ASQ scores again significantly predicted

religious belief. Here both measures of theory of mind abilities, viz., EQ and the Reading the Mind in the Eyes test, mediated the relationship between ASD and religious belief, and both did so independently. (Other studies, however, have failed to replicate the result with the Reading the Mind in the Eyes test (Caldwell-Harris et al., 2011; Jack et al., 2016; Lindeman et al., 2015; Vonk & Pitzen, 2016). The Reading the Mind in the Eyes test is concerned with the *accuracy* of mindreading, as opposed to the basic maturationally natural inclination to mentalize (Vonk & Pitzen, 2016), which is the salient issue in the hypotheses of interest here.) As with Study 3, religious attendance also predicted belief in God. Lower levels of education did so as well.

On the one hand, these findings are insufficient to establish causal relations, but, on the other hand, they stood up to a variety of methodological checks and controls. The studies replicated the central findings suggesting that deficits in theory of mind capacities not only interfere with individuals' intuitive understanding of religious representations of agents with counter-intuitive properties but also substantially diminish the probabilities of their believing in them. Findings in other studies show anywhere from a modest to a moderate effect on religious belief of facility with theory of mind as well (e.g., Willard & Norenzayan, 2013).

Further Evidence Bearing on the Three Hypotheses

Ever growing interest in ASD, in CSR, in these hypotheses, and in related hypotheses about the evolution of religions has inspired further empirical research, especially once Norenzayan and his colleagues' paper appeared. At first glance this additional research may appear less encouraging, as few obtain findings either as univocal or as potent as those in the Caldwell-Harris or the Norenzayan groups' studies. That said, in light of some conceptual clarifications that we will take up in what follows and

appearances sometimes to the contrary notwithstanding, the findings are, arguably, supportive overall, certainly of hypothesis (1) and (2), but even of hypothesis (3).

Among these more recent studies, Rafael Wlodarski and Eiluned Pearce's (2016) research offers the most direct support of hypothesis (1). Wlodarski and Pearce enlisted 298 adult American participants via Mechanical Turk and had them complete a series of questionnaires, which measured, among other things, both theory of mind and religiosity. They used a short version of EQ for assessing mentalizing capacity and the Santa Clara (Plante et al., 2002) and Duke University (Koenig & Bussing, 2010) measures of religiosity. Their findings were unambiguous. Mentalizing, as measured by EQ, was a significant predictor of religiosity, and its predictive value was independent of the other variables that they explored in their study.

Probably the study yielding the least congenial findings for the three hypotheses is Leif Ekblad and Lluis Oviedo's first experiment surveying 2138 volunteers on the Aspie Quiz website – a site where people can take an on-line survey to check whether they likely qualify as "neurodiverse" (ND), which includes ASD but many other cognitive disorders as well, such as Attention Deficit Hyperactivity Disorder and Dyslexia (2017, p. 288). On most of the factors that they examined, Ekblad and Oviedo found that their ND participants tended to be slightly but significantly *more religious* than their participants who scored as neuro-typical, contrary to hypotheses (2) and (3) and, perhaps, to hypothesis (1) as well. Surprisingly, though, the authors offer comments throughout the paper that are certainly consistent with hypothesis (2).[1] They suggest that their ND participants are far more likely than their neuro-typical participants to "suffer difficulties in adapting to the standard religious socialization," that they acquire their religious representations primarily through "social learning," and that they develop "preferred private religious beliefs" (2017, pp. 294 and 295). Still, they include no talk of any constructed ersatz

[1] Given the subtitle of their paper, "Defective or Different?" and opting for *different*, Ekblad and Oviedo would, no doubt, prefer that hypothesis (2) was differently named.

theory of mind, and they explicitly question hypothesis (1), holding that it is not even clear that theory of mind "is related to . . . a belief in God" (2017, p. 288).

