COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–30

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12435

Core Intuitions About Persons Coexist and Interfere With Acquired Christian Beliefs About God

Michael Barlev, Spencer Mermelstein, Tamsin C. German

Department of Psychological & Brain Sciences, University of California, Santa Barbara

Received 10 November 2015; received in revised form 4 June 2016; accepted 20 June 2016

Abstract

This study tested the hypothesis that in the minds of adult religious adherents, acquired beliefs about the extraordinary characteristics of God coexist with, rather than replace, an initial representation of God formed by co-option of the evolved person concept. In three experiments, Christian religious adherents were asked to evaluate a series of statements for which core intuitions about persons and acquired Christian beliefs about God were consistent (i.e., true according to both [e.g., "God has beliefs that are true"] or false according to both [e.g., "All beliefs God has are false"]) or inconsistent (i.e., true on intuition but false theologically [e.g., "God has beliefs that are false"] or false on intuition but true theologically [e.g., "All beliefs God has are true"]). Participants were less accurate and slower to respond to inconsistent versus consistent statements, suggesting that the core intuitions both coexisted alongside and interfered with the acquired beliefs (Experiments 1 and 2). In Experiment 2 when responding under time pressure participants were disproportionately more likely to make errors on inconsistent versus consistent statements than when responding with no time pressure, suggesting that the resolution of interference requires cognitive resources the functioning of which decreases under cognitive load. In Experiment 3 a plausible alternative interpretation of these findings was ruled out by demonstrating that the response accuracy and time differences on consistent versus inconsistent statements occur for God—a supernatural religious entity—but not for a natural religious entity (a priest).

Keywords: Core cognition; Person concept; God concept; Theological incorrectness; Religious beliefs

1. Introduction

The tendency to attribute supernatural entities (e.g., gods, spirits, ancestor spirits, and divine beings) with person-like characteristics is widespread among present and past

Correspondence should be sent to Michael Barlev or Tamsin C. German, Department of Psychological & Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106-9660. E-mail: michael.barlev@psych.ucsb.edu or tamsin.german@psych.ucsb.edu

human cultures; indeed, it is noted in writings dating as far back as ancient Greece (e.g., Boyer, 1994a,b, 2001). However, it was only with relatively recent theoretical advances in cognitive science that this tendency could be explained via the evolved, universal information-processing architecture of the human mind: Supernatural entities are attributed with person-like characteristics because they are formed by co-opting the evolved person concept (also referred to as a "person template"; e.g., Boyer, 2001; Boyer & Ramble, 2001). The person concept consists of default inferences about persons, such as about their physicality, biology, and psychology, which reliably develop from a skeletal set of inferences about persons present in infancy and from associated learning adaptations (e.g., Baillargeon, 2004; Baillargeon, Scott, & Bian, 2016; Carey, 1985, 2009; Inagaki & Hatano, 2002, 2006; Spelke, 1990).

However, supernatural entities are also believed to have extraordinary characteristics which are inconsistent with default inferences about persons.[1] The exact characteristics depend on the supernatural entity and the theological tradition. For example, in all mainstream Christian denominations God is believed to be omniscient, while persons are intuitively believed to have limited perceptual and mental abilities (e.g., Boyer, 1994a,b, 2001; Boyer & Ramble, 2001). In this study we investigate the hypothesis proposed by Sperber and colleagues (Sperber, 1985, 1996, 1997, 2000; Mercier & Sperber, 2009; also see Boyer, 1994a,b, 2001; Boyer & Ramble, 2001; Barrett & Keil, 1996; Barrett, 1998, 1999), according to which characteristics attributed to supernatural entities which are inconsistent with default inferences about persons (a) do not replace these inferences but (b) coexist with them in the minds of religious believers.[2]

1.1. Previous research

A variety of studies have examined the psychological, biological, and physical characteristics adults attribute to God, concluding that they are quite willing to attribute to God certain human characteristics, especially certain kinds of mental states (e.g., Epley, Converse, Delbosc, Monteleone, & Cacioppo, 2009; Gray, Gray, & Wegner, 2007; Shtulman, 2008; Shtulman & Lindeman, 2016; see Heiphetz, Lane, Waytz, & Young, 2016, for a recent review of psychological attributions to God by children and adults). For example, Gray et al. (2007) suggested that there are two distinct dimensions of mental state attributions: Agency (consisting of mental states such as self-control, morality, memory, emotion, recognition, planning, communication, and thought) and Experience (consisting of mental states such as hunger, fear, pain, pleasure, rage, desire, personality, consciousness, pride, embarrassment, and joy), and that adults are significantly more likely to attribute to God Agency than Experience. Additionally, in their study, participants reporting stronger religious beliefs were more likely to attribute Agency to God (see their Supplementary Material); elsewhere other individual difference variables such as attachment style (e.g., Granqvist & Kirkpatrick, 2008; Kirkpatrick, 2005) were found to be associated with different mental state attributions.

Although these studies are compatible with the coexistence hypothesis, they are also compatible with two (not mutually exclusive) alternatives. First, it is possible that some

adults in these studies did not acquire the relevant theological doctrines to fully replace all default inferences about persons in their representations of God. The development of an understanding of the extraordinary characteristics of God, such as omniscience, is compatible with this alternative. The ability to attribute infallibility (a component of omniscience) to God is acquired in childhood: Lane, Wellman, and Evans (2010) found that at the age when children begin attributing false beliefs to persons on explicit tasks (between 4 and 5 years) they also attribute false beliefs to God, but that this latter tendency decreases with age; Lane, Wellman, and Evans (2012) found that religiously schooled children begin attributing infallibility to God at earlier ages than secularly schooled children, and that in children who attribute fallibility to persons (e.g., by explicitly attributing to persons false beliefs), familiarity with information about God is associated with earlier attributions of infallibility to God. Incidentally, religiously schooled children in their study attributed infallibility to a being other than God at earlier ages as well (e.g., Mr. Smart, who was described as a man who knows everything). This suggests that religiously schooled children may be earlier to understand extraordinary mental states in general, not solely God's extraordinary mental states (see also Lane, Wellman, & Evans, 2014).

However, and critically, a full understanding of omniscience requires many years to develop: The breadth of omniscience (having knowledge of all domains) is not fully appreciated before middle childhood, and the depth of omniscience (having all knowledge within a specific domain) is not fully appreciated before late adolescence to early adulthood (Lane et al., 2014). Thus, it is possible, given the years it takes to acquire a full understanding of certain extraordinary characteristics (e.g., omniscience), that a full understanding of only some extraordinary characteristics develops by early adulthood (the age typically examined in studies of adults' attributions of characteristics to God); studies showing explicit attributions of human characteristics to God may, therefore, merely show that adults do not fully know or understand the nature of God's extraordinary characteristics in Christian theology.

Second, in studies investigating the attribution of various characteristics to God using self-report methodologies, it is possible that some adults have acquired the relevant theological doctrines but intentionally deviated from them in their reports. For example, in her ethnographic study of Evangelical Christians in Chicago and the Bay Area, Luhrmann (2012) found that many of her participants attributed person-like mental states to God. But, when pressed, they would acknowledge that these attributions deviated from the theology of their group; these attributions, they said, allowed them to experience God more closely and intimately.

The primary line of research to have investigated the coexistence hypothesis (there termed "theological incorrectness") which controls for the above alternatives are the studies of memory confusions in religious adults by Barrett and colleagues (Christian adults in Barrett & Keil, 1996; Hindu adults in Barrett, 1998; see Barrett, 2000, for a review). For example, in Barrett and Keil (1996) participants asked to recall narratives, such as about God intervening to answer a prayer, were shown to mistakenly add physical and/or psychological limitations to God's actions not present in the original narratives (e.g., that to intervene God has to finish answering another prayer or stop another action, such as

helping an angel work on a crossword puzzle), and which were not in accord with the participants' self-reported theological beliefs. Barrett and Keil interpreted these memory confusions as showing that in recalling the narratives about God, participants mistakenly relied on their intuitions about persons (e.g., sequential action).

However, one major critique of the studies by Barrett and colleagues is that a person-like representation of God was implied in the narratives themselves (e.g., Shtulman, 2008; see Hyde, 1990, for a similar critique of studies in which children are asked to draw an image of God). As Shtulman (2008) notes:

> God was described in other stories as pushing a large stone, looking at the rock, listening to the birds, enjoying the smell, and helping an angel work on a crossword puzzle. Any participants who might have disagreed with the anthropomorphic implications of these statements were still required to reason on their basis. To these participants, stories about a looking, listening, helping God would be as incongruent with their personal beliefs as stories about a looking, listening, helping teapot, yet one could hardly fault them for drawing anthropomorphic inferences consistent with the stories' premises.

Shtulman (2008) considers it plausible that this language may have contributed to the person concept-based responses in recall of the narratives in the Barrett and Keil (1996) and Barrett (1998) studies.[3]

1.2. The current study

The primary goal of this study is to provide a novel test of the coexistence hypothesis in religious beliefs, using the case of Christian beliefs about God. The methodology used is the sentence verification task of Shtulman and Valcarcel (2012), which consists of an explicit measure of response accuracy, and an implicit measure of response time. In the task, participants are required to endorse or reject statements of two broad kinds: consistent statements that are true or false according to both core intuitions about persons and Christian theology about God (e.g., "God has beliefs that are true"; "All beliefs God has are false"), and inconsistent statements that are either true intuitively but false theologically (e.g., "God has beliefs that are false") or false intuitively but true theologically (e.g., "All beliefs God has are true").

