COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 43 (2019) e12784

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12784

The Embodied God: Core Intuitions About Person Physicality Coexist and Interfere With Acquired Christian Beliefs About God, the Holy Spirit, and Jesus

Michael Barlev,[a] Spencer Mermelstein,[b] Adam S. Cohen,[c] Tamsin C. German[b]

[a]*Department of Psychology, Arizona State University*
[b]*Department of Psychological & Brain Sciences, University of California, Santa Barbara*
[c]*Department of Psychology, University of Hawai'i at Mānoa*

Received 17 July 2018; received in revised form 30 July 2019; accepted 1 August 2019

Abstract

Why are disembodied extraordinary beings like gods and spirits prevalent in past and present theologies? Under the intuitive Cartesian dualism hypothesis, this is because it is natural to conceptualize of minds as separate from bodies; under the counterintuitiveness hypothesis, this is because beliefs in minds without bodies are unnatural—such beliefs violate core knowledge intuitions about person physicality and consequently have a social transmission advantage. We report on a critical test of these contrasting hypotheses. Prior research found that among adult Christian religious adherents, intuitions about person psychology coexist and interfere with theological conceptualizations of God (e.g., infallibility). Here, we use a sentence verification paradigm where participants are asked to evaluate as true or false statements on which core knowledge intuitions about person physicality and psychology and Christian theology about God are inconsistent (true on one and false on the other) versus consistent (both true or both false). We find, as predicted by the counterintuitiveness hypothesis but not the Cartesian dualism hypothesis, that Christian religious adherents show worse performance (lower accuracy and slower response time) on statements where Christian theological doctrines about God's physicality (e.g., incorporeality, omnipresence) conflict with intuitions about person physicality. We find these effects for other extraordinary beings in Christianity—the Holy Spirit and Jesus—but not for an ordinary being (priest). We conclude that it is unintuitive to conceptualize extraordinary beings as disembodied, and that this, rather than inherent Cartesian dualism, may explain the prevalence of beliefs in such beings.

Keywords: God; Holy Spirit; Religious beliefs; Dualism; Disembodied mind; Counterintuitiveness; Representational coexistence

Correspondence should be sent to Michael Barlev, Department of Psychology, Arizona State University, PO Box 871104, Tempe, AZ 85287-1104, USA. E-mail: mlbarlev@gmail.com

1. Introduction

Why are disembodied beings prevalent in past and present theologies? Abrahamic religions, particularly their mystical traditions, focus not only on an incorporeal and omnipresent God, but on a plethora of disembodied spiritual beings such as angels and demons. The Afro-Brazilian syncretic religion Candomblé focuses on disembodied spirits (*entidades* or "entities") that can possess human bodies (Cohen, 2007). While in Candomblé possessing spirits are to be worshipped, in Abrahamic traditions possessing spirits also exist, but they are evil demons that are to be ritualistically exorcised (e.g., *dybbuk* in Judaism). Japanese practitioners of the traditional religion Shintoism worship animistic spirits and the spirits of the deceased (*kami*) in special shrines. Lastly, adherents of many New Age spiritual movements focus on interactions with a spiritual world including with spirits of the deceased, and on spiritual experiences such as out-of-body experiences wherein the mind of the experiencer is believed to temporarily leave her or his body (Barlev, Kinsella, German, Taves, & Paloutzian, 2015; Kinsella, 2017). What explains the prevalence of beliefs in beings without physical bodies, and related beliefs such that these beings can enter human bodies and that the mind of a person can temporarily leave her or his body?

The most prominent explanation was advanced by Bloom (2005). Bloom begins with the theory, corroborated by decades of research on babies, according to which the brain consists of evolved and reliably developing core knowledge mechanisms (e.g., Carey, 2009; Spelke & Kinzler, 2007). While perceptual modalities carve up the world around us, the brain consists of additional inferential mechanisms—core knowledge mechanisms—that go beyond what is perceived. When you throw a projectile, you must predict its trajectory of motion through space. For example, you predict that its trajectory will be spatio-temporally continuous—that it will move in a connected path and will continue to exist even when out of sight—and that it could not occupy the same physical space as another object at the same time. Infants understand these and other core properties of the physicality and spatio-temporal mechanics of objects, such as that objects are cohesive (i.e., objects are bounded wholes, and neither separate nor join together), and that objects act on each other if and only if they come into contact (i.e., objects do not interact at a distance). See Baillargeon (2004) and Spelke (1990) for reviews.

Our world additionally consists of "intentional agents" or entities the observed behavior of which can be explained and predicted by unobservable mental states, rather than by the physics of objects moving through space. Before a hunter throws a spear, you can predict where the spear will be thrown by attributing to the hunter the goal of killing an animal that is in his or her sight. Infants understand that the behavior of intentional agents can be explained by goals, and infants predict that intentional agents will act in accordance with those goals (Woodward, 1998) and will achieve those goals through action that is rational (Csibra et al., 1999). See Baillargeon, Scott, and Bian (2016) for a review.

The adult core person concept or template consists of such inferences about "minds" or the mental states of persons ("folk psychology") and their "bodies" or their physicality

and spatio-temporal mechanics ("folk physics"). The core person concept reliably develops in all neurotypical humans from a skeletal set of core knowledge inferences that are present in infancy and from associated learning mechanisms (e.g., Baillargeon, 2004; Baillargeon, Scott, & Bian, 2016; Carey, 2009; Spelke, 1990; Spelke & Kinzler, 2007).

Bloom suggests that although adults by default represent persons as embodied, it is intuitive or natural for us to conceptualize of minds as separate from bodies, and it is around such conceptions that more elaborated beliefs about disembodied beings like the Christian God form. Bloom (2005) is here building on the suggestion that representations of extraordinary beings are initially formed by co-opting the person concept (e.g., Barrett, 1998, 1999; Barrett & Keil, 1996; Boyer, 2001). The developmental findings by Lane and colleagues on how children represent God's mind are compatible with this suggestion. For example, Lane, Wellman, and Evans (2010) showed that children younger than 5 conceptualize the psychological characteristics of God and of persons similarly: On verbal response tasks, children younger than 5 who explicitly attributed constrained knowledge to persons (e.g., their mom) did also to God (for a review see Heiphetz, Lane, Waytz, & Young, 2016). Thus, according to intuitive or default Cartesian dualism, beliefs in disembodied beings are prevalent because they are a byproduct of the way in which core knowledge mechanisms inherently conceive of the relation between minds and bodies (also see, e.g., Bloom, 2007; Forstmann & Burgmer, 2015; 2017; Hood, Gjersoe, & Bloom, 2012; Kuhlmeier, Bloom, & Wynn, 2004).