In light of the non-representative sample they employ (among other things, people come to the Aspie Quiz website of their own accord), that the score cutoff on the Aspie Quiz "was originally set so that 80% of diagnosed ASD would have their diagnosis confirmed" (Ekblad & Oviedo, 2017, p. 289), and the many studies we have already reviewed with findings that seem to corroborate hypothesis (1), their expressions of extreme skepticism about it seem, perhaps, a bit premature. On the other hand, their suggestion and that of Ingela Visuri (2019) that what religious experiences and understandings people with ASD might have are *different* from those of the neuro-typical population is perfectly consistent with hypothesis (1) and (2).

Multiple studies provide evidence for a modest but positive relationship between theory of mind abilities and religious belief or, conversely, for a relationship between impaired theory of mind and a lack of religious belief, yet the authors are reluctant to characterize their findings that way (e.g., Jack et al., 2016 and Lindeman et al., 2015). Although none of the findings in the papers reviewed in the remainder of this section and in the next constitute replications of either the Caldwell-Harris et al., 2011 or the Norenzayan et al., 2012 or the Wlodarski and Pearce, 2016 findings, mostly they do tilt in the same direction and, we shall argue, the analyses the authors offer are generally consistent with hypothesis (1) and (2). Nonetheless, the authors cast their findings as contrary to hypothesis (3) (when they address it at all) and, in some cases, as revealing variables that eclipse any influence that theory of mind abilities might seem to exert, appearing to render hypotheses (1) and (2) as pertaining to a causal intermediary at most. Crucially, though, those interpretations mostly depend upon trying to treat empathizing and mentalizing as independent of one another – a conceptual maneuver which we find less than convincing for reasons we present below. All of this said, these observations are *not* meant to

downplay the fact that their studies have generated additional intriguing findings that point to further variables that also influence outcomes on these fronts.

Anthony Jack and his colleagues gloss empathizing as "moral concern" and argue that, in fact, it is this variable that stands behind and subsumes any effect on these fronts that might be attributed to theory of mind or its impairment (Jack et al., 2016, p. 13). They carried out eight studies with -- in all but one that looked at sixty-nine undergraduates -- a few hundred (ranging between 155 and 527) participants in each, who were recruited from Amazon's Mechanical Turk. They obtained powerful associations between religious belief and moral concern in four, and in two others they argue that a negative correlation between moral concern and analytic thinking explains the negative relation their findings revealed between the latter and religious belief.

That Jack and his colleagues frame these findings as contrary to positions like hypotheses (2) and (3) is puzzling. Two conceptual points are worth making. First, certainly for our purposes here, the relabeling of empathy as "moral concern" is incidental. The Empathic Concern subscale of the Interpersonal Reactivity Index (Davis, 1983) that Jack and his colleagues employ to obtain their measure of moral concern is, *straightforwardly* and *explicitly, a measure of empathy*!

Second, Jack and his colleagues raise larger theoretical questions about affective dimensions of theory of mind in contrast to its cognitive dimensions, on which, for example, Baron-Cohen focuses. Jack and his colleagues seek to strongly differentiate empathizing (moral concern) as affective from mentalizing as cognitive, and they appeal to that distinction as the basis for what they take to be the negative import of their findings for positions (very much like hypotheses (2) and (3)) that they presume are wedded to an exclusively cognitive conception of mentalizing.

Multiple considerations suggest that the two are unlikely to be easily disentangled. Consider, first, the performance of people with ASD in the Ultimatum Game. In the Ultimatum Game both players

know the rules and the relevant circumstances from the outset. The first player is given a sum of money and is required to make an offer to the second player that can range from offering none of the money to the second player, to offering all of it, or offering any amount in between. The second player can, then, either accept the first player's offer or reject it. If the second player accepts the offer, then the two players keep the sums entailed by the first player's offer. If, on the other hand, the second player rejects the offer, then *neither* player gets to keep any of the money.

About one third of first players with ASD, whether children or adults, offer *nothing* to second players (Henrich et al., 2005). Presumably, they make such unusual (non-)offers, because they are unable to empathize with the second players and cannot imagine how they will react. Lacking empathy for their fellow players (which, incidentally, squares with the findings of the studies under examination in this section), they make proposals in that game that diverge strikingly from those of the general population. A prominent factor in their lack of self-understanding and in their inability to empathize is their mindreading impairments.