As per the outline of the person concept in the Introduction, intuitive beliefs about persons were derived from reliably developing default inferences about persons that have been well established by research with infants and toddlers. Acquired beliefs about God were derived from common Christian theological doctrines about God's mental and physical characteristics that are shared among mainstream Christian denominations. Additionally, previous studies suggest that the acquisition of particular beliefs about God's characteristics begins as early as the preschool years (e.g., see the studies by Lane et al., 2010, 2012 reviewed in section 1.1). The doctrines selected were informed by these studies so as to increase the likelihood that they will be known to adult religious adherents.

This study examined four predictions which follow from the coexistence hypothesis:

1. If core intuitions coexist with acquired beliefs which are inconsistent with them, then they might also interfere with those beliefs, and this interference might cause inconsistent statements to be responded to less accurately and more slowly than consistent statements. In principle, coexistence is possible without interference (and interference might be so weak as to be imperceptible with our methods), but as interference necessarily requires coexistence, to demonstrate interference would also be to demonstrate coexistence. The first prediction therefore aims to support and extend the coexistence hypothesis as shown for scientific beliefs (e.g., Shtulman & Valcarcel, 2012) to the domain of religious beliefs.
2. If acquired beliefs and core intuitions are in conflict, then it is plausible that there are cognitive mechanisms involved in resolving this conflict, and that the efficiency of the functioning of these mechanisms might be decreased by putting participants under cognitive load. For example, previous findings suggest that when participants are put under time pressure their tendency to endorse intuitive but erroneous teleological (purpose- or function-based) explanations for natural phenomena is increased (e.g., Kelemen & Rosset, 2009; Kelemen, Rottman, & Seston, 2012). Therefore, we predicted that when put under time pressure participants would show a decrease in accuracy on inconsistent statements more so than on consistent statements.
3. Kelemen and Rosset (2009) showed that executive inhibition is negatively associated with intuitive but erroneous teleological explanations, and they found that a measure of inhibition (the behavioral Stroop task) was one predictor of scientific accuracy on their task. Similarly, Lindeman and Aarnio (2007) argued that ontological confusions are intuition-based, and recent evidence shows that the tendency to make ontological confusions is negatively associated with inhibition (Svedholm & Lindeman, 2013; see also Lindeman, Riekki, & Hood, 2011). If inhibition is (at least part of) the mechanism which resolves conflicts between intuitions and acquired beliefs, then it is plausible that individual differences in inhibition, as indexed by the behavioral Stroop task (Kelemen & Rosset, 2009; Kelemen et al., 2012; Riekki et al., 2011; Svedholm & Lindeman, 2013), would be related to performance on the sentence verification task, and more so than other executive functions such as working memory (Broadway & Engle, 2010; Redick et al., 2012).
4. We consider it unlikely that any amount of practice with acquired beliefs could replace inconsistent intuitions (e.g., Goldberg & Thompson-Schill, 2009; Kelemen et al., 2012; Shtulman & Harrington, 2016), but some findings suggest that practice could attenuate the effects of interference from inconsistent intuitions (e.g., Kelemen & Rosset, 2009; these findings and others are discussed in more detail in section 5.4). We therefore predicted that individual differences in practice with acquired religious and scientific beliefs (indexed with measures of religion education and science education) would be related to performance on the sentence verification task.

In contrast to the predictions listed above, the alternative hypothesis whereby acquired theology replaces default inferences but adults intentionally deviate from it in self-report measures by intentionally applying person-like characteristics to God predicts that: (a) under time pressure there should be an identical decrease in response accuracies on consistent and inconsistent statements, because participants are given less time to select a response, and (b) for correct responses—those responses where person-like characteristics would not have been applied—there should be no difference in response times between consistent and inconsistent statements.

A secondary goal of this study was to conduct a full replication of the Shtulman and Valcarcel (2012) experiment assessing the parallel case of the coexistence of early- and later-acquired scientific beliefs. Early-acquired scientific beliefs are often inconsistent with culturally transmitted (later-acquired) scientific beliefs, yet they are common in children and were historically common among the educated adults of various cultures (Vosniadou & Brewer, 1992). For example, 6-year-olds hold beliefs about the shape of the earth which are inconsistent with the culturally transmitted model of a spherical earth with people living all around it, but are consistent with a mental model of a flat earth with people living on its top surface (Vosniadou & Brewer, 1992). Vosniadou and Brewer (1992; also see Baumard & Boyer, 2013) have proposed that similarities between children in early scientific beliefs are due to developmental constraints caused by universal core intuitions (in the case of beliefs about the shape of the earth, intuitions that unsupported objects fall); the acquisition of novel scientific beliefs, similarly to the acquisition of religious beliefs, may therefore require resolving a conceptual conflict with core intuitions.

Finally, in the current climate of concern over the replicability of findings across science, but especially in psychological science (e.g., Ioannidis, 2005; Nosek & Lakens, 2014; see also other papers in that special issue), undertaking replications of existing findings alongside attempted extensions is one valuable additional tool available to the psychological scientific community that might eventually offset the problem of (lack of) replication. Accordingly, the new religion statements were intermixed in the current study with the entire set of science statements used by Shtulman and Valcarcel (2012) so as to determine if that result replicated, albeit in a slightly different design. It was predicted that the same pattern of findings would emerge.

2. Experiment 1

2.1. Method

2.1.1. Participants

Participants were 44 university students (56% female), ranging in age from 18 to 24 (M = 20), and drawn from two different samples: (a) two local churches, one Catholic and one Charismatic, that serve an almost exclusively college-aged population (these participants were paid for their time), and (b) the psychology participant pool at the University of California, Santa Barbara (these participants received class credit). Fifty-

four percent of participants identified as White, 25% identified as Hispanic or Latino, and 20% identified as Asian.

In order to ensure primary exposure to and belief in Christianity, participants were preselected to have been brought up within, and currently identify with, Christianity. Five participants did not match these criteria (all having been brought up or currently identified as atheist or agnostic), despite the initial preselection, and were not included in the final sample ($N = 39$), who were assigned pseudorandomly to receive the task under speeded instructions ($n = 20$) or unspeeded instructions ($n = 19$). Of the final sample, 95% of participants were brought up within the same Christian denomination with which they currently identified, and 95% identified as religious believers (indexed as a minimum rating of "slightly religious" on a religiosity question).

Of the final sample, 50% of participants identified as Roman Catholic, 41% identified as simply "Christian," and 9% identified as one of a number of Protestant Christian denominations (e.g., Baptist, Lutheran). Most participants who identified as "Christian" in this sample reported being affiliated with the local Charismatic church. On a 4-point Likert scale (range 0–3; *not at all*, *slightly*, *moderately*, *very*), participants on average reported being moderately religious ($M = 2.20$, $SD = 0.83$) and moderately spiritual ($M = 2.08$, $SD = 0.87$), and the two were highly correlated ($r = .522$, $p = .001$).

2.1.2. Design

The primary dependent variable was response accuracy, and with respect to this dependent variable, the design was a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) × 2 (Instructions: Speeded vs. Unspeeded) factorial with within-subjects repeated measures on the first two factors.

A secondary dependent variable, response time, was collected for participants in the speeded instructions condition. For this dependent variable, the design was a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) factorial with within-subjects repeated measures.

2.1.3. Materials

The religion statements (48 in total) were constructed in groups of four statements, with each group targeting a particular characteristic of God that is inconsistent with a core intuition about persons (see section 1.2 for a more detailed discussion). Following Shtulman and Valcarcel (2012), each group of statements was constructed such that there was one that was true on both intuition and theology, one that was true on neither, one that was true only on intuition, and one that was true only on theology. In this way one set of statements (consistent on intuition and theology) served as a baseline to which the other (inconsistent on intuition and theology) could be compared, and within each group there was an equal number of statements that were objectively true or false. Additionally, the four statements within each group were balanced in terms of overall sentence structure, complexity, and length in words.[4] Example statements appear in Table 1, and a full list of all religion statements can be found in the Supplementary Material.

The science statements were the same 200 statements used by Shtulman and Valcarcel, covering 10 areas of mathematics and science (astronomy, evolution, fractions, genetics, germs, matter, mechanics, physiology, thermodynamics, and waves). Example statements appear in Table 2, and a full list of all science statements can be found in the Supplementary Material.

Because the executive functions and practice measures were included to evaluate their relationship with interference, a sentence verification task interference score was calculated for each participant as the difference between consistent and inconsistent statements (for both response accuracies and times), with higher scores indicating stronger interference.

Additional materials included (a) a 144-item behavioral Stroop task (modified from Stroop, 1935) which included the following three conditions (48 items per condition): Congruent (the words RED, BLUE, GREEN, and YELLOW appearing in red, blue, green, and yellow color, respectively), Incongruent (the words RED, BLUE, GREEN, and YELLOW appearing in a color different than the one they spell), and Neutral (the words LOT, SHIP, KNIFE, FLOWER—length matched and frequency matched to the color words, appearing in colors); a Stroop response time interference score is then calculated by averaging the difference between the Incongruent and Neutral, and Incongruent and Congruent conditions, (b) a running span working memory task (Broadway & Engle, 2010), and (c) a short survey seeking demographic information, self-report measures of religiosity and spirituality, extent of participants' religious education, and extent of math and science education (indexed by asking participants to list all math and science courses they have taken in college). An education composite score (range 0–4) was calculated by summing the number of content areas—mathematics, biology, chemistry, and physics—participants took at least one course in.

Table 1
Sample statements from the domain of religion

Consistency	Intuition	Theology	Religion Statements
Consistent	T	T	God has beliefs that are true
	F	F	All beliefs God has are false
Inconsistent	T	F	God has beliefs that are false
	F	T	All beliefs God has are true
Consistent	T	T	God can hear what I say out loud
	F	F	God can't hear what I say out loud
Inconsistent	T	F	God can't hear what I say to myself
	F	T	God can hear what I say to myself
Consistent	T	T	God can be present at my church and at other churches as well
	F	F	God is never present at my church, nor is He present anywhere else
Inconsistent	T	F	Sometimes God is at my church, and sometimes He is at other churches
	F	T	God is at all times both at my church and at other churches

Note. Consistent statements are true on both intuition and theology; inconsistent statements are true on one and false on the other.