M. Barlev and A. Shtulman (under review) similarly suggest that representations of disembodied beings are formed by co-opting the person concept, but, contra inherent Cartesian dualism, and following models on the relationship between folk psychology and folk physics such as by Leslie (1994), that it is not natural to represent persons as disembodied. In fact, M. Barlev and A. Shtulman (under review) suggest that representations of disembodied persons are inconsistent with core intuitions about person physicality and are therefore counterintuitive (Boyer, 2001). The prevalence of such representation may at least partially be explained by this (expanding on the framework advanced by Sperber, 1997, 2000, and Boyer, 2001): Representations of persons as disembodied are quarantined in a specialized meta-representational "bubble" where they recruit attention and memory (Banerjee, Haque, & Spelke, 2013; Barrett & Nyhof, 2001; Boyer & Ramble, 2001) as people try to reconcile them with their pre-existing beliefs, which gives them a social transmission advantage. That is, according to this epidemiological model (Sperber, 1996), representations of disembodied persons are culturally prevalent precisely because they are counterintuitive.[1]

At the crux of these contrasting hypotheses then is whether it is natural or not to represent persons or extraordinary beings such as the Christian God as disembodied. This study aims to explore this question using a task modified from Barlev, Mermelstein, and German (2017, 2018). Previously, Barlev et al. tested the hypothesis that in the minds of adult Christian religious adherents, representations of God based on core person psychology coexist and interfere with acquired theology about God. Barlev et al. did not specifically focus on core person physicality and therefore do not speak to whether core knowledge intuitions about person physicality interfere with acquired theology about God

as, for example, omnipresent (present everywhere at the same time) or incorporeal (has no physical or bodily existence).

Barlev et al. used a statement verification paradigm where participants evaluated as true or false statements that were consistent or inconsistent between core person psychology and acquired Christian theology about God (e.g., "God has beliefs that are true" is true according to both intuition and theology [consistent]; "God has beliefs that are false" is true intuitively but false theologically [inconsistent]). The task instructions asked participants to evaluate each statement according to theology as accurately and as quickly as they could. The logic of this sentence evaluation paradigm was as follows: If representations of God are based on the person template, then core knowledge intuitions about persons may interfere with acquired Christian theology about God on items where these are inconsistent, but not where they are consistent. The representational interference or conflict on inconsistent items needs to be resolved, and it was predicted that failures to do so will yield more incorrect responses on inconsistent items as compared to consistent items; it was additionally predicted that this process will yield slower responding on inconsistent items (including on correct responses, i.e., when this representational conflict was successfully resolved). Barlev et al. found behavioral evidence compatible with just such conflict: worse performance—lower accuracy and slower response time (RT)—on inconsistent versus consistent statements (Barlev et al., 2017, 2018), including in Christian religious believers with a lifetime of religious adherence (Barlev et al., 2018).

This study uses the same experimental paradigm but with the addition of statements targeting God's physicality (e.g., "God is able to be at my church and at other churches" which is true intuitively and theologically [consistent], and "God is at my church when He is not at other churches" which is true intuitively but false theologically [inconsistent]). If representations of God are based on the person template and it is unnatural to represent persons as disembodied, in line with the counterintuitiveness hypothesis, then adult Christian religious adherents should show representational coexistence and interference on statements targeting God's physicality as well as psychology. But if it is natural to represent persons as disembodied minds, in line with the intuitive Cartesian dualism hypothesis, then participants should show representational interference on statements targeting God's psychology only. We report on a critical test of these alternatives in Exp. 1 and 2.

In Exp. 3, we explore the primacy of the person concept in forming representations of extraordinary beings more broadly. While in mainstream Christian theology members of the Trinity are believed to be consubstantial (of the same essence), they are also believed to be three distinct beings. Recently, Sharp, Rentfrow, and Gibson (2017) suggested that Christian religious adherents personify members of the Trinity to varying extents: Jesus—theologically believed to be God incarnate in a human body—was conceptualized as most like a person, and the Holy Spirit was conceptualized as least like a person. We wondered whether, given this, the formation of the concept of the Holy Spirit, in contrast to the concepts of God and Jesus, did not co-opt the person template. We explore this possibility in Exp. 3 by, in addition to a God condition (replication of Exp. 1), including conditions in which the target extraordinary being is the Holy Spirit or Jesus. If the Holy

Spirit co-opts the person template, then participants should show representational interference (indexed by accuracy and RT performance on the sentence verification task) on statements targeting the Holy Spirit's psychology and physicality; but if the Holy Spirit does not co-opt the person template, then they might not show representational interference on either.

Finally, in Exp. 4, we provide a check for the possibility that any observed findings are due to low-level biases in the structure of the statements used here, rather than to representational coexistence and interference. In this control experiment, "God" is replaced with "my priest."[2] If the observed pattern of findings is due to such low-level biases, then statements where lower accuracy and higher RT were observed for God (Exp. 1–3) should show the same effects for "my priest"; alternatively, if the pattern of findings in Exp. 1–3 is due to representational interference and coexistence, then there should not be such effects on statements where the extraordinary being was replaced with the ordinary being "my priest."

2. Experiment 1

2.1. Methods

2.1.1. Power analysis

A priori power analyses for this and all other experiments reported here were computed using G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). The RT effects in past experiments using the same sentence verification task (Barlev, Mermelstien, & German, 2017, 2018) were significantly smaller than the accuracy effects, so the power analyses were computed on the basis of the RT effects. The expected RT effect sizes were in the range of $d = 0.30$–0.40. However, unlike in past experiments which were administered to undergraduate students and community members in the laboratory, this experiment was administered to Amazon Mechanical Turk (MTurk) workers online. We, therefore, to be conservative, used the lower end of this range ($d = 0.30$). This showed that for a power of 95% and an alpha of 5%, we needed, at minimum, 122 participants.

2.1.2. Participants

Participants were Christian religious adherents recruited from MTurk. Participants completed a brief survey for a small monetary reward, with those meeting the prescreening criteria invited to participate in a longer survey for a bonus of $1 (this is commensurate with payments on MTurk). Participants were not told what the prescreening criteria were. Participants were prescreened to have grown up, and currently identify, as Roman Catholics, Mainline Protestants, or Protestant Evangelicals (all mainstream Christian denominations that are well represented on MTurk), and to identify as at least slightly religious and slightly spiritual—this pre-screening strategy was to increase the likelihood that participants currently believed in mainstream Christian theology. Participants were excluded for having religion Sentence Verification Task scores that were not statistically

different from chance, because in previous studies we found that chance responding was suggestive of not believing in mainstream Christian theology or of otherwise having an idiosyncratic representation of God, or of not believing in the existence of God at all (see Constraints on Generality for details). We chose this exclusion strategy because it is conservative in assigning equal weight to responses on consistent and inconsistent items; across the four experiments reported here, a total of 36 out of 470 (7.66%) participants were excluded on the basis of this. Participants were also excluded for having taken the study on a smartphone device rather than on a personal computer, because the different device interface on a smartphone may have caused substantially noisier RTs on the Sentence Verification Task (given the software with which the Sentence Verification Task was administered, smartphones required touchscreen responding, while personal computers, even those with touchscreen capabilities, defaulted to keyboard responding).

The final sample of $N = 127$ (68% F) had a mean age of 41 (range 18–74). Participants identified as White (85%), Hispanic or Latino (7%), East, Southeast, or South Asian (6%), and Black (2%). Participants identified as Roman Catholic (31%), Protestant Evangelical (35%), and Mainline Protestant (35%). Eighty-three percent currently identified with the religious tradition with which they grew up. Finally, on a scale of 0 = Not at all to 100 = Very, participants identified as $M = 74$ ($SD = 21$) on religiosity and $M = 78$ ($SD = 19$) on spirituality ($r = .82$, $p < .001$), and on the Christian Orthodoxy Scale (Hunsberger, 1989; also see Fullerton & Hunsberger, 1982), which is conceptualized as a measure of the extent to which respondents accept the central doctrines of Christian theology, participants scored $M = 2.44$ ($SD = 0.82$; from -3 = Strongly disagree to 3 = Strongly agree).