The results of Jennifer Vonk and Jerrica Pitzen's (2016) studies provide further support for skepticism about how smoothly the affective and cognitive dimensions of mentalizing can be disentangled when reflecting on their impact on religiosity. They, as we, maintain that theory of mind is a complex collection of capacities and draw a distinction between the basic motivation to use theory of mind as opposed to further questions about accuracy in its use, holding, as do we, that the former is the more important consideration for assessing the hypotheses in question. They argue that empathy and emotional intelligence are better measures of that motivation than "variance in actual ToM [Theory of Mind] capability" (Vonk & Pitzen, 2016, p. 2).

Their two studies asked large samples (437 and 388, respectively) of undergraduates to complete a variety of questionnaires concerning various measures of religiosity and of theory of mind,

including EQ, Reading the Mind in the Eyes, and Emotional Intelligence in the first study and adding four additional theory of mind tests in the second. The findings of their first study offered support for their contention that emotional intelligence, in particular, was the significant theory of mind measure correlated with their measures of emotionally based and intrinsic religiosity. Following up on that finding, they added the Situational Test of Emotion Understanding and the Situational Test of Emotion Management (McCann & Roberts, 2008) in their second study in order to obtain more extensive assessments of participants' emotional intelligence. The results of their second study, however, only complicated the picture, since its more sophisticated measures of emotional intelligence *failed* to provide any significant relationships.

Marjaana Lindeman and her colleagues' work employs both EQ and the Reading the Mind in the Eyes test, which, as a further indication of that entanglement between cognitive and affective dimensions of theory of mind, they describe as getting at "*cognitive empathic* ability" (Lindeman et al., 2015, p. 67, emphasis added; also see Rothstein, 2011). Although, as we noted earlier, they did not replicate the positive findings with the Reading the Mind in the Eyes test that the Norenzayan group obtained, Lindeman and her colleagues, in two separate studies, did replicate the significant positive correlation of empathizing with religiosity (Lindeman et al., 2015; Lindeman & Lipsanan, 2016). Lindeman and Jari Lipsanen (2016, p. 5) remark that "empathizing was positively correlated with religiosity, replicating earlier findings . . . The result supports current theories of religiosity in which understanding other people's minds has been considered the main factor in explaining religiosity." Like Norenzayan (2013), they also spotlight the fact that theory of mind impairments are neither necessary nor sufficient for a person to become an atheist. They too hold that many factors can lead to unbelief or, for that matter, to belief as well. Still, their findings also suggest that theory of mind capacities can play a substantial role in those outcomes.

Paul Reddish and his colleagues (2016) undertook a study that resembles Study 1 in Norenzayan et al., 2012. They compared the mentalizing abilities and religiosity of a small sample (19 participants) of high-functioning adolescents and young adults who had been *diagnosed* with ASD (HFA) (all of whom were twelve years of age or older) and a matched sample of twenty-seven typically developing participants (TD). All of the participants were from Singapore. The Reddish group employed multiple measures of theory of mind abilities and of religiosity. Although they found differences in the same direction that Norenzayan and his colleagues did on both mindreading capacities and religiosity between the two groups, those differences, except with regard to the attractiveness of prayer, were not as dramatic and did not rise to the level of statistical significance.

Their most surprising finding was a slightly negative relationship between mentalizing abilities overall and many of their measures of religiosity. They concede (and we agree) that this otherwise anomalous finding may well be a function of the exclusively *explicit* measures of mentalizing that they employ (Reddish et al., 2016, p. 108). Given that those findings are contrary to the findings in all of the papers reviewed above, some caution may be in order.[2]

Reddish and his colleagues (2016, p. 106) also acknowledge that their study's small sample size may well have introduced biases in their findings that disappear in research like Studies 2, 3, and 4 in Norenzayan et al., 2012, which employ between eight and nineteen times more participants. Still, they stress how minor the differences are that they find between their TD and HFA participants. They comment (2016, p. 106) that "our data seem to suggest that HFA still have the cognitive capabilities to think about and interact with gods. Any impairment in mentalizing they do have does not appear to influence the aspects of religiosity that we measured – apart from attraction to prayer – to any substantial degree." That all seems fair enough, but in addition to the facts that (a) their findings still tilt

[2] They are also contrary to Norenzayan and his colleagues' findings (2012) about the role of theory of mind in explaining the so-called "gender gap" in religiosity.

in the directions that the three hypotheses predict, (b) their finding about attraction to prayer tilts significantly so, and (c) these differences prove statistically significant in studies with larger samples, what their results may disclose is not some debilitating problem with any of the three hypotheses so much as three features of their study that will help to sharpen our view of just what, in fact, is at stake here.