Table 2
Sample statements from the domain of science

Consistency	Intuition	Science	Science Statements
Consistent	T	T	Rocks are composed of matter
	F	F	Numbers are composed of matter
Inconsistent	T	F	Fire is composed of matter
	F	T	Air is composed of matter
Consistent	T	T	People turn food into energy
	F	F	Rocks turn food into energy
Inconsistent	T	F	Plants turn food into energy
	F	T	Bacteria turn food into energy
Consistent	T	T	Humans are descended from tree-dwelling creatures
	F	F	Humans are descended from plants
Inconsistent	T	F	Humans are descended from chimpanzees
	F	T	Humans are descended from sea-dwelling creatures

Note. Consistent statements are true on both intuition and science; inconsistent statements are true on one and false on the other. Statements are from Shtulman and Valcarcel (2012).

2.1.4. Procedure

In a quiet testing room, groups of up to six participants took the experiment at semi-private computer testing stations. Participants in both the speeded and unspeeded instructions conditions completed, in this order, the sentence verification task, the behavioral Stroop task, the working memory task, and the survey. In the speeded instructions condition, the instructions to the sentence verification task emphasized both response accuracy and speed (in multiple parts of the instructions participants were told to "respond as quickly as you can, while making as few mistakes as you can" and that "speed and accuracy are both very important"), and responses were collected via key presses to facilitate faster and less deliberate responding (presented via E-Prime software). In the unspeeded instructions condition, the instructions emphasized accuracy only, and responses were presented in survey form (presented via Qualtrics software) to facilitate slower and more deliberate responding. In both instructions conditions, the sentence verification task items were presented one by one and in a randomized order, and in the speeded instructions condition whether the right or left hand were used to respond "true" or "false" was randomized between participants.

2.2. Results

2.2.1. Sentence response accuracy

The primary hypothesis under test was that participants will be more accurate responding to items in which core intuitions are consistent with acquired beliefs in the domains of religion and science (see the Supplementary Material for response accuracies on individual items). The sentence response accuracy data were entered into a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) × 2 (Instructions: Unspeeded vs. Speeded) mixed analysis of variance (ANOVA) with repeated measures on

the first two factors, revealing main effects of Domain ($F_{1,37}$ = 348.7, p < .001, partial η^2 = .90) and Consistency ($F_{1,37}$ = 226.2, p < .001, partial η^2 = .86) qualified by an interaction between Domain and Consistency ($F_{1,37}$ = 50.5, p < .001, partial η^2 = .58); there was no main effect of Instructions, and the Instructions factor did not enter into any two- or three-way interactions (all Fs < 1.3, all ps = n.s.), so no further analyses involving it are reported. The interaction between Domain and Consistency is shown for both Instruction factors in Fig. 1.

Simple main effect analyses confirmed that participants performed better on the religion than on the science items for both consistent and inconsistent items (t(38) = 12.42, p < .001, d = 2.75, and t(38) = 15.93, p < .001, d = 2.78, respectively), and that the interaction resulted from the size of the effect for consistency being more than twice as large for the science items than for the religion items (t(38) = 16.90, p < .001, d = 2.81, and t(38) = 6.06, p < .001, d = 1.15, respectively).[5]

2.2.2. *Sentence response time*[6]

Sentence response time data were collected for those participants who received speeded instructions (see the Supplementary Material for response times on individual items). Response time data for correct responses were entered into a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) repeated measures ANOVA, revealing a main effect of Consistency ($F_{1,19}$ = 40.88, p < .001, partial η^2 = .68); there was no main effect of Domain and no interaction (both Fs < 1.0, ps = n.s.). The main effect of Consistency is shown in Fig. 2. Simple main effect analyses confirmed that

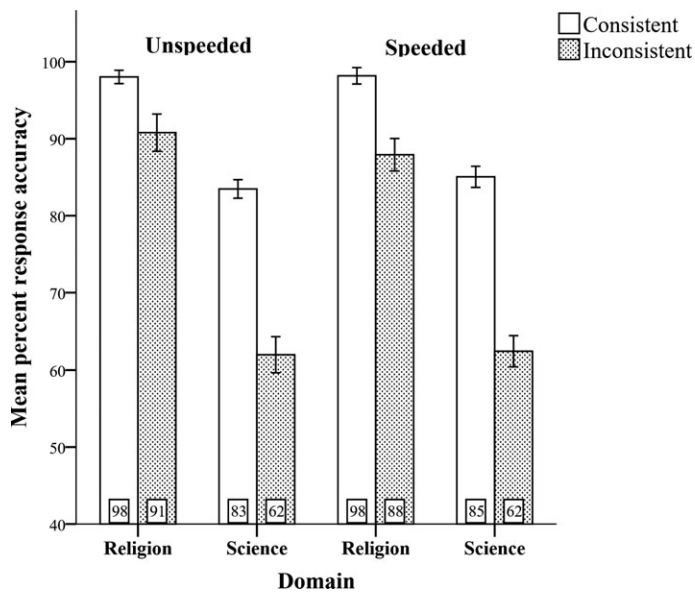

Fig. 1. Mean percent response accuracy on consistent and inconsistent items in the domains of religion and science for both the speeded and unspeeded instructions conditions. Error bars: ±1 *SE*.

participants were faster on consistent than inconsistent items for both religion and science items ($F_{1,19}$ = 11.58, p < .005, d = 0.34 and $F_{1,19}$ = 29.55, p < .001, d = 0.61, respectively), and that there were no differences in response times between science and religion items for either consistent or inconsistent items (both Fs < 1.0, ps = n.s.).

2.2.3. Associations with measures of executive functions

The inhibition[7] (Stroop response time interference scores: M = 95 ms, SD = 72 ms, for correct responses only) and working memory (M = 25.62, SD = 12.26) measures were entered into a correlational analysis with interference scores on accuracy (both speeded and unspeeded instructions) and response time (speeded instructions only) on the sentence verification task, for both the religion and science items. Neither of the two executive functions measures were correlated with either the accuracy or response time interference scores in either the religion or science domains (all ps = n.s.). The executive functions measures also were not correlated with each other (p = n.s.).

2.2.4. Associations with measures of education

The composite science education measure (M = 2.67, SD = 1.03) did not correlate with either accuracy (both speeded and unspeeded instructions) or response time (speeded instructions only) interference scores for science items (both ps = n.s.). Shtulman and Valcarcel (2012) similarly collected data on the math and science courses their participants took, and using a slightly different variable (total number of courses taken, rather

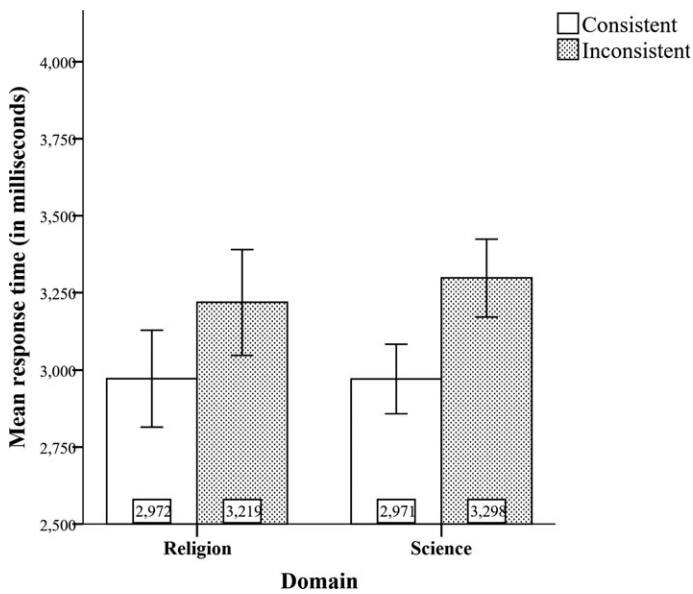

Fig. 2. Mean response time (in milliseconds) on consistent and inconsistent items in the domains of religion and science for the speeded instructions condition only. Error bars: ±1 *SE*.

than the composite score for content areas used here) did not find that it predicted any of the effects reported in their study.

Both religion education variables were strongly skewed, and Spearman's rank-order correlations were therefore used instead of Pearson's correlations, revealing no correlation between church attendance or theology study and accuracy (both speeded and unspeeded instructions) or response time (speeded instructions only) interference scores for religion items (all *p*s = n.s.). However, church attendance was correlated with accuracy on consistent religion items, and marginally with accuracy on inconsistent religion items ($r = .376$, $p < .05$, and $r = .284$, $p = .079$, respectively).

3. Experiment 2

The results of Experiment 1 support our first prediction by replicating and extending the findings of Shtulman and Valcarcel (2012), to show that, as in the domain of science, there is conflict between core intuitions and acquired beliefs in the domain of religion. Note that there was a large effect of domain on the accuracy scores reported in Experiment 1, with religion items being responded to more accurately than science items. While a comparison of the science and religion items was not a focus of this study, the difference was expected, given that the acquired religion items were specifically derived from common Christian theological doctrines about God's mental and physical characteristics so as to increase the likelihood that they will be known to adult religious adherents.

Our second prediction that the difference in response accuracies between consistent and inconsistent items would be greater under instruction to respond quickly was not supported; participants responded with almost identical levels of accuracy under speeded and unspeeded instructions.

The lack of an effect for the instructions manipulation in the current study is not critical for evaluating the coexistence hypothesis, particularly given the strong effects of consistency on response accuracy and speed. Indeed, in Kelemen and Rosset (2009) and Kelemen et al. (2012), there is evidence of the coexistence of teleological intuitions alongside scientific beliefs at the slowest response rates imposed (e.g., in the unspeeded condition in the two experiments reported in Kelemen & Rosset endorsements of teleological explanations range from 29% to 42% vs. 47% to 54% in their speeded condition).