2.1.3. Design

The IV was the characteristic of God (physicality vs. psychology) tested within subjects. The DVs were the accuracy and RT interference scores (calculated as the difference in performance between consistent and inconsistent statements).

2.1.4. Materials

Statements ($N = 52$) about the psychology and physicality of God were constructed in quartets, with each quartet concerning a particular theological doctrine about God. The doctrines chosen, to increase the likelihood that they will be known to adult religious adherents, were very common Christian theological doctrines that are shared among mainstream Christian denominations: for example, infallibility (that God does not have false beliefs), omnipresence (that God is present everywhere at the same time), or incorporeality (that God has no physical or bodily existence); studies suggest that the acquisition of Christian theological doctrines such as infallibility begins as early as the preschool years (e.g., Lane et al., 2010). In each quartet there was a pair of **consistent** statements (true according to both intuitions about persons and Christian theology about God, or false according to both) and a pair of **inconsistent** statements (true intuitively but false theologically, or false intuitively but true theologically). Thus, within each quartet consistent statements served as a baseline with which the inconsistent statements were compared. Additionally, within each quartet there were two true and two false statements according to Christian theology. The four statements

within each quartet were balanced in terms of overall sentence structure, complexity, and length in words. See Table 1 for sample statements.

Correct responses to statements about the psychology and physicality of God were coded based on the theology level (e.g., "false" responses to the statement "God is at my church when He is not at other churches" were coded as correct). Accuracy and RT interference scores were calculated as the mean difference between consistent and inconsistent statements, such that performance on consistent statements was a baseline with which performance on inconsistent statements was compared. Accuracy and RT interference scores different from zero are interpreted as the presence of an effect, and scores above zero are interpreted as worse performance (lower accuracy and/or slower RT) on inconsistent versus consistent statements.

2.1.5. Procedure

Participants were tested online using Inquisit 5 software. The scripts for all experiments reported here are available through the Open Science Framework (OSF): https://osf.io/428ca/?view_only=2f0d422d98be4a3bb2ba9ec0930423ec. The study lasted between 15 min to an hour, with an average of about 30 min. Instructions to the Sentence Verification Task emphasized both response accuracy and speed. Items were presented one by one and in a randomized order. Responses were collected via key presses. Participants were instructed to respond with their dominant hand, and whether the index or ring finger was used to respond "true" or "false" was randomized between participants.

2.2. Results

We report Bayesian analyses in the main manuscript and Frequentist analyses in Appendix S1. All analyses for the experiments reported here were performed using R 3.4.3

Table 1
Sample person physicality and psychology statements

Domain	Consistency	Intuition	Theology	Statements
Physicality	Consistent	T	T	God is able to be at my church and at other churches
		F	F	God is never at my church nor is He at other churches
	Inconsistent	T	F	God is at my church when He is not at other churches
		F	T	God is at all times at my church and at other churches
	Consistent	T	T	God can occupy the physical space inside a church
		F	F	God can never occupy the physical space inside a church
	Inconsistent	T	F	God can never occupy the physical space inside a boulder
		F	T	God can occupy the physical space inside a boulder
Psychology	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

Note. Consistent statements are true on both intuition and theology; inconsistent statements are true on one and false on the other. Psychology statements are modified from Barlev et al. (2017).

and JASP 0.8.5.1 (JASP Team, 2017). A Bayes factor (BF) is a measure of relative evidence, quantifying how well data are predicted by one hypothesis relative to a competing hypothesis. If the hypotheses of interest are an alternative (H_1) and a null (H_0), then $BF_{10} = 6$ would mean the data are six times more likely under the alternative than the null. Whereas a *p*-value is typically used to arrive at a binary decision about a single null hypothesis (reject or fail to reject the null depending on whether $p < \alpha$), a BF is a continuous measure of evidence that can be used to update our beliefs about hypotheses in light of new data. For example, if we believe H_1 and H_0 are equally probable ahead of the data [$p(H_1) = 0.5$ and $p(H_0) = 0.5$], then $BF_{10} = 6$ indicates odds of 6:1 in favor of H_1, which, via Bayes Rule, would update our beliefs to $p(H_1|D) = 6/7 = 0.857$ and $p(H_0|D) = 1/7 = 0.143$. Although researchers may hold different beliefs about the prior probabilities p(H1) and p(H0), BF_{10} depends only on the data and simply tells us how to reallocate our confidence between two competing hypotheses, no matter what prior probabilities we assign them.

Although there are common guidelines for evaluating BFs (e.g., Jeffreys, 1961, suggest that values of 1–3 are anecdotal evidence in favor of H_1, values of 3–10 are moderate evidence in favor of H_1, and values of 10–30 are strong evidence in favor of H_1; also see Held & Ott, 2018), there are some advantages to eschewing such guidelines: (a) doing so avoids the categorical thinking that such guidelines invite, which is not so different from the binary thinking invited by *p*-value cutoffs, and (b) the reader can see how easy it is to go from BFs to belief updating (posterior probabilities).

Bayesian analyses were performed using the default priors set in JASP 0.8.5.1. The data files and R scripts for all experiments reported here are available through OSF.

First, data points that were above or below 3 *SD* from the mean RT of each statements were removed (in no experiment were these more than 3.5% of data points). Second, RT interference scores were calculated using both correct and incorrect responses. See Appendix S1 for multiverse analyses (Steegen, Tuerlinckx, Gelman, & Vanpaemel, 2016) that include other outlier removal strategies, including the strategies of not removing outliers at all and of calculating response interference using correct responses only.

The full list of items for all experiments along with accuracy and RT data are available in Appendix S1.

2.2.1. God is represented as an embodied person

We used a Bayesian paired sample *t* test to evaluate whether the accuracy and RT interferences scores on psychology and physicality were the same or different. While the null hypothesis that physicality and psychology interference scores were the same was more likely for accuracy ($BF_{10} = 0.15$), the experimental hypothesis that they were different was more likely for RT ($BF_{10} = 5.91$). The RT interference score for physicality was more than twice as large as that for psychology.

We then used Bayesian one-sample *t* tests to evaluate whether the accuracy and RT interferences scores on psychology and physicality were different from zero. The experimental hypothesis that they were was more likely than the null for both physicality (accuracy $BF_{10} > 300$, RT $BF_{10} > 300$) and psychology (accuracy $BF_{10} > 300$, RT $BF_{10} = 133.60$) see Figs. 1 and 2. The above findings are compatible with the hypothesis

that the acquired Christian theological God concept coexists and conflicts with an embodied person conception and not with the hypothesis that God is intuitively conceptualized as disembodied.