Pondering Intuitive Versus Reflective Mentalizing and Explicit Versus Implicit Measures and Affective Versus Cognitive Measures of Mentalizing

The first feature of interest is that the participants with ASD in the Reddish group's study are *extremely* high functioning on the theory of mind measures they employ. The HFA participants exhibited "milder autism symptomology" than most people with ASD and "comparable IQ and language ability" to the members of the TD sample (Reddish et al., 2016, p. 96). The fact that only four of their HFA participants failed the second order false belief task is probably the best indication of just how high functioning Reddish and his colleagues' HFA participants were.[3]

Of all of the measures of mentalizing they used, including the first and second order false belief tasks, their TD and HFA groups differed significantly on only one, which tested participants' abilities to manage "nonliteral use of language such as sarcasm, jokes, and white lies" (Reddish et al., 2016, p. 108). This inspired the Reddish group's apt observation that the gods rarely indulge in sarcasm, but, on the other hand, they seem to have ignored the fact that the deities routinely traffic in narratives and metaphors. From all of this, the Reddish group concludes that "the mentalizing ability needed for successful interaction with humans in complex social situations . . . does not make a major impact on

[3] This is especially striking in light of the fact that *four of their TD participants* also failed the second order false belief task! (See Miller, 2012.)

people's religious behaviour and cognition" (2016, p. 108). This seems to underestimate the complexity that religious materials can present. Metaphors aside, ponder how much must be kept in mind about minds to make sense of even brief narratives -- such as Rebekah's plotting for Jacob to steal Isaac's blessing intended for Esau (an episode that does not even involve any agents with counter-intuitive properties).

Reddish and his colleagues' judgment that their HFA participants obtain "certain mentalizing capabilities through a different cognitive route" (2016, p. 107) than their TD participants accentuates a critical aspect of the social cognition content bias hypothesis (1) and the impairment of religious understanding hypothesis (2). In short, they concern *intuitive* theory of mind capacities. By contrast the HFA participants depend on *conscious reflection*. They call upon their ersatz theory of mind diligently accumulated over at least one or more decades of social interactions. The key point is that their firsthand testimony and Baron-Cohen's research (e.g., 2003, p. 141) both suggest that people with ASD quickly find this sort of conscious reflective consultation of their ersatz theory of mind, even for routine purposes, wearisome. That suggests that they have little, if any, *intuitive* command of these resources. And *that* suggests that if they eventually acquire any intuitive facility with this look-up table at all, it arises from extensive experience and practice with it. Any intuition they possess in this domain is via practiced naturalness, not via the maturational naturalness that the TD participants enjoy (McCauley, 2011).

That occasions interest in a second feature of their study, which, to their credit, the Reddish team themselves acknowledge (2016, p. 108). Their study, and *all* of the studies reviewed in this and the previous section, employ *explicit* measures of religiosity exclusively. Basically, in this and those other studies participants answer a lot of questionnaires.[4] By contrast, *implicit* experimental measures, from

[4] The Jack et al., 2016 (p. 5) studies employ but a *single* explicit measure, viz., participants' responses to the question: "Do you believe in the existence of either God or a universal spirit?"

reaction time to priming and more, are the preferred means for tapping the operations of humans' intuitive capacities. This is all the more advisable when participants have ASD. The Reddish group cites research indicating that people with ASD, not surprisingly, have problems with precisely the sort of awareness and identification of their own mental states that is often indispensable for capture by explicit measures (Hill et al., 2004; Lind, 2010).