A likely explanation for this outcome is that instructions to respond quickly alone were not a strong enough manipulation to put participants under time pressure in the speeded instructions condition. In some previous studies that used a speeding manipulation, instructions to respond quickly were accompanied by time limits on participants' response windows (e.g., Kelemen & Rosset, 2009; Kelemen et al., 2012; Riekki et al., 2011; Svedholm & Lindeman, 2013). In Experiment 2, therefore, response time limits were added to the speeding manipulation.

3.1. Method

3.1.1. Participants

Participants were 75 university students (80% female), ranging in age from 18 to 24 (M = 19), and drawn from the psychology participant pool at the University of California, Santa Barbara. All participants received class credit for their time. Thirty-eight percent of participants identified as White, 29% identified as Hispanic or Latino, 29% identified as Asian, and 4% identified as "Other."

Participants were preselected according to criteria stricter than in Experiment 1. In addition to the selection criteria used in Experiment 1, participants also needed to have been brought up within the same Christian denomination with which they currently identified and to identify as religious believers (indexed as a minimum rating of "slightly religious" on a religiosity question). Four participants did not match these criteria, despite the initial preselection, and were not included in the final sample (N = 71), who were assigned pseudorandomly to receive the task under time limit (n = 32) or no time limit (n = 39).

Of the final sample, 40% of participants identified as Roman Catholic, 39% identified as nondenominational Christian, and 21% identified as one of a number of Protestant Christian denominations (e.g., Presbyterian, Baptist). Unlike in Experiment 1 where most of the participants who identified as "Christian" reported being affiliated with the local Charismatic church, none of the participants who identified as "Christian" in this experiment did. On a 4-point Likert scale (range 0–3) participants on average reported being moderately religious (M = 1.93, SD = 0.64) and moderately spiritual (M = 1.83, SD = 0.83), and the two were highly correlated (r = .571, p < .001).

3.1.2. Design

As in Experiment 1, the primary dependent variable was response accuracy, and a secondary dependent variable was response time, which was collected for participants in both conditions. The design was a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) × 2 (Condition: Time Limit vs. No Time Limit) factorial design with within-subjects repeated measures on the first two factors.

3.1.3. Materials

The materials used were identical to those used in Experiment 1 with the exception of the behavioral Stroop task, which was modified to resemble the task used by Riekki et al. (2011) and Svedholm and Lindeman (2013). The goal of this was to rule out a possible interpretation of the failure to find an association between the sentence verification task and the version of the behavioral Stroop task used in Experiment 1: In informal debriefings, some participants reported using a response strategy whereby they only attended to the color in which words appeared by either directing their gaze to the periphery of the display or by squinting their eyes, thereby blurring the words which appear at the center of the display. In this modified version of the Stroop task, the Neutral condition was replaced with a Word-Naming condition in which the words RED, BLUE, GREEN, and

YELLOW appear in black, and participants are required to respond to the words that appear, which controls for the above response strategy by requiring them to direct their gaze to the center of the display. The Congruent condition was replaced with a Color-Naming condition in which a string of Xs appears in red, blue, green, or yellow color, and participants are required to respond to the color in which the Xs appear. The Incongruent condition (the words RED, BLUE, GREEN, and YELLOW appearing in a color different than the one they spell) remained the same. A Stroop response time interference score is then calculated by subtracting response times on the Color-Naming condition from the Incongruent condition.

3.1.4. Procedure

The procedure was the same as in Experiment 1, except that in the speeded condition in addition to instructions emphasizing both response accuracy and speed in the sentence verification task, participants were told that each statement will appear for a short duration, and that the durations will be of variable times. The actual times (M = 3,298 ms, SD = 962 ms; range 1,605–6,749 ms) were determined through pretesting (N = 15) as the average reading time plus two standard deviations of each statement. The rationale for using variable times was that a fixed time limit would be insensitive to any overall differences in reading times for the different sentences; variable times allowed us to approximately equate the time available for responding after reading. In the unspeeded condition, the instructions emphasized response accuracy only and statements appeared until participants responded. In both conditions, responses were collected via key presses (presented via E-Prime software).

3.2. Results

3.2.1. Sentence response accuracy[8]

As in Experiment 1, the primary hypothesis under test was that participants would be more accurate responding to items in which core intuitions are consistent with acquired beliefs in the domains of religion and science. The accuracy data were subjected to a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) × 2 (Condition: Time Limit vs. No Time Limit) mixed ANOVA with repeated measures on the first two factors, revealing main effects of Domain ($F_{1,69}$ = 480.2, p < .001, partial η^2 = .87), Consistency ($F_{1,69}$ = 576.4, p < .001, partial η^2 = .89), and Condition ($F_{1,69}$ = 60.5, p < .001, partial η^2 = .47), qualified by interactions between Domain and Consistency ($F_{1,69}$ = 60.0, p < .001, partial η^2 = .46) and Consistency and Condition ($F_{1,69}$ = 4.8, p < .05, partial η^2 = .06). The interaction between Domain and Condition was not statistically significant ($F_{1,69}$ = 2.4, p = .123, partial η^2 = .03), and the factors did not enter into a three-way interaction ($F_{1,69}$ = 2.7, p = .105, partial η^2 = .04). The interaction between Domain and Consistency is shown for both the unspeeded (no time limit) and speeded (time limit) conditions in Fig. 3.

Simple main effect analyses replicated the findings reported in Experiment 1 by confirming that participants performed better on the religion items than on the science items

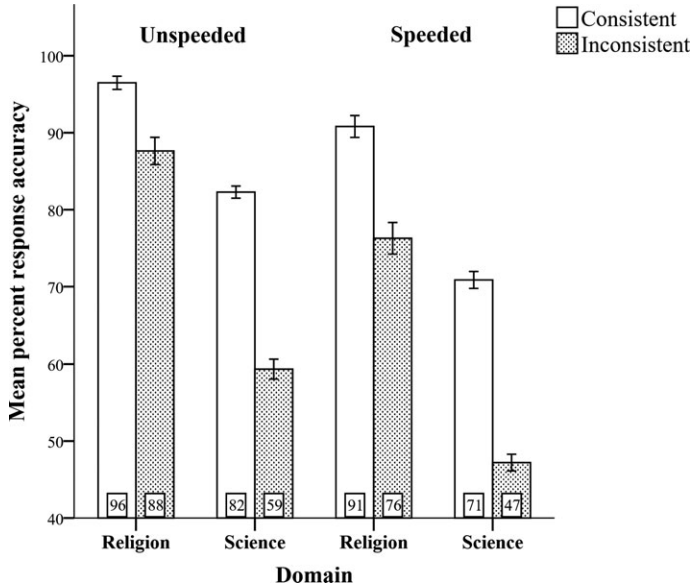

Fig. 3. Mean percent response accuracy on consistent and inconsistent items in the domains of religion and science for both the time limit and no time limit conditions. A mean response accuracy of 50% represents chance responding. Error bars: ±1 *SE*.

for both consistent and inconsistent items ($t(70) = 14.92$, $p < .001$, $d = 2.17$, and $t(70) = 20.32$, $p < .001$, $d = 2.56$, respectively), and that the interaction between Domain and Consistency resulted from the size of the effect for consistency being more than twice as large for the science items than for the religion items ($t(70) = 30.01$, $p < .001$, $d = 2.67$, and $t(70) = 9.30$, $p < .001$, $d = 1.12$, respectively).[9]

An examination of the mean differences, separated by condition, between consistent and inconsistent religion ($M_{\text{unspeeded}} = 5.67\%$, vs. $M_{\text{speeded}} = 11.35\%$) and science ($M_{\text{unspeeded}} = 11.41\%$, vs. $M_{\text{speeded}} = 12.13\%$) items revealed that the two-way interaction between Consistency and Condition was primarily carried by the religion items.[10] The response accuracy data suggest that this was caused by a floor effect in responses to inconsistent science items in the time limit condition which were at chance responding.

An analysis of the timed out responses in the time limit condition further supports this interpretation. A 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) repeated measures ANOVA with proportion of incorrect responses that were due to time outs as the DV revealed main effects of Domain ($F_{1,38} = 5.75$, $p < .05$, partial $\eta^2 = .13$) and Consistency ($F_{1,38} = 15.48$, $p < .001$, partial $\eta^2 = .29$) and no interaction. A larger number of incorrect responses on science items than on religion items were due to time outs ($M_{\text{science}} = 25.60\%$ vs. $M_{\text{religion}} = 18.70\%$), and a larger number of incorrect responses on consistent items than on inconsistent items were due to time outs ($M_{\text{consistent}} = 27.90\%$ vs. $M_{\text{inconsistent}} = 16.40\%$). A plausible interpretation of this pattern is that overall the science items were more difficult than the religion items, but as they

were allowed approximately similar (or often shorter, given that the science sentences were often shorter) response windows, participants who tried to think about them for too long before responding were timed out. In contrast, inconsistent items (particularly inconsistent science items) were significantly more difficult than consistent items, so much so that participants might often have chosen an answer at random, causing faster response times and fewer time outs.

3.2.2. Sentence response time[11]

Response time data for correct responses in the unspeeded condition only[12] were entered into a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) repeated measures ANOVA, revealing a main effect of Consistency ($F_{1,31}$ = 37.86, $p < .001$, partial $\eta^2 = .55$) and an interaction between Domain and Consistency ($F_{1,31}$ = 21.98, $p < .001$, partial $\eta^2 = .42$). Simple main effect analyses revealed that the main effect of Consistency was carried by the science items ($t(31) = 9.22$, $p < .001$, $d = 0.84$, and $t(31) = 1.21$, p = n.s., $d = 0.14$, for science and religion, respectively), and that despite the lack of a main effect of Domain in the preceding analysis ($F_{1,31}$ = 1.53, p = n.s., partial $\eta^2 = .05$), the response time for inconsistent religion items was significantly faster than for inconsistent science items ($t(31) = 3.36$, $p = .002$, $d = 0.51$). The mean response times are shown in Fig. 4.