3. Experiment 2: Pre-registered replication of Exp. 1

3.1. Methods

3.1.1. Power analysis

Exp. 2 of this study was a pre-registered replication (https://osf.io/pf9vz/) with undergraduate students. Because access to participants was limited to those enrolled in a research methods course, the replication tested only for accuracy effects; in Exp. 1, as in our past experiments using the same sentence verification task, accuracy yielded a larger effect size than RT and would therefore be less likely to result in a Type II error. Additionally, factoring in the direction of the effect in Exp. 1, power was calculated for a one-sided test. (Note that the findings reported here are unchanged if the pre-registered one-sided tests are replaced with two-sided tests.) Although 95% power (and a two-sided test) would have been strongly preferred, the limited pool of participants required setting a lower value. With power set to 80%, effect size set to $d = 0.55$ (based on the combined accuracy effects for physicality and psychology items from Exp. 1), and alpha set to 5%, at minimum 22 participants were needed. After applying exclusion criteria, the final sample for the replication fell short of this, resulting in $n = 19$. A post hoc power analysis with $n = 19$, $d = 0.55$, and alpha = 0.05 for a one-tailed paired t test yielded power = 75%.

3.1.2. Participants

Participants were Christian religious adherents recruited from the University of Hawai‘i at Mānoa. Participants were enrolled in a research methods course at the time of

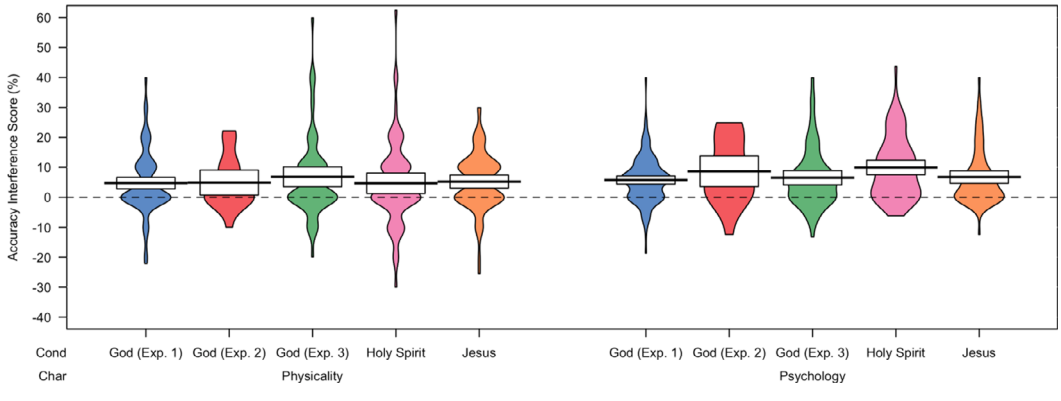

Fig. 1. Pirate plots of mean accuracy interference scores (%) on physicality and psychology items. Inference bands correspond to 95% CIs.

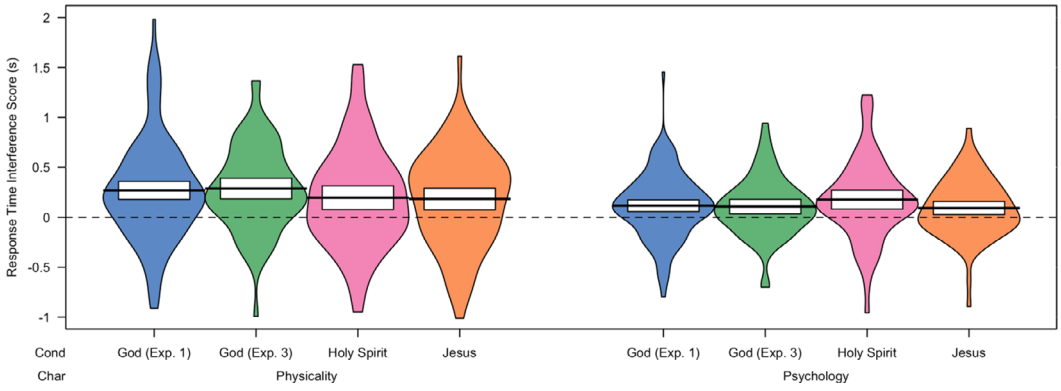

Fig. 2. Pirate plots of mean response time (RT) interference scores (s) on physicality and psychology items. Inference bands correspond to 95% CIs.

participation, and unlike the pre-screening process in all other experiments reported here, everyone in the course was tested and the exclusion criteria—identical to those in Exp. 1—were applied after data collection.

The final sample of $N = 19$ (68% F) had a mean age of 20 (range 18–26). Participants identified as White (16%), Hispanic or Latino (11%), East, Southeast, or South Asian (53%), Black (5%), and Other (16%). Participants identified as Roman Catholic (79%), Protestant Evangelical (11%), and Mainline Protestant (11%). Ninety-five percent currently identified with the religious tradition with which they grew up. Finally, participants identified as $M = 69$ ($SD = 21$) on religiosity and $M = 74$ ($SD = 23$) on spirituality ($r = .63$, $p < .005$). The Christian Orthodoxy Scale was not administered to these participants.

3.1.3. Design

The IV was, as in Exp. 1, the characteristic of God (physicality vs. psychology) tested within subjects. The DV was the accuracy interference score (calculated as the difference in performance between consistent and inconsistent statements).

3.1.4. Materials and procedure

The materials and procedure were identical to those in Exp. 1, except that participants were tested in the lab rather than online.

3.2. Results

The pre-registration specified the use of Frequentists analyses, rather than Bayesian analyses, and comparisons of consistent versus inconsistent item accuracies, rather than analyses of the accuracy interference scores (a subtraction of accuracy on inconsistent items from accuracy on consistent items; note that computing paired-sample t tests on consistent versus inconsistent items is mathematically identical to computing one-sample

t test on their difference score). We therefore report Frequentist comparisons of consistent versus inconsistent item accuracies, with Bayesian analyses of the interference scores reported alongside them, to facilitate easy comparisons with the other experiments reported in this study. See Appendix S1 for Frequentists analyses on consistent versus inconsistent item accuracies for Exp. 1 that were used as the basis for the pre-registration.

A successful replication was defined as meeting the following three criteria: (a) statistically significant *p*-values in the same direction as in Exp. 1, (b) effect sizes that differ from zero, and are not different from the effect sizes of Exp. 1, and (c) Evidence Updating Bayes Factors that are greater than 3 in favor of the alternative hypotheses relative to the null hypotheses.

3.2.1. The p*-values are statistically significant and are in the same direction as in Exp. 1*

Frequentist paired-sample *t* tests (one-tailed) revealed that for psychology items, accuracy was significantly higher for consistent (M = 96.5%, SD = 5.0%) than inconsistent (M = 87.9%, SD = 10.3%) items; $t(18) = 3.55$, $p = .001$, $d = 0.81$, 95% CI = [0.37, ∞].[3] Similarly, for physicality items, accuracy was significantly higher for consistent (M = 94.9%, SD = 8.8%) than inconsistent (M = 90.0%, SD = 11.8%) items; $t(18) = 2.46$, $p = 0.012$, $d = 0.57$, 95% CI = [0.15, ∞]. The *p*-values from Exp. 2 were significant, and in the same direction, as those in Exp. 1, satisfying the first criterion for a successful replication.

Additionally, to facilitate comparisons with the other experiments reported in this study, we analyzed the data with Bayesian one-sample *t* tests. We evaluated whether the accuracy interference scores on psychology and physicality were different from zero. The experimental hypothesis that they were different from zero was more likely than the null for both psychology (accuracy BF_{10} = 36.00) and physicality (accuracy BF_{10} = 4.99).