This second consideration is paramount. The hypotheses in CSR that have inspired this research concern maturationally natural cognitive capacities that typically arise effortlessly and unconsciously in the first half dozen years of life across every known cultural setting. They are not taught. They appear to be largely independent of general intelligence. They undergird what is usually an instantaneous and automatic *intuitive understanding* of other minds, their representations, and (many of) those representations' implications for agents' actions. The presumptive limitations that theory of mind deficits introduce concern such intuitive understanding and the spontaneous inferences it informs, whether in standard social interaction or in thoughts about and putative interactions with religious agents possessing counter-intuitive properties.

Still, the Reddish group (2016, p. 108) registers its surprise (in the light of the findings of Caldwell-Harris et al., 2011 and of Study 1 of Norenzayan et al., 2012) that the responses of the HFA and TD participants to their explicit measures turn out as similarly as they do. We are less surprised for at least two reasons.

The first of those two reasons looks back to reservations about Reddish and his colleagues' use of explicit measures. Bering's (2002) review of memoirs of people with ASD does suggest that explicit measures may suffice to gain insight about their putative disaffection with religions. Justin Barrett remarks, however, that "determining how people think about God in real-time ordinary situations only on the basis of the language people use is tricky" (Barrett, 2012, p. 158). As we have emphasized

before, people say things for all sorts of reasons. From one setting to the next, any number of variables may cause someone to make one claim as opposed to another. With religions in particular, people often say fairly similar things because what they say is so carefully scripted and so rigorously policed. (Careful scripting, of course, may well appeal to people with ASD.) Although they are not uncontroversial, implicit measures on balance are probably better means for gaining insight about participants' intuitive understanding in any domain.

The additional reason raises the third feature of interest about the Reddish group's study. Not only were the measures of religiosity that they utilized not implicit, most of them were only tangentially concerned with cognition, let alone with intuitive understanding and inference. Seven of their eight measures -- frequency of praying, attraction to prayer, efficacy of prayer, sense of agency while praying, frequency of attendance at religious services, felt closeness to god, and, arguably, even strength of belief -- focused on either *practices, feelings,* or *feelings about practices*. Their failure to obtain significant differences between their two groups of participants on the one inarguably cognitive measure they employed, viz., anthropomorphism of god traits, reverts back to the fact that this was an *explicit* measure. It tested "the degree to which participants hold *explicit* anthropomorphic concepts of supernatural agents" (Reddish et al., 2016, p. 100, emphasis added). The literature on theological incorrectness (such as Barrett & Keil, 1996, Slone, 2004, and Cohen & Barrett, 2008) shows that participants, in doctrinal religions at least, are *unlikely* to hold *explicit* anthropomorphic representations of their gods. Since the groups did not significantly differ with regard to religious attendance, we might infer that they had roughly equal familiarity with their religions' doctrines, forms, and practices. Grounds exist for skepticism about the reliability of such an explicit measure of anthropomorphism in an advanced society like Singapore, where the literacy rate is above 95%.

The Reddish group's findings comparing a sample of nineteen *particularly able people with ASD,* who were deploying painstakingly acquired ersatz theory of mind skills, with nineteen matched TD

participants, based on *explicit measures only* of what are *mostly practices and feelings* pertaining to religious matters revealed differences between their participants in the predicted direction and, on one measure, a difference that was statistically significant. As the Reddish group submits, those findings hardly constitute compelling support for our three hypotheses. Moreover, that they obtained religious affirmations from participants with ASD, on the face of it, counts as evidence contrary to the mindblind atheism hypothesis (3). Still, in the face of all of the studies reviewed here, neither do their findings with a comparatively small sample size, about what they acknowledge (again, to their credit) is a "narrow slice of religious behaviour and cognition" (2016, p. 109), amount to a refutation of any of those hypotheses either.

Coda

Virtually all of the works reviewed above include passages that underscore the complexity of the central concepts at stake (e.g., Vonk & Pitzen, 2016, p. 2), whether theory of mind, autism, or religiosity. For the purposes of empirical studies, that complexity is only multiplied by the many alternative means available for the measurement of each. Consequently, no matter how clear cut the empirical findings in a dozen or two studies might have proved, they would be unlikely to settle these matters once and for all. On balance, though, the available empirical evidence to this point provides substantial support for the social cognition content bias hypothesis (1) and for the impaired religious understanding hypothesis (2). The evidence pertaining to the mind-blind atheism hypothesis (3) is, perhaps, somewhat less univocal but often supportive as well.