3.2.3. Associations with measures of executive functions

As in Experiment 1, inhibition (Stroop response time interference scores: M = 178 ms, SD = 110 ms, for correct responses only) and working memory (M = 22.75, SD = 10.46)

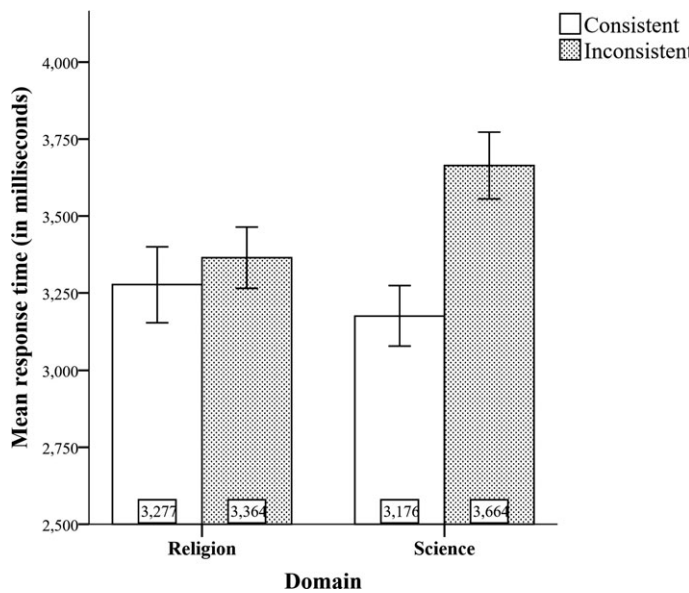

Fig. 4. Mean response time (in milliseconds) on consistent and inconsistent items in the domains of religion and science for the no time limit condition only. Error bars: ±1 *SE*.

were entered into a correlational analysis with interference scores on accuracy (analyzed separately in the time limit and no time limit conditions) and response time (analyzed in the no time limit condition only) on the sentence verification task, for both the religion and science items.[13] Neither inhibition nor working memory were correlated with either the accuracy or response time interference scores in either the religion or science domains (all *p*s = n.s.). As in Experiment 1, the executive functions measures were not correlated with each other (p = n.s.), but the higher mean response time interference score on the behavioral Stroop task was nearly identical to that reported elsewhere in similar populations (e.g., Miyake et al., 2000), supporting our interpretation that the new version of the task used in this experiment controlled for the use of response strategies reported by some participants in Experiment 1.

3.2.4. Associations with measures of education

The composite science education variable (M = 2.32, SD = 1.08) did not correlate with interference scores for the science items, either in terms of accuracy (analyzed separately in the time limit and no time limit conditions) or response time (analyzed in the no time limit condition), both *p*s = n.s.

Both religion education variables were strongly skewed, and Spearman's rank-order correlations were therefore used instead of Pearson's correlations, revealing no correlation between church attendance or theology study with interference scores for the religion items, either in terms of accuracy (analyzed separately in the time limit and no time limit conditions) or response time (analyzed in the no time limit condition), all *p*s = n.s. However, for consistent and inconsistent religion items, church attendance was correlated with accuracy ($r = .623$, $p < .001$, and $r = .483$, $p < .01$, respectively) in the no time limit condition.

4. Experiment 3

The primary prediction of this study, that inconsistent statements would be responded to less accurately and more slowly than consistent statements, was supported in Experiments 1 and 2. The goal of Experiment 3 was to control for the possibility that this pattern was caused by systematic low-level biases in the statements used, rather than by coexistence and interference of core intuitions on acquired beliefs. For example, processing of the inconsistent versus consistent statements within each group might have been more difficult due to a systematic syntactic bias which was not controlled for during statement construction (e.g., passive vs. active, or affirmative vs. negative, sentences; e.g., see Wason, 1959, for a classic study of response time differences between affirmative and negative sentences). Note that although no such systematic low-level biases were identified a priori, we nonetheless considered it important to include this control experiment. To test this, in a subset of the religion statements in Experiment 3, the supernatural entity "God" was replaced with a nonsupernatural religious agent ("my priest"); it was predicted that as participants hold no acquired beliefs inconsistent with core intuitions

about this nonsupernatural agent, there would be no differences in accuracy or response time on the modified "inconsistent" and "consistent" statements (these terms were retained for ease of comparison).

4.1. Method

4.1.1. Participants

Participants were 37 university students (81% female) ranging in age from 18 to 21 (M = 19) drawn from the psychology participant pool at the University of California, Santa Barbara. All participants received class credit for their time. Thirty-eight percent of participants identified as White, 24% identified as Hispanic or Latino, 8% identified as Black, 19% identified as Asian, and 11% identified as "Other."

Participants were preselected according to the same criteria used in Experiment 2 (see section 3.1.1 for details). Four participants did not match these criteria, despite the initial preselection, and were therefore not included in the final sample (N = 33).

Of the final sample, 30% of participants identified as Roman Catholic, 37% identified as nondenominational Christian, and 33% identified as one of a number of Protestant Christian denominations (e.g., Methodist, Pentecostal). On a 4-point Likert scale (range 0–3), participants on average reported being moderately religious (M = 1.85, SD = 0.51) and moderately spiritual (M = 1.61, SD = 0.70), and the two were highly correlated (r = .44, p = .01).

4.1.2. Design

As in Experiments 1 and 2, the primary dependent variable was response accuracy and a secondary dependent variable was response time. The design was a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) factorial design with within-subjects repeated measures.

4.1.3. Materials

A subset of the religion statements (sets T3, T5, T6, T8, and T9; see the Supplementary Material) was modified by replacing the word "God" with "my priest." The religion statements modified in this manner were not selected at random. Rather, they were selected because they could be modified and still remain coherent. For example, a number of statements could not be modified because they pertained to God listening to prayers, and prayers are directed toward God but not toward priests. Example statements appear in Table 3.

The modified religion statements were presented in random order along with the remaining unmodified religion statements and the entire set of science statements. No measures of executive functions or education were administered in this experiment.

Table 3
Sample modified statements from the domain of religion

Intuition	Theology	Religion Statements	Modified Religion Statements
Consistent			
T	T	God has beliefs that are true	My priest has beliefs that are true
F	F	All beliefs God has are false	All beliefs my priest has are false
Inconsistent			
T	F	God has beliefs that are false	My priest has beliefs that are false
F	T	All beliefs God has are true	All beliefs my priest has are true

Note. The terms "consistent" and "inconsistent" do not apply to the modified religion statements, but they are used for illustrative purposes only. The modified religion statements are true or false according to intuition only.

4.1.4. Procedure

The procedure was the same as in Experiment 2, except that all participants were assigned to the same unspeeded condition (identical to the unspeeded condition in Experiment 2).

4.2. Results

4.2.1. Sentence response accuracy

A paired-samples *t* test revealed no response accuracy difference between the modified "consistent" and "inconsistent" religion items ($M_{\text{consistent}}$ = 77.30%, $SD_{\text{consistent}}$ = 15.47%, vs. $M_{\text{inconsistent}}$ = 73.90%, $SD_{\text{inconsistent}}$ = 18.53%; $t(32) = .69$, $p > .5$, $d = 0.20$). Note that the response accuracies for the modified religion items are relatively low. We think that this is because in replacing "God" with "my priest," we introduced ambiguity into the pragmatic interpretations of some items, which were designed to evaluate beliefs about God rather than about priests. For example, the items "my priest knows of various things that happened in the past" or "my priest doesn't know of things that happened in the past" may be interpreted as proclamations about knowledge in general, in which case the correct response to the first item is "true" and to the second item "false," because people have at least some knowledge of the past. Alternatively, these items may be interpreted as proclamations about the priest's competence, in which case responses to these items are dependent on particular priests. In any case, what is critical for our purposes is that items were evaluated relative to others in their group—the relevant data are within-group differences between "consistent" and "inconsistent" items, not overall accuracy.

In contrast, a paired-samples *t* test on the unmodified versions of these same items (using the "God" rather than "my priest" wording) from the no time limit condition from Experiment 2 revealed a response accuracy difference between the consistent and inconsistent items ($M_{\text{consistent}}$ = 96.60%, $SD_{\text{consistent}}$ = 5.45%, vs. $M_{\text{inconsistent}}$ = 88.70%, $SD_{\text{inconsistent}}$ = 13.62%; $t(31) = 3.14$, $p < .01$, $d = 0.76$).

Next, to examine if the main findings from Experiments 1 and 2 replicated in this experiment the response accuracy data were subjected to a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) repeated measures ANOVA with

the unmodified religion items only. The analysis revealed main effects of Domain ($F_{1,32}$ = 195.41, p < .001, partial η^2 = .86) and Consistency ($F_{1,32}$ = 307.20, p < .001, partial η^2 = .91), qualified by an interaction between Domain and Consistency ($F_{1,32}$ = 53.23, p < .001, partial η^2 = .62).