3.2.2. The effect sizes of the replication experiment differ from zero and are not different from the effect sizes of Exp. 1 (based on 95% CIs)

The effect sizes of the replication experiment differed from zero for both the psychology items, $d = 0.81$ and 95% CI = [0.37, ∞], and the physicality items, $d = 0.57$ and 95% CI = [0.15, ∞]. Additionally, the effect sizes of the replication experiment fell within the confidence intervals of Exp. 1 for both psychology, 95% CI = [0.55, 0.94], and physicality, 95% CI = [0.25, 0.61], suggesting that the effects were of similar magnitude. The finding that the effect size estimates from Exp. 2 differed from zero and were similar in magnitude to those in Exp. 1 satisfies the second criterion for a successful replication.

3.2.3. The Evidence Updating Bayes Factors are greater than 3 in favor of the alternative hypotheses relative to the null hypotheses

Finally, given limitations of directly comparing *p*-values and/or effect sizes, and difficulties of interpreting replications due to differences in power between the original and replication experiments (Simonsohn, 2015; Verhagen & Wagenmakers, 2014), we

complemented the previous two criteria to evaluating replication success by an Evidence Updating Bayes Factor (EU-BF). The EU-BF addresses whether an effect is present in a replication experiment factoring in the data from the original experiment (Ly et al., 2018). The EU-BF is calculated as follows: BF_{10} (replication | original) = BF_{10} (original + replication)/BF_{10} (original). As with the Frequentist analyses, we tested directional hypotheses. The test favored the alternative hypothesis that accuracy interference scores for both psychology (EU-BF_{10} = 87.74 and physicality (EU-BF_{10} = 9.85) were different than zero in a positive direction, relative to the null hypothesis of being different from zero in a negative direction or not different from zero. The EU-BF thus provides complementary evidence and satisfies the third criterion for a successful replication.

4. Experiment 3

4.1. Methods

4.1.1. Power analysis

Exp. 1 extended the RT effect documented in Barlev et al. (2017, 2018) to MTurk workers. The effect size found was a bit greater than 0.40 (vs. the initially estimated $d = 0.30$). We therefore used $d = 0.40$ to calculate power for Exp. 3. This showed that for a power of 95% and an alpha of 5%, at minimum 70 participants for each of the three conditions in this experiment were needed.

4.1.2. Participants

Participants were Christian religious adherents recruited from MTurk. Pre-screening and exclusion criteria were identical to those in Exp. 1. The final sample of $N = 219$ (63% F) had a mean age of 39 (range 18–74). Participants identified as White (86%), Hispanic or Latino (6%), East, Southeast, or South Asian (4%), and Black (4%). Participants identified as Roman Catholic (37%), Protestant Evangelical (37%), and Mainline Protestant (26%). Ninety-three percent currently identified with the religious tradition with which they grew up. Finally, participants identified as $M = 70$ ($SD = 22$) on religiosity and $M = 76$ ($SD = 20$) on spirituality ($r = .68$, $p < .001$), and on the Christian Orthodoxy Scale participants scored $M = 2.51$ ($SD = 0.80$).

4.1.3. Design

The IV was the characteristic of God, the Holy Spirit, or Jesus (physicality vs. psychology) tested within subjects. Additionally, the target of each statement (God vs. the Holy Spirit vs. Jesus) was manipulated between subjects. The DVs were, as in Exp. 1, the accuracy and RT interference scores (calculated as the difference in performance between consistent and inconsistent statements).

4.1.4. Materials and procedure

The materials and procedure were identical to those in Exp. 1, except that similarly constructed statements ($n = 64$) concerning mathematics and science were added as distractors (from Shtulman & Valcarcel, 2012). The distractor science and mathematics items were mixed with the religion items. Further, in the Holy Spirit and Jesus conditions, "God" was replaced by these extraordinary beings.

4.2. Results

4.2.1. The Holy Spirit and Jesus, like God, co-opt an embodied person concept

We used a 3 (Condition: God vs. Holy Spirit vs. Jesus) × 2 (Characteristic: Psychology vs. Physicality) Bayesian mixed ANOVA to evaluate whether there were differences in response interference among the three members of the Trinity. The experimental hypothesis that there were such differences was less likely than the null (Condition: accuracy $BF_{10} = 0.05$, RT $BF_{10} = 0.07$). Similar to Exp. 1, we found weak evidence that interference scores for physicality were different from those for psychology (Characteristic: accuracy $BF_{10} = 1.16$, RT $BF_{10} = 3.00$), with the RT interference score for physicality being nearly twice as large as for psychology. Finally, the experimental hypothesis that Condition and Characteristic entered into an interaction was less likely than the null (accuracy $BF_{10} = 0.04$, RT $BF_{10} = 0.04$).

We then used Bayesian one-sample t tests to evaluate whether the accuracy and RT interference scores on psychology and physicality were different from zero for each of the three members of the Trinity. The experimental hypothesis that they were was more likely than the null for all three beings for physicality (God: accuracy $BF_{10} = 184.70$, RT $BF_{10} > 300$; Holy Spirit: accuracy $BF_{10} = 3.84$, RT $BF_{10} = 17.90$; Jesus: accuracy $BF_{10} > 300$, RT $BF_{10} = 20.93$) and psychology (God: accuracy $BF_{10} > 300$, RT $BF_{10} = 6.69$; Holy Spirit: accuracy $BF_{10} > 300$, RT $BF_{10} = 59.47$; Jesus: accuracy $BF_{10} > 300$, RT $BF_{10} = 4.60$). The above replicates the findings of Exp. 1 with God and extends them to the cases of the Holy Spirit and Jesus.

5. Experiment 4

5.1. Methods

5.1.1. Power analysis

We used $d = 0.40$ to calculate power for this experiment (see Section 4.1.1. for details). This showed that for a power of 95% and an alpha of 5%, at minimum 70 participants were needed.

5.1.2. Participants

Participants were Christian religious adherents recruited from MTurk. Pre-screening and exclusion criteria were identical to those in Exp. 1, except that in this experiment

participants were additionally pre-screened to currently identify with a church, and to know the priest of that church at least slightly. The final sample of $N = 69$ (70% F) had a mean age of 39 (range 18–75). Participants identified as White (94%), Hispanic or Latino (1%), East, Southeast, or South Asian (1%), and Black (3%). Participants identified as Roman Catholic (33%), Protestant Evangelical (42%), and Mainline Protestant (25%). Ninety-four percent currently identified with the religious tradition with which they grew up. Finally, participants identified as $M = 72$ ($SD = 19$) on religiosity and $M = 78$ ($SD = 21$) on spirituality ($r = .72$, $p < .001$), and on the Christian Orthodoxy Scale participants scored $M = 2.57$ ($SD = 0.76$).

5.1.3. Design

The IV was the characteristic of God or "my priest" (physicality vs. psychology) tested within subjects. Additionally, the target of each statement ("my priest" vs. God) was manipulated within subjects. The DVs were, as in Exp. 1 and 3, the accuracy and RT interference scores (calculated as the difference in performance between consistent and inconsistent statements).

5.1.4. Materials and procedure

The materials and procedure were identical to those in Exp. 3, except that "God" was replaced with "my priest" in all physicality statements, and in a subset of psychology statements that could be modified thus and still make sense. For example, since God listens to and answers prayers but not priests, statements concerning God's ability to listen to prayers were not modified. The stimuli set for this experiment was thus a mix of "God" ($n = 16$) and "my priest" ($n = 36$) statements. Correct responses to "my priest" statements were coded based on the intuition level (e.g., "true" responses to the statement "My priest is at my church when he is not at other churches" were coded as correct).