We have argued that it is the relative *absence* among people with ASD of any intuitive familiarity with the minds of the gods that resonates with some of CSR's most fundamental theoretical insights. ASD looks to be the exception that proves the rule.[5]

References

Atran, S., & Norenzayan, A. (2004). Religion's evolutionary landscape: Counter-intuition, commitment, compassion, communion. *Behavioral and Brain Sciences*, Vol. 27, 713-777.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. Cambridge: MIT Press.

Baron-Cohen, S. (2003). *The Essential Difference: Male and Female Brains and the Truth about Autism*. New York: Basic Books.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001a). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, Vol. 42, 241-251.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001b) The autism spectrum quotient (AQ): evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, Vol. 31, 5–17.

Barrett, J. L. (2004). *Why Would Anyone Believe in God?* Walnut Creek, CA: Alta Mira Press.

Barrett, J. L. (2012). *Born Believers: The Science of Children's Religious Belief*. New York: Free Press.

Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a Non-natural Entity: Anthropomorphism in God Concepts. *Cognitive Psychology*, Vol. 31 (3), 219-247.

Bloom, P. (2007). Religion is natural. *Developmental Science*, Vol. 10, 147-151.

Boyer, P. (2001). *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books.

[5] We wish to express our gratitude to Justin Barrett, Paul Reddish, and Radek Kundt for their comments on this paper.

Brezis, R. S. (2012). Autism as a case for neuroanthropology: Delineating the role of theory of mind in religious development. In D. H. Lende & G. Downey (Eds.), *The Encultured Brain: An Introduction to Neuroanthropology*, (pp. 291–314). Cambridge MA: MIT Press.

Caldwell-Harris, C. L., Murphy, C. F., Velazques, T., & McNamara, P. (2011). Religious belief systems of persons with high functioning autism. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society,* Austin, pp. 3362-3366.

Cohen, E., & Barrett, J. L. (2008). Conceptualizing spirit possession: Ethnographic and experimental evidence. *Ethos*, Vol. 36, 245-266.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, Vol. 44 (1), 113-126.

Ekblad, L., & Oviedo, L. (2017). Religious cognition among subjects with autism spectrum disorder (ASD): Defective or different? *Clinical Neuropsychiatry*, Vol. 14 (4), 287-296.

Gervais, W. M., & Henrich, J. (2010). The Zeus Problem: Why Representational Content Biases Cannot Explain Faith in Gods. *Journal of Cognition and Culture*, Vol. 10 (3-4), 383-389.

Lindeman, M., & Lipsanen, J. (2016). Diverse Cognition Profiles of Religious Believers and Nonbelievers. *International Journal for the Psychology of Religion*, Vol. 26 (3), 185-192.

Lindeman, M., Svedholm-Häkkinen, A. M., & Lipsanen, J. (2015). Ontological confusions but not mentalizing abilities predict religious belief, paranormal belief, and belief in supernatural purpose. *Cognition*, Vol. 134, 63-76.

Henrich, J., Boyd, R., Bowles, S., Gintis, H., Fehr, E., Camerer, C., McElreath, R., Gurven, M., Hill, K., Barr, A., Ensminger, J., Tracer, D., Marlow, F., Patton, J., Alvard, M., Gil-White, F., & Henrich, N. (2005). "Economic Man" in Cross-Cultural Perspective: Ethnography and Experiments from 15 Small-Scale Societies. *Behavioral and Brain Sciences*, Vol. 28 (6), 795-855.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred status as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, Vol. 22 (3), 1-32.

Hill, E., Berthoz, S., & Frith, U. (2004). Cognitive processing of own emotions in individuals with autism spectrum disorder and in their relatives. *Journal of Autism and Developmental Disorders*, Vol. 34, 229-235.