Simple main effect analyses confirmed that as in Experiments 1 and 2 participants performed better on the religion items than on the science items in both the consistent (M_{religion} = 97.19%, SD_{religion} = 6.67%, vs. M_{science} = 81.61%, SD_{science} = 5.66%; $t(32)$ = 10.78, p < .001, d = 2.52) and inconsistent conditions (M_{religion} = 85.06%, SD_{religion} = 9.68%, vs. M_{science} = 57.61%, SD_{science} = 7.82%; $t(32)$ = 13.77, p < .001, d = 3.12). The interaction entailed the size of the effect for consistency being more than twice as large for the science items ($M_{\text{consistent}}$ = 81.61%, $SD_{\text{consistent}}$ = 5.66%, vs. $M_{\text{inconsistent}}$ = 57.61%, $SD_{\text{inconsistent}}$ = 7.82%; $t(32)$ = 19.18, p < .001, d = 3.52) than for the religion items ($M_{\text{consistent}}$ = 97.19%, $SD_{\text{consistent}}$ = 6.67%, vs. $M_{\text{inconsistent}}$ = 85.06%, $SD_{\text{inconsistent}}$ = 9.68%; $t(32)$ = 8.83, p < .001, d = 1.46).

4.2.2. Sentence response time[14]

Paired-samples t tests revealed no response time differences on correct responses between the modified "consistent" and "inconsistent" items ($M_{\text{consistent}}$ = 3,770 ms, $SD_{\text{consistent}}$ = 1,178 ms, vs. $M_{\text{inconsistent}}$ = 3,541 ms, $SD_{\text{inconsistent}}$ = 992 ms; $t(32)$ = .96, p = n.s., d = 0.21), and no response time differences on correct responses on the unmodified versions of these same items from the no time limit condition from Experiment 2 ($M_{\text{consistent}}$ = 3,277 ms, $SD_{\text{consistent}}$ = 693 ms, vs. $M_{\text{inconsistent}}$ = 3,364 ms, $SD_{\text{inconsistent}}$ = 563 ms; $t(31)$ = 1.21, p = n.s., d = 0.14); however, response times on consistent and inconsistent items were in the predicted direction for the unmodified items from Experiment 2 (consistent < inconsistent), and in the opposite direction for the modified items (consistent > inconsistent).

Next, to examine if the main findings from Experiments 1 and 2 replicated in this experiment the response time data on correct responses were subjected to a 2 (Domain: Religion vs. Science) × 2 (Consistency: Consistent vs. Inconsistent) repeated measures ANOVA with the unmodified religion items only. The analysis revealed main effects of Domain ($F_{1,32}$ = 7.44, p = .01, partial η^2 = .19) and Consistency ($F_{1,32}$ = 68.12, p < .001, partial η^2 = .68) qualified by an interaction between Domain and Consistency ($F_{1,32}$ = 8.36, p < .01, partial η^2 = .21).

Simple main effect analyses revealed that participants were faster on the science than on the religion items for consistent items (M_{science} = 2,991 ms, SD_{science} = 632 ms, vs. M_{religion} = 3,284 ms, SD_{religion} = 715 ms; $t(32)$ = 3.99, p < .001, d = 0.43), and that there was no difference in response times between the science and religion items for inconsistent items (M_{religion} = 3,576 ms, SD_{religion} = 791 ms, vs. M_{science} = 3,538 ms, SD_{science} = 779 ms; $t(32)$ = .49, p = n.s., d = 0.05). The interaction entailed the size of the effect for consistency being nearly twice as large for the science items ($M_{\text{consistent}}$ = 2,991 ms, $SD_{\text{consistent}}$ = 632 ms, vs. $M_{\text{inconsistent}}$ = 3,538 ms, $SD_{\text{inconsistent}}$ = 779 ms; $t(32)$ = 9.19, p < .001, d = 0.77) than for the religion items ($M_{\text{consistent}}$ = 3,284 ms, $SD_{\text{consistent}}$ = 715 ms, vs. $M_{\text{inconsistent}}$ = 3,576 ms, $SD_{\text{inconsistent}}$ = 791 ms; $t(32)$ = 3.91, p < .001, d = 0.39).

5. General discussion

5.1. Support for the coexistence of inconsistent acquired beliefs and core intuitions

The primary goal of this study was to test the hypothesis that in the minds of adult religious adherents acquired Christian beliefs about God coexist with, rather than replace, an initial representation of God formed by co-option of the evolved person concept. The experiments reported here utilized a task where participants were required to evaluate statements that were true or false according to both core intuitions about persons and Christian theology (consistent statements), and statements that were either true intuitively but false theologically or false intuitively but true theologically (inconsistent statements). In Experiments 1 and 2, participants were less accurate and slower in evaluating inconsistent versus consistent statements, thereby demonstrating both coexistence and interference in the domain of religion.

Furthermore, by replicating and expanding on the findings of Shtulman and Valcarcel (2012) with earlier- and later-acquired scientific beliefs, this study brings attention to the theoretical parallels in coexistence in the domains of religion and science, which have thus far been studied mostly independently.

In contrast to previous studies purporting to show that adults attribute person characteristics to God, this study controlled for two plausible alternative interpretations. According to one, attributing person-like characteristics to God is caused by some adults not having acquired the relevant theological doctrines to fully know which characteristics God is thought to have. However, the very high performance on inconsistent religion statements and near perfect performance on consistent religion statements suggest that the acquired beliefs about God examined in this study were known to the participants tested (as indeed was predicted in deriving these beliefs from common Christian theological doctrines about God's mental and physical characteristics—see section 1.2).

Additionally, an examination of response accuracies on individual religion statements (see the Supplementary Material) shows that for the most part errors are relatively evenly distributed among the different statements. That is, there was a small but reliable probability that participants would make an error on any given statement regardless of the characteristic the statement pertained to. The alternative whereby participants did not know certain parts of the relevant theology predicts clustering of errors around only few statements, for example, statements pertaining to characteristics of God talked about ambiguously and/or rarely.

According to another alternative interpretation, attributing person-like characteristics to God is caused by some adults intentionally deviating from theology (i.e., speaking of God metaphorically; see, e.g., Luhrmann, 2012). This alternative predicts, for correct responses, no difference in response times between consistent and inconsistent statements. However, in both the speeded instructions condition of Experiment 1 and the no time limit condition of Experiment 2, inconsistent religion statements were responded to

significantly more slowly than consistent religion statements. Additionally, this alternative predicts an identical decrease in response accuracies on consistent and inconsistent statements when participants respond under time pressure, but in the time limit condition in Experiment 2 response accuracies decreased significantly more for inconsistent than for consistent religion statements.

Finally, in Experiment 3, by showing that the difference between consistent and inconsistent statements disappeared when replacing the term "God" with "my priest" in a subset of the statements, we controlled for the possibility that the findings reported here were caused by systematic low-level biases (e.g., syntactic biases) in the statements used rather than by coexistence.

5.2. Support for the existence of mechanisms that resolve conflicts between inconsistent beliefs

If core intuitions coexist and interfere with acquired beliefs which are inconsistent with them, then certain mechanisms should exist to resolve the interference or conflict created by tasks in which both representations are engaged (e.g., Kelemen & Rosset, 2009; Kelemen et al., 2012; Riekki et al., 2011; Svedholm & Lindeman, 2013). A second prediction derived from the coexistence hypothesis which was tested here was that the effect of consistency on response accuracy will increase when participants are made less able to resolve conflicts between inconsistent beliefs.

In Experiment 2, a subset of participants was assigned to a speeded responding condition where a time limit was set on responses, and another subset received the task with no such time limit (unspeeded responding condition). The findings demonstrated that when participants are made to respond quickly, they disproportionately make more errors on inconsistent versus consistent religion statements. While this finding does not as yet highlight any specific mechanism (also see section 5.3), the effect of cognitive load in the form of time pressure does support the prediction that those cognitive mechanisms that resolve conflicts between inconsistent beliefs have limited efficiency, and intuitive beliefs are more likely to lead to errors in responding when the conflict between them and acquired beliefs is less able to be resolved.

Note that these findings are compatible with, although they do not uniquely support, the suggestion by Sperber and colleagues (Mercier & Sperber, 2009; Sperber, 1997, 2000; also see Barrett, 1999) and by dual-process theorists (Evans, 2003, 2008; Evans & Stanovich, 2013; also see Mercier & Sperber, 2011) that acquired beliefs inconsistent with core intuitions exist in a specialized meta-representational format distinct from that of intuitive beliefs (see our note 2). If this is the case, then under cognitive load religious adherents are more likely to rely on erroneous intuitive beliefs not only because of a limitation of conflict resolution mechanisms, but because intuitive beliefs require significantly fewer processing resources than meta-represented religious beliefs (see Evans, 2003, 2008; Evans & Stanovich, 2013).

5.3. No correlations with executive inhibition

A third prediction derived from the coexistence hypothesis which was tested here was that executive inhibition, as indexed by the behavioral Stroop task, is the process that resolves the interference or conflict of core intuitions with acquired beliefs. Previous studies are equivocal with respect to this question. For example, on the one hand, Kelemen and Rosset (2009) showed that variance in endorsing intuitive but erroneous teleological explanations was uniquely explained with the behavioral Stroop task (ibid, p. 141), and similarly, Svedholm and Lindeman (2013) showed that a measure of ontological confusions (argued by these authors to be based on intuitions) was strongly correlated with the behavioral Stroop task (albeit using a different configuration of task conditions and coding than those used by Kelemen & Rosset, 2009). In addition, Zaitchik, Iqbal, and Carey (2014) found evidence that executive functions, including inhibition, uniquely explained variance in performance, controlling for age and verbal IQ, on a biological reasoning task in 5- to 7-year-olds. Zaitchik, Iqbal, and Carey argued that executive functions are necessary for both the acquisition of beliefs about biology at this age range, and for performance on their task in children who have acquired these beliefs (this is because to perform accurately on their task children had to resolve a conflict between their intuitions about biology and newly acquired biological beliefs).

On the other hand, Kelemen et al. (2012) found no relationship between teleological intuitions and the behavioral Stroop task in their study of college students, professional scientists, and community members (ibid, p. 1079).