5.2. Results

5.2.1. The accuracy and RT effects observed in this study are likely due to representational coexistence and interference, not low-level biases in the statements

We used Bayesian one-sample *t* tests to evaluate whether the accuracy and RT interference scores on the modified "my priest" psychology and physicality items were different from zero. The null hypothesis that interference scores were not different from zero was more likely than the experimental hypothesis for physicality RT ($BF_{10} = 0.14$) and for psychology accuracy ($BF_{10} = 0.13$) and RT ($BF_{10} = 0.27$). The experimental hypothesis that interference scores were different from zero was more likely than the null for physicality accuracy ($BF_{10} = 2.84$), but the effect was in the opposite direction to that in Exp. 1 and 3.

We next performed the following additional checks. First, we confirmed that interference scores on the unmodified "God" psychology items were different from zero (accuracy $BF_{10} = 71.06$, RT $BF_{10} = 7.34$). We thereby showed that the null effects for the

"my priest" items were not due to a general failure to replicate the interference effects in this experiment. Second, we confirmed that interference scores on the same subset of items that were modified in this experiment ("my priest") were for the most part different from zero in Exp. 1 and 3 where they are unmodified (i.e., where they read "God"): Exp. 1, physicality: accuracy $BF_{10} > 300$, RT $BF_{10} > 300$; psychology: accuracy $BF_{10} > 300$, RT $BF_{10} = 0.12$, and Exp. 3, physicality: accuracy $BF_{10} = 184.70$, RT $BF_{10} > 300$; psychology: accuracy $BF_{10} = 84.90$, RT $BF_{10} = 0.14$. We thereby showed that the effects for the "my priest" items observed in this experiment were not due to a failure to replicate the interference effects in that subset of items.

The above analyses suggest that it is very unlikely that performance differences between consistent and inconsistent statements in Exp. 1 and 3 were caused by low-level biases in the structure of statements.

6. Discussion

The present series of experiments aimed to evaluate two contrasting hypotheses for why representations of disembodied extraordinary beings are prevalent in present and past theologies: intuitive Cartesian dualism (e.g., Bloom, 2005) and counterintuitiveness (M. Barlev & A. Shtulman, under review; Boyer, 2001). At the crux of these alternatives is whether it is natural to represent persons—or extraordinary beings such as the Christian God which co-opt the person template—as disembodied. We used a sentence verification paradigm where participants evaluated as true or false statements that were consistent or inconsistent between core knowledge intuition about person physicality or psychology and acquired Christian theological doctrines about God (e.g., that God is present everywhere at the same time or that God has no physical or bodily existence). If it is unnatural or unintuitive to conceptualize of persons (and by extension, of extraordinary beings like God) as disembodied, then participants would show representational interference on statements targeting God's physicality as well as psychology. But if it is intuitive to conceptualize persons as disembodied (as argued by inherent Cartesian dualism), then participants would show representational interference on statements targeting God's psychology only. Across multiple experiments, including a pre-registered replication (Exp. 2), we found support for the suggestion by M. Barlev and A. Shtulman (under review) that it is unintuitive to conceptualize God as disembodied: Christian religious adherents showed evidence of representational coexistence and interference (lower accuracy and slower RT) on statements where core knowledge intuitions about person physicality and acquired Christian theological doctrines about God's physicality (e.g., omnipresence and incorporeality) were inconsistent versus consistent. The same was also found for statements where core knowledge intuitions about person psychology and acquired Christian theological doctrines about God's psychology (e.g., infallibility) were inconsistent versus consistent (thereby replicating Barlev et al., 2017, 2018).

The implication of these findings to the academic study of religious thought and behavior are wide-ranging. First, as already noted, these findings suggest that the

prevalence of theological representations of disembodied beings can be understood within the explanatory framework advanced by Boyer (2001). We thus would predict, for example, that disembodied being concepts, like other counterintuitive concepts, would preferentially recruit attention and memory (e.g., Banerjee, Haque, & Spelke, 2013; Barrett & Nyhof, 2001; Boyer & Ramble, 2001) as would the sources of such concepts (S. Mermelstein, M. Barlev, & T.C. German, under review). We would also predict that the prevalence of such concepts could additionally be partially understood by a differential motivation to talk about them as people try to reconcile them with their pre-existing beliefs (S. Mermelstein, M. Barlev, & T.C. German, under review).

Second, these findings shed light on discrepancies between theological representations of extraordinary beings as disembodied and the speech and actions of ordinary people (M. Barlev & A. Shtulman, under review). For example, as Hodge (2008; also see Nikkel, 2015) argued, Christian theology notwithstanding, Christian religious adherents also conceptualize the deceased as embodied (e.g., as is evident in artistic depictions). See Slone (2004) for other such discrepancies between formal theology and "on-the-ground" beliefs and behaviors. Indeed, historically, in Abrahamic traditions as well as in other ancient religions, the deceased were theologically—as well as in the speech and actions of laypeople—conceptualized as embodied and the afterlife as a physical world (e.g., the underworld in Greek theology). Finally, we find a variety of practices that are hard to explain if disembodied beings like spirits were not also conceptualized as embodied: Japanese practitioners of the traditional religion Shintoism leave offerings like food and sake for animistic spirits and the spirits of the deceased, and practitioners of Candomblé leave cigarettes and alcohol for possessing spirits (here, these beings are not only conceived as embodied, but as requiring food and water and as pleased by carnal desires).

Note that although we did not have specific a priori predictions about possible differences between the physicality and psychology representational interference effects, we found evidence suggesting that the RT effects for physicality were larger than those for psychology in Exp. 1 ($BF_{10} = 5.91$). However, the accuracy data in Exp. 1 showed no such difference ($BF_{10} = 0.15$; the effects in Exp. 3 were in the "anecdotal evidence" range of between $\frac{1}{3}$ and 3 and are therefore not readily interpretable). We are therefore not ready to comment on whether differences between the physicality and psychology effects exist, except to note that whether such differences do or do not exist does not have obvious implications to our thesis. Future research is needed to further explore this possibility.

Further, following the suggestion by Sharp et al. (2017) that among the members of the Trinity, the Holy spirit is conceptualized as least like a person, we explored whether the Holy Spirit, like Jesus and God, co-opts the person concept. The same behavioral signatures of representational coexistence and interference from both folk psychology and physicality were found with the Holy Spirit as with Jesus and God (Exp. 3), suggesting that the person concept is co-opted to form representations of all three extraordinary beings. First, this implies that although Christian theology and, correspondingly, conceptualizations by Christian religious adherents, focus on certain differences between the three members of the Trinity (Sharp et al., 2017), all three are grounded in the same core

person concept. Second, this implies that, as suggested by Boyer (2001), the person concept may play a central role in forming representations of extraordinary beings broadly; nonetheless, we still consider it plausible that there are extraordinary beings that co-opt other concepts in addition to the person concept. We discuss this further under Future Directions.