Jack, A.I., Friedman, J.P., Boyatzis, R.E., & Taylor, S.N. (2016). Why Do You Believe in God? Relationships between Religious Belief, Analytic Thinking, Mentalizing and Moral Concern. *PLoS ONE*, Vol. 11 (3), e0149989. doi:10.1371/journal.pone.0149989

Koenig, H. G., & Büssing, A. (2010). The Duke University religion index (DUREL): A five-item measure for use in epidemiological studies. *Religion*, Vol. 1 (1), 78-85.

Lind, S. (2010). Memory and the self in autism spectrum disorder: A review and theoretical framework. *Autism*, Vol. 14, 430-456.

McCann, C., & Roberts, R. D. (2008). New Paradigms for Assessing Emotional Intelligence: Theory and Data. *Emotion,* Vol. 8, 540-551.

McCauley, R. N. (2000). The Naturalness of Religion and the Unnaturalness of Science. In F. Keil & R. Wilson (Eds.), *Explanation and Cognition*, (pp. 61-85). Cambridge: MIT Press.

McCauley, R. N. (2011). *Why Religion Is Natural and Science Is Not.* New York: Oxford University Press.

McCauley, R. N., & Graham, G. (2020). *Hearing Voices and Other Unusual Experiences: What Mental Abnormalities Can Teach Us about Religions.* New York: Oxford University Press.

Miller, S. A. (2012). *Theory of Mind: Beyond the Preschool Years*. New York: Psychology Press.

Norenzayan, A. (2013). *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton: Princeton University Press.

Norenzayan, A., Gervais, W. M., & Trzesniewski, K. H. (2012). Mentalizing Deficits Constrain Belief in a Personal God. *PLoS ONE*, Vol. 7 (5), e36880. doi:10.1371/journal.pone.0036880

Paloutzian, R. F., & Ellison, C. W. (1991). *Manual for Spirtual Well-Being Scale.* Nyack: Life Advance, Inc.

Plante, T. G., Vallaeys, C. L., Sherman, A. C., & Wallston, K. A. (2002). The development of a brief version of the Santa Clara Strength of Religious Faith Questionnaire. *Pastoral Psychology*, Vol. 50 (5), 359-368.

Reddish, P., Tok, P., & Kundt, R. (2016). Religious Cognition and Behavior in Autism: The Role of Mentalizing, *International Journal for the Psychology of Religion*, Vol. 26 (2), 95-112.

Rothstein, J. (2011). *The Affective Foundations of Moral Cognition and Justification: A Naturalistic Account.* PhD. Dissertation, Department of Philosophy, Emory University.

Saroglou, V. (2002). Religion and the Five Factors of Personality: A Meta-analytic Review. *Personality and Individual Differences*, Vol. 32 (1), 15-25.

Saroglou, V. (2010). Religiousness as a cultural adaptation of basic traits: A five-factor model perspective. *Personality and Social Psychology Review*, Vol. 14, 108-125.

Slone, D. J. (2004). *Theological Incorrectness: Why Religious People Believe What They Shouldn't*. New York: Oxford University Press.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.

Visuri, I. (2019). Sensory Supernatural Experiences in Autism. *Religion, Brain & Behavior.* DOI: https://doi.org/10.1080/2153599X.2018.1548374

Vonk, J., & Pitzen, J. (2016). Believing in other minds: Accurate mentalizing does not predict religiosity. *Personality and Individual Differences*, Vol. 115, 70-76. DOI: http://dx.doi.org/10.1016/j.paid.2016.06.008

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., Goldenfeld, N., Delaney, J., Fine, D., Smith, R. & Weil, L. (2006). Development of short forms of the Empathy Quotient (EQ-Short) and the Systemizing-Quotient (SQ-Short). *Personality and Individual Differences*, Vol. 41, 929–940.

Willard, A. K., & Norenzayan, A. (2013). Cognitive biases explain religious belief, paranormal belief, and belief in life's purpose. *Cognition*, Vol. 129, 379-391.

Wlodarski, R., & Pearce, E. (2016). The God Allusion: Individual variation in agency detection, mentalizing and schizotypy and their associations with religious behaviors. *Human Nature*, Vol. 27, 160-172.