The results reported here using task conditions and coding similar to those adopted by Kelemen and colleagues in Experiment 1, and Svedholm and Lindeman (2013) in Experiment 2, repeatedly failed to show a relationship between performance on the sentence verification task and the behavioral Stroop task, further muddying the pattern of results on this question. We think it likely that this failure to find a correlation between performance on the sentence verification task and the behavioral Stroop task is due to the fact that executive inhibition can be measured in a variety of ways and, more broadly, a range of different executive functions jointly contribute to performance on complex tasks (e.g., see the latent variable analysis and thorough discussion in Miyake et al., 2000). In the future, researchers might consider the use of a more varied and rigorous battery of inhibitory and other executive functions measures.

5.4. No correlations with practice

A final prediction derived from the coexistence hypothesis which was tested here was that practice with acquired beliefs would attenuate the effect of interference from inconsistent intuitions. Experiment 1 and the time limit condition of Experiment 2 found that regular church attendance was positively correlated with accuracy on consistent and inconsistent religion items; however, none of the practice measures used correlated with accuracy or response time interference scores in either the religion or science domains. The failure to find such an effect in this study parallels that of Shtulman and Valcarcel

(2012). In contrast, the investigation by Kelemen and Rosset (2009) in a similar population of university undergraduates did find a relationship between science education (indexed by questionnaires on geoscience and natural selection) and a tendency to make teleological errors.

A different way to evaluate practice with acquired beliefs, which has been used in previous studies, is to compare novices with experts (e.g., university undergraduates with professors); however, these studies are equivocal. Goldberg and Thompson-Schill (2009) find a smaller bias in preferentially ascribing animacy to animals than to plants in biology professors versus undergraduates, but they find similar biases in these groups in, for example, preferentially ascribing animacy to moving than to nonmoving artifacts, which are categories not studied in biology departments. More recently, Shtulman and Harrington (2016) found, using a sentence verification task identical to that of Shtulman and Valcarcel (2012), that the accuracy difference between consistent and inconsistent science items was lower in science professors (9%) than in humanities professors (13%), and in turn lower in humanities professors than in community members of similar ages (20%).

The findings of Shtulman and Harrington (2016) and Goldberg and Thompson-Schill (2009) suggest that practice with acquired beliefs can explain differences in interference, albeit within circumscribed domains. On the other hand, while Kelemen et al. (2012) demonstrated that the tendency to make teleological errors was greater in undergraduates than in science professors, they did not find such a difference between science and humanities professors; this suggests that differences between undergraduates and professors other than differences in practice might be responsible for differences in interference between these groups. The observed difference between humanities professors and community members in Shtulman and Harrington (2016) is compatible with this interpretation as well. In the future, researchers might consider the use of more nuanced indices and combinations of multiple indices to further investigate questions pertaining to the effects of practice on interference during coexistence; for instance, both questionnaires as in Kelemen and Rosset (2009) and populations that more strongly differ in practice or expertise as in Kelemen et al. (2012) and Goldberg and Thompson-Schill (2009).

5.5. Implications to theories of religious beliefs and behavior, and future directions

The evidence provided in this study for the coexistence hypothesis has significant implications to foundational theories of the cultural transmission of supernatural concepts, including those concepts deemed religious. Boyer's "cognitive optimum" theory (1994a,b, 2001; Boyer & Ramble, 2001) postulates that the historical and cross-cultural prevalence of specific supernatural beliefs (e.g., beliefs in agents with extraordinary mental characteristics) is in part due to a memory and transmission advantage of beliefs which are inconsistent with core intuitions (there termed "counterintuitive" beliefs). The current study suggests that, as required by the cognitive optimum theory, conflicts between counterintuitive beliefs and core intuitions cannot be permanently resolved, and counterintuitive beliefs may therefore retain their memory and transmission advantage within and between individual minds.

Additionally, the coexistence hypothesis explains two "on-the-ground" observations by social scientists and humanists (see Sperber, 1985, for an early discussion of these observations under the more general question of why people hold "irrational" beliefs): discrepancies between an individual's reported beliefs, and between reported beliefs and behavior (also see Slone, 2004). For example, Christian religious believers often simultaneously describe God as person-like (e.g., loving, fallible) and abstract (not able to be described with human emotion terms and infallible), and although God is in Christian theology all-knowing, believers nonetheless tell Him their prayers (the contradiction, of course, is that if God is all-knowing, believers do not need to tell Him anything—He already knows everything; see Lane, Evans, Brink, & Wellman, 2016, for children's developing understanding of extraordinary communication). Future studies are needed to further investigate the "on-the-ground" conditions under which core intuitions versus acquired theological beliefs are verbalized, and the variable role of each in regulating religious behavior.

Finally, while in this study we assumed that the person concept is co-opted to form a representation of God (e.g., see Boyer, 2001; Boyer & Ramble, 2001), at least one other hypothesis is potentially compatible with our findings: Bloom (2004) and others (e.g., Shtulman, 2008; Shtulman and Lindeman, 2016) hypothesized that the concept co-opted to form representations of supernatural agents, including the Christian God, is that of a disembodied mind (however, see Hodge, 2008, unpublished data). In this study, we did not carefully differentiate between different characteristics of God, but post hoc analyses on statements pertaining to God's physicality versus psychology showed that the reported effects held for both (which is compatible with co-option of the person concept and not of a disembodied mind concept). However, we hesitate to draw strong conclusions on this question, since the current study was not designed to evaluate it.

In any case, the findings we report regarding coexistence are independent from hypotheses about the exact concept that is being co-opted. Future studies are needed to further investigate a range of interesting specific hypotheses, including those regarding the different possible concepts which might be co-opted in representations of God and other supernatural entities—including the possibility of individual differences in the concept co-opted for a given supernatural entity—in both Christianity and other religious traditions (see, for example, the work of Cohen, 2007, 2008, who hypothesized that among believers in the Afro-Brazilian syncretic cult Candomblé, representations of certain possessing spirits co-opt the pathogen concept).

In conclusion, beyond providing novel evidence for the coexistence hypothesis in the domain of religion, and bringing attention to the theoretical parallels in coexistence in religion and science, we aim to highlight the explanatory utility of cognitive science to religious phenomena. Recent reviews have disproportionately focused on the present limitations of cognitive science in explaining supernatural beliefs and beliefs deemed religious (e.g., see Purzycki & Willard, 2015, and the commentaries to their article); with the current study we aim to reorient focus to the contributions cognitive science can make to the study of religion (and culture more broadly) and to possible novel directions of empirical and theoretical investigation.

Acknowledgments

The authors thank Ann Taves, Chris Morales, and other members of the Religion, Experience, and Mind lab group at the University of California, Santa Barbara, for early conceptual advice; Sangin Kim for statistical advice; Andrew Shtulman and two anonymous reviewers for thoughtful and thorough feedback on earlier versions of this manuscript; and staff of the Goleta and Santa Barbara churches that facilitated participant recruitment and testing for this study.

Notes

1. We do not evaluate in this study whether acquisition of these beliefs constitutes conceptual change or belief revision (see Carey, 2009, for a thorough discussion of this distinction).
2. Sperber and colleagues further hypothesize that (c) acquired beliefs that are inconsistent with core intuitions exist in a specialized metarepresentational "bubble" which isolates them from core concepts that exist in a mental database of beliefs. In contrast to the database of beliefs which can be accessed unconsciously and spontaneously, metarepresentations can only be accessed consciously (e.g., Sperber, 1997, 2000; Mercier & Sperber, 2009; for a related discussion of dual-process theory, see Evans, 2003, 2008; Evans & Stanovich, 2013; also see Mercier & Sperber, 2011). We do not evaluate this hypothesis which, while interesting, is beyond the scope of the experiments reported here.
3. A second objection raised by Shtulman concerns the overall rate of anthropomorphic responding in the studies by Barrett and colleagues for narratives involving religious agents, which he points out was (a) lower than 100% and (b) not substantially different from the rate of anthropomorphic responding observed for narratives involving a supercomputer (a nonreligious agent with full information access; ibid, p. 1125). However, a problem with evaluating this objection is that there is no consensus on what amount of anthropomorphism—beyond demonstrating that it exists—is required to support the coexistence hypothesis, or how minimal such responding needs to be to refute it.
4. The four items constituting one of the groups - group T12 in the Supplementary Material - were, due to experimenter error, of slightly uneven lengths in words. All response patterns reported in this study remain the same if this group of items is excluded.
5. Shapiro–Wilk tests revealed that several of the response accuracy distributions were not normal (consistent religion items in the unspeeded instructions condition: $SW = .455$, $df = 20$, $p < .001$; consistent and inconsistent religion items in the speeded instructions condition: $SW = .599$, $df = 19$, $p < .001$, $SW = .681$, $df = 19$, $p < .001$, respectively). The simple main effects were therefore analyzed with

nonparametric tests to supplement the parametric tests reported here. A series of planned comparisons using the Wilcoxon signed-rank test confirmed all findings reported here (all *Z*s less than −4.83, all *p*s < .001).