Last, a control experiment (Exp. 4) ruled-out the possibility that the findings reported here were caused by low-level biases in the structure of the statements used. If the pattern of effects reported here was due to such low-level biases, then statements where worse performance was observed for the extraordinary being "God" should show the same effects for the ordinary being "my priest." We compared statements that in Exp. 1–3 were coded as "consistent" and "inconsistent" (these terms no longer apply to the "my priest" items because they do not have a theological layer; also note that the "my priest" items are true or false according the intuitive layer), and we did not find worse performance on "inconsistent" versus "consistent" statements.

6.1. Constraints on generality

We identify, following the recommendation of Simons, Shoda, and Lindsay (2017), target populations for the reported findings (also see Future Directions), and information that is important for future replications. The findings reported in the present study with several samples of Christian religious adherents recruited from MTurk (predominantly white) and from a predominantly East, Southeast, or South Asian undergraduate population at a university in Hawai'i, expand on previous findings with several samples of young-adult undergraduates at a Southern California university (Barlev et al., 2017), and of an age-variant adult sample recruited from Christian churches in Southern California (Barlev et al., 2018). Our general findings on representational coexistence and interference have been highly replicable regardless of sample, and of various methodological differences between studies including whether participants were tested in the laboratory (Barlev et al., 2017, 2018; Exp. 2 of the present study) or online (Exp. 1, 3, and 4 of the present study), and of whether filler science statements were included with the religion statements (e.g., Exp. 1–2 vs. Exp. 3–4 of this study).

A couple of notes for future replications: (a) We found that following the participant prescreening and exclusion criteria described here and in Barlev et al. (2017, 2018) is important. If a participant does not believe in mainstream Christian theology (e.g., Latter Day Saints believe that God has a physical body) or has an idiosyncratic conceptualization of God, she may, for example, respond "false" to some statements that are coded as "true" according to Christian theology not because of representational coexistence and interference from a core knowledge person concept, but because that is her theological conceptualization of God; similarly, if a participant does not believe in the existence of God, she may respond "false" to some or all statements concerning God's capacities or press the "true" and "false" keys randomly. Indeed, in our previous studies, when exclusion criteria such as identifying as at least slightly religious were applied after data have been collected (rather than used to prescreen participants as in Exp. 1, 3, and 4), we

observed that participants who reported being not at all religious were also very likely to have religion Sentence Verification Task scores that were around zero or not statistically different from chance. (b) We found that when constructing new statements, it is difficult in all instances to predict how these statements will be interpreted by participants. For example, consider the statements "God can sometimes see what I'm doing" and "God can sometimes hear what I'm saying," both of which were coded in Barlev (2018) as true intuitively and true theologically. Although the predicted answer to these statements was "true," many participants responded that these statements were false. Barlev (2018) found that the modifier "sometimes" was interpreted as "only sometimes" by many participants. We recommend that researchers who construct new statements pre-test them, and also that researchers get qualitative data through interviews or focus groups from the population being studied on how they interpret these new statements. We additionally recommend the use of controls, like the "my priest" condition in Exp. 4 of this study (and Exp. 3 in Barlev et al., 2017), to make sure that this way of constructing statements does not bias the structure of the statements toward the experimental hypothesis.

6.2. Future directions

Consider the ethnographic study by Cohen (2007, 2008) of Candomblé, an Afro-Brazilian syncretic religion that is focused on ritualized interactions with possessing spirits. Cohen suggested that in Candomblé there are two distinct forms of spirit possession: executive possession, in which a spirit is believed to take over the body of the possessed, speaking and acting through him or her, and pathogenic possession, in which a spirit is believed to be the cause of illness and misfortune. Cohen suggested that in the minds of practitioners of Candomblé, these two forms may be conceptualized differently: while executive possession spirits may co-opt the person concept, pathogenic possession spirits may co-opt the pathogen concept (see, e.g., Rozin, Haidt, & McCauley, 2008). In this study, we found that in the minds of Christian religious adherents, the Holy Spirit, like Jesus and God, co-opts the person concept. We take this to show the primacy of the person concept in forming extraordinary being concepts, and we are therefore somewhat skeptical of the possibility that pathogenic possession spirits do not co-opt the person concept. A different possibility may be that pathogenic possession spirits co-opt both the person concept and the pathogen concept. Future research into possessing spirits in Candomblé is needed to explore this; future research into how the mind forms evolutionarily new concepts is needed to explore the possibility that two or more core concepts can be co-opted to form a new concept (thus far, researchers, including us, have focused on the co-option of a single concept; see also for example Cohen, Sasaki, & German, 2015, on how the Theory of Mind mechanism is co-opted to process public representations like maps).

Regarding our finding that God, the Holy Spirit, and Jesus all co-opt the person concept in the minds of Christian religious believers, future research is needed to explore whether this generalizes to other religions. For example, although 83% of Jews believe in the existence of God or a universal spirit (vs. 98% of Protestants and 97% of Catholics), only 25% of Jews believe that God is a person with whom people can have a

relationship (vs. 72% of Protestants and 60% of Catholics) with 50% believing that God is an impersonal force (vs. 19% of Protestants and 29% of Catholics) (Pew Research Center, 2008). Although we are somewhat skeptical of this possibility, the person concept may not be co-opted, or be co-opted by a substantially smaller number of religious adherents, to form the Jewish versus the Christian God concept.

Finally, as in Barlev et al. (2017, 2018), not all participants showed behavioral signatures of representational coexistence and interference—this was especially so for response accuracy (which as was already noted is a less sensitive measure of representational coexistence and interference than RT), visible in the pirate plots as a wider distribution around zero effect. Future research is needed to explore individual difference variables that might explain such differences in performance on the sentence verification task and what this might reveal about the underlying representations and/or the process that resolves conflict between inconsistent representations.

7. Conclusions

Why are theological representations of disembodied beings culturally prevalent? Intuitive Cartesian dualism (e.g., Bloom, 2005) proposes that this is because it is natural to conceptualize of persons—and extraordinary beings that co-opt the person concept—as disembodied minds. However, our findings that core knowledge intuitions about person physicality coexist and interfere with Christian theological representations of God as incorporeal and omnipresent are incompatible with this proposal. Our findings instead suggest that the Christian God—and disembodied beings that co-opt the person concept more broadly like the Holy Spirit—are unnatural. Indeed, they are counterintuitive (Boyer, 2001) because they violate core knowledge intuitions about person physicality; perhaps, as M. Barlev and A. Shtulman (under review) have suggested, it is exactly this that explains their social transmission advantage and their prevalence across the world's religions.

Acknowledgments

We thank James A. Hampton and three anonymous reviewers for helpful comments.

Author contributions

MB and TCG developed the study concept. MB, SM, and TCG developed the study materials. MB and SM collected and analyzed the data for Exp. 1, 3, and 4. ASC collected and analyzed the data for Exp. 2. MB drafted the manuscript, and all authors suggested critical edits and approved the final version of the manuscript for submission.