6. A very small number of response time data points (<1%) were removed for being more than 3 *SD* above or below the mean response time.
7. Three participants did not complete the behavioral Stroop task, and two participants' scores were removed for being more than 2 *SD* above the mean Stroop response time interference score; two participants' scores were removed for being more than 2 *SD* above or below the mean working memory score.
8. In the time limit condition, responses that were timed out were included in the analysis and considered incorrect. The response patterns reported in this section remained the same if instead responses that were timed out were excluded from the analysis.
9. Shapiro–Wilk tests revealed that several of the response accuracy distributions were not normal (consistent and inconsistent religion items in the no time limit condition: $SW = .731$, $df = 32$, $p < .001$, $SW = .872$, $df = 32$, $p = .001$, respectively; consistent religion items and inconsistent science items in the time limit condition: $SW = .874$, $df = 39$, $p < .001$, $SW = .941$, $df = 39$, $p < .05$, respectively). The simple main effects were therefore analyzed with nonparametric tests to supplement the parametric tests reported here. A series of planned comparisons using the Wilcoxon signed-rank test confirmed all findings reported here (all *Z*s < −6.67, all *p*s < .001).
10. A separate examination of the religion and science items each via a 2 (Consistency: Consistent vs. Inconsistent) × 2 (Condition: Time Limit vs. No Time Limit) repeated measures ANOVA further confirmed that despite the absence of a three-way interaction, the two-way interaction between Consistency and Condition was carried by the religion items (religion: $F_{1,69} = 5.1$, $p < .05$, partial $\eta^2 = .07$; science: $F_{1,69} = .22$, p = n.s., partial $\eta^2 = .00$). The results from this mixed ANOVA were additionally replicated with an extended linear-mixed effects model, which accommodated different error variances between the two Consistency factors (the error variance was greater for consistent than for inconsistent items in both Condition factors). The interaction for Consistency and Condition for the religion domain remained significant at $p < 0.05$.
11. A very small number of response time data points (<2%) were removed for being more than 3 *SD* above or below the mean response time.
12. Response times in the speeded condition were truncated by the time limit manipulation and were therefore excluded from this analysis.
13. Seven participants did not complete the behavioral Stroop task, and three participant's scores were removed for being more than 2 *SD* above the mean Stroop response time interference score; four participants' scores were removed for being more than 2 *SD* above or below the mean working memory score.
14. A very small number of response time data points (<2%) were removed for being more than 3 *SD* above or below the mean response time.

References

Baillargeon, R. (2004). Infants' physical world. *Current Directions in Psychological Science*, *13*(3), 89–95.

Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, *67*, 159–186. doi:10.1146/annurev-psych-010213-115033

Barrett, J. L. (1998). Cognitive constraints on Hindu concepts of the divine. *Journal for the Scientific Study of Religion*, *37*(4), 608–619.

Barrett, J. L. (1999). Theological correctness: Cognitive constraint and the study of religion. *Method & Theory in the Study of Religion*, *11*, 325–339.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences*, *4*(1), 29–34.

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

Baumard, N., & Boyer, P. (2013). Religious beliefs as reflective elaborations on intuitions : A modified dual-process model. *Current Directions in Psychological Science*, *22*(4), 295–300. doi:10.1177/0963721413478610

Bloom, P. (2004). *Descartes' Baby: How the science of child development explains what makes us human*. New York: Basic Books.

Boyer, P. (1994a). Cognitive constraints on cultural representations: Natural ontologies and religious ideas. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain-specificity in culture and cognition* (pp. 391–411). New York: Cambridge University Press.

Boyer, P. (1994b). *The naturalness of religious ideas: A cognitive theory of religion*. Berkeley, CA: University of California Press.

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

Boyer, P., & Ramble, C. (2001). Cognitive templates for religious concepts: Cross-cultural evidence for recall of counter-intuitive representations. *Cognitive Science*, *25*, 535–564.

Broadway, J., & Engle, R. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*(2), 563–570.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Cohen, E. (2007). *The mind possessed: The cognition of spirit possession in an Afro-Brazilian religious tradition*. New York: Oxford University Press.

Cohen, E. (2008). What is spirit possession? Defining, comparing, and explaining two possession forms. *Ethnos*, *73*(1), 101–126.

Epley, N., Converse, B. A., Delbosc, A., Monteleone, G. A., & Cacioppo, J. T. (2009). Believers' estimates of God's beliefs are more egocentric than estimates of other people's beliefs. *Proceedings of the National Academy of Sciences*, *106*(51), 21533–21538. doi:10.1073/pnas.0908374106

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. doi:10.1016/j.tics.2003.08.012

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278. doi:10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241. doi:10.1177/1745691612460685

Goldberg, R. F., & Thompson-Schill, S. L. (2009). Developmental "roots" in mature biological knowledge. *Psychological Science*, *20*(4), 480–487. doi:10.1111/j.1467-9280.2009.02320.x

Granqvist, P., & Kirkpatrick, L. A. (2008). Attachment and religious representations and behavior. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 906–933). New York: Guilford Press.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. doi:10.1126/science.1134475

Heiphetz, L., Lane, J. D., Waytz, A., & Young, L. L. (2016). How children and adults represent God's mind. *Cognitive Science*, *40*, 121–144. doi:10.1111/cogs.12232

Hodge, K. M. (2008). Descartes' mistake: How afterlife beliefs challenge the assumption that humans are intuitive Cartesian substance dualists. *Journal of Cognition and Culture*, *8*, 387–415. doi:10.1163/156853708X358236

Hodge, K. M. (Unpublished Data). Theory of mind, mind-body dualism, and their (mis) application in the cognitive science of religion.

Hyde, K. E. (1990). *Religion in childhood and adolescence: A comprehensive review of the research*. Birmingham, AL: Religious Education Press.

Inagaki, K., & Hatano, G. (2002). *Young children's naıve thinking about the biological world*. New York, NY: Psychology Press.

Inagaki, K., & Hatano, G. (2006). Young children's conception of the biological world. *Current Directions in Psychological Science*, *15*, 177–181. doi:10.1111/j.1467-8721.2006.00431.x

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), 696–701. doi:10.1371/journal.pmed.0020124

Kelemen, D., & Rosset, E. (2009). The human function compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143. doi:10.1016/j.cognition.2009.01.001

Kelemen, D., Rottman, J., & Seston, R. (2012). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*(4), 1074–1083. doi:10.1037/a0030399

Kirkpatrick, L. A. (2005). *Attachment, evolution, and the psychology of religion*. New York: Guilford.

Lane, J. D., Evans, E. M., Brink, K. A., & Wellman, H. M. (2016). Developing concepts of ordinary and extraordinary communication. *Developmental Psychology*, *52*(1), 19–30. doi:10.1037/dev0000061

Lane, J. D., Wellman, H. M., & Evans, E. M. (2010). Children's understanding of ordinary and extraordinary minds. *Child Development*, *81*(5), 1475–1489. doi:10.1111/j.1467-8624.2010.01486.x

Lane, J. D., Wellman, H. M., & Evans, E. M. (2012). Sociocultural input facilitates children's developing understanding of extraordinary minds. *Child Development*, *83*(3), 1007–1021. doi:10.1111/j.1467-8624.2012.01741.x

Lane, J. D., Wellman, H. M., & Evans, E. M. (2014). Approaching an understanding of omniscience from the preschool years to early adulthood. *Developmental Psychology*, *50*(10), 2380–2392.

Lindeman, M., & Aarnio, K. (2007). Superstitious, magical, and paranormal beliefs: An integrative model. *Journal of Research in Personality*, *41*(4), 731–744. doi:10.1016/j.jrp.2006.06.009

Luhrmann, T. M. (2012). *When God talks back: Understanding the American Evangelical relationship with God*. New York: Alfred A. Knopf. doi:10.14318/hau3.3.021

Mercier, H., & Sperber, D. (2009). Intuitive and reflective inferences. In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 147–170). New York: Oxford University Press. doi:10.1093/acprof:oso/9780199230167.001.0001

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–74. http://doi.org/10.1017/S0140525X10000968

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi:10.1006/cogp.1999.0734

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. doi:10.1027/1864-9335/a000192

Purzycki, B. G., & Willard, A. K. (2015). MCI theory: A critical discussion. *Religion, Brain & Behavior*, *1–42*, doi:10.1080/2153599X.2015.1024915

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. doi:10.1027/1015-5759/a000123

Riekki, T., Lindeman, M., & Hood, B. M. (2011). Is weaker inhibition associated with supernatural beliefs? *Journal of Cognition and Culture*, *11*, 231–239. doi:10.1163/156853711X570038

Shtulman, A. (2008). Variation in the anthropomorphization of supernatural beings and its implications for cognitive theories of religion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34* (5), 1123–1138. doi:10.1037/0278-7393.34.5.1123

Shtulman, A., & Harrington, K. (2016). Tensions between science and intuition across the lifespan. *Topics in Cognitive Science*, *8*(1), 118–137.

Shtulman, A., & Lindeman, M. (2015). Attributes of God: Conceptual foundations of a foundational belief. *Cognitive Science*, *40*(3), 635–670. doi:10.1111/cogs.12253

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*, 209–215. doi:10.1016/j.cognition.2012.04.005

Slone, D. J. (2004). *Theological incorrectness: Why religious people believe what they shouldn't*. New York: Oxford University Press.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56. doi:10.1016/0364-0213(90)90025-R

Sperber, D. (1985). *On anthropological knowledge: Three essays*. Cambridge, UK: Cambridge University Press.

Sperber, D. (1996). *Explaining culture: A naturalistic approach*. Oxford, UK: Blackwell Publishers Ltd.

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind and Language*, *12*(1), 67–83.

Sperber, D. (Ed.) (2000). *Metarepresentations: A multidisciplinary perspective*. New York: Oxford University Press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Svedholm, A. M., & Lindeman, M. (2013). The separate roles of the reflective mind and involuntary inhibitory control in gatekeeping paranormal beliefs and the underlying intuitive confusions. *British Journal of Psychology*, *104*, 303–319. doi:10.1111/j.2044-8295.2012.02118.x

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535–585. doi:10.1016/0010-0285(92)90018-W

Wason, P. C. (1959). The processing of positive and negative information. *Quarterly Journal of Experimental Psychology*, *11*(2), 92–107.

Zaitchik, D., Iqbal, Y., & Carey, S. (2014). The effect of executive function on biological reasoning in young children: An individual differences study. *Child Development*, *85*(1), 160–175.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Full list of religion and science statements.