Declaration of conflicting interests

The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Notes

1. We do not argue that all counterintuitive concepts have a transmission advantage. For example, an additional criterion discussed in Boyer (2001) is conserved inferential potential or that counterintuitive concepts trigger useful inferences that are not blocked by the core knowledge violation(s). A counterintuitive concept may be a cognitive dead-end (Boyer, 2001) if it does not allow us to make any such useful inferences (e.g., a deity that can see into the future but who always forgets what he sees). The critical finding for our purposes, replicated in many studies with children and adults and across different populations, is that all else being equal counterintuitive concepts have a marked transmission advantage over otherwise similar but ordinary concepts.
2. Strictly, this usage does not recognize the wide variety of clergy terms used in Christianity: the term "priest" refers to some ordained ministers of some Christian denominations such as the Roman Catholic Church. But an adherent belonging to these denominations might have a clergyperson who is not a priest (e.g., a deacon or a bishop), and further, adherents belonging to other denominations might have other clergypersons altogether (e.g., ordained ministers in Lutheranism are often referred to as pastors). We chose, for simplicity, to only use the term "priest" in Exp. 4. We found no evidence that this affected the results of Exp. 4: participants, regardless of their denominational affiliation, responded to "my priest" items sensibly, and there were no noteworthy differences on performance on the sentence verification task between the three denominations represented in Exp. 4 (Roman Catholics, Protestant Evangelicals, and Mainline Protestants).
3. The upper bound on some of the CIs reported here is infinity because of the use of one-tailed t-tests. A two-sided CI uses a +/− margin of error (MoE). A one-sided CI uses either a +MoE or a −MoE, meaning only an upper or a lower bound will be defined. That is, a MoE around the point estimate is only defined on one side of the estimate because the hypothesis test is directional. For example, given the alternative $mu1 - mu2 > 0$, all positive values are compatible with it, so the upper bound is positive infinity no matter what the point estimate is.

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. https://doi.org/10.1146/annurev-psych-010213-115033

Banerjee, K., Haque, O. S., & Spelke, E. S. (2013). Melting lizards and crying mailboxes: Children's preferential recall of minimally counterintuitive concepts. *Cognitive Science*, *37*(7), 1251–1289. https://doi.org/10.1111/cogs.12037

Barlev, M. (2018). How the mind builds evolutionarily new concepts. Unpublished doctoral dissertation, University of California, Santa Barbara.

Barlev, M., Kinsella, M., German, T. C., Taves, A., & Paloutzian, R. F. (2015). Anomalous experiences and paranormal attributions in a new spiritual movement. Paper presented at the Psychology of Religion and Spirituality Preconference at the Society for Personality and Social Psychology, Long Beach, CA.

Barlev, M., Mermelstein, S., & German, T. C. (2017). Core intuitions about persons coexist and interfere with acquired Christian beliefs about God. *Cognitive Science*, *41*(S3), 425–454. https://doi.org/10.1111/cogs.12435

Barlev, M., Mermelstein, S., & German, T. C. (2018). Representational co-existence in the God concept: Core knowledge intuitions of God as a person are not revised by Christian theology despite lifelong experience. *Psychonomic Bulletin & Review*, *25*, 2330–2338.

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

Barrett, J., & Nyhof, M. (2001). Spreading non-natural concepts: The role of intuitive conceptual structures in memory and transmission of cultural materials. *Journal of Cognition and Culture*, *1*(1), 69–100. https://doi.org/10.1163/156853701300063589

Barlev, M., & Shtulman, A. (under review). Minds, bodies, spirits, and gods: Does widespread belief in disembodied beings imply that we are inherent dualists? https://psyarxiv.com/e9cw4/

Bloom, P. (2005). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.

Bloom, P. (2007). Religion is natural. *Developmental Science*, *10*(1), 147–151. https://doi.org/10.1111/j.1467-7687.2007.00577.x

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.

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

Cohen, A. S., Sasaki, J., & German, T. C. (2015). Are beliefs special because they are beliefs or because they are representations? *Cognition*, *136*, 49–63.

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. https://doi.org/10.1080/00141840801927558

Csibra, G., Gergely, G., Bíró, S., Koós, O., & Brockbank, M. (1999). Goal attribution without agency cues: The perception of "pure reason" in infancy. *Cognition*, *72*, 237–267.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Forstmann, M., & Burgmer, P. (2015). Adults are intuitive mind-body dualists. *Journal of Experimental Psychology: General*, *144*(1), 222–235.

Forstmann, M., & Burgmer, P. (2017). Antecedents, manifestations, and consequences of belief in mind-body dualism. In C. M. Zedelius, B. C. N. Müller, & J. Schooler (Eds.), The science of lay theories: How beliefs shape our cognition, behavior, and health (pp. 181–205). New York: Springer. http://doi.org/10.1007/978-3-319-57306-9

Fullerton, T. J., & Hunsberger, B. (1982). A unidimensional measure of Christian orthodoxy. *Journal for the Scientific Study of Religion*, *21*(4), 317–326.

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

Held, L., & Ott, M. (2018). On p-values and Bayes factors. *Annual Review of Statistics and Its Application*, *5*, 393–419.

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. https://doi.org/10.1163/156853708X358236

Hood, B., Gjersoe, N. L., & Bloom, P. (2012). Do children think that duplicating the body also duplicates the mind? *Cognition*, *125*(3), 466–474. https://doi.org/10.1016/j.cognition.2012.07.005

Hunsberger, B. (1989). A short version of the Christian orthodoxy scale. *Journal for the Scientific Study of Religion*, *28*(3), 360–365.

JASP Team. (2017). JASP (Version 0.8.1.2) [Computer software].

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kinsella, M. (2017). Near-death experiences and networked spirituality: The emergence of an afterlife movement. *Journal of the American Academy of Religion*, *85*(1), 168–198. https://doi.org/10.1093/jaarel/lfw037

Kuhlmeier, V., Bloom, P., & Wynn, K. (2004). Do 5-month-old infants see humans as material objects? *Cognition*, *94*, 95–103.

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

Leslie, A. (1994). ToMM, ToBY, and Agency: Core architecture and domain specificity. 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.

Ly, A., Etz, A., Marsman, M., & Wagenmakers, E.-J. (2018). Replication Bayes factors from evidence updating. *Behavioral Research Methods*, https://doi.org/10.3758/s13428-018-1092-x

Mermelstein, S., Barlev, M., & German, T. C. (under review). he told me about a singing cactus: The sources of counterintuitive concepts are recalled better than those of ordinary concepts. https://psyarxiv.com/6cp8e

Nikkel, D. H. (2015). The dualistic, discarnate picture that haunts the cognitive science of religion. *Zygon*, *50*(3), 621–646. https://doi.org/10.1111/zygo.12194

Pew Research Center. (2008). U.S. Religious landscape survey: Religious beliefs and practices: Diverse and politically relevant. Pew Forum on Religion & Public Life. Available at religions.pewforum.org

Rozin, P., Haidt, J., & McCauley, R. C. (2008). Disgust. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 757–776). New York: Guilford Press.

Sharp, C. A., Rentfrow, P. J., & Gibson, N. J. S. (2017). One God but three concepts: Complexity in Christians' representations of God. *Psychology of Religion and Spirituality*, *9*(1), 95–105.

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

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on Generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123–1128. https://doi.org/10.1177/1745691617708630

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, *26*(5), 559–569.

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. https://doi.org/10.1016/0364-0213(90)90025-R

Spelke, E., & Kinzler, K. (2007). Core knowledge. *Developmental Science*, *10*, 89–96.

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

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.

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Verhagen, J., & Wagenmakers, E. J. (2014). Bayesian tests to quantify the result of a replication attempt. *Journal of Experimental Psychology: General*, *143*(4), 1457.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*, 1–34.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1: Supplementary Materials for The Embodied God.