Cognition 204 (2020) 104398

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Science demands explanation, religion tolerates mystery

Emily G. Liquin[a,*], S. Emlen Metz[b], Tania Lombrozo[a]

[a] Department of Psychology, Princeton University, Peretsman Scully Hall, Princeton, NJ 08540, USA
[b] University of California, Berkeley, Campbell Hall, University Drive, Berkeley, CA 94720, USA

ARTICLE INFO

Keywords:
Explanation
Explanatory inquiry
Science
Religion
Mystery
Belief

ABSTRACT

Some claims (e.g., that the Earth goes around the Sun) seem to call out for explanation: they make us wonder "why?". For other claims (e.g., that God exists), one might accept that the explanation is a mystery. In the present research, we investigate "need for explanation" and "mystery acceptability" across the domains of science and religion, as a window onto differences between scientific and religious cognition more broadly. In Study 1, we find that scientific "why" questions are judged to be in greater need of explanation and less adequately answered by appeals to mystery than religious "why" questions. Moreover, this holds for both religious believers and non-believers. In Study 2, we find that these domain differences persist after statistically controlling for confidence in the premises of scientific and religious "why" questions (e.g., that "the Earth goes around the Sun" and that "there is a God"). In Study 3, we match levels of confidence within-participants, and we find that domain differences in need for explanation and mystery acceptability are systematically related to domain differences in epistemic commitments (whether an explanation is within human comprehension, whether the same explanation is true for everyone) and explanatory norms (whether an explanation should be pursued), which could signal domain differences in epistemic and social functions, respectively. Together, these studies shed light on the role of explanatory inquiry across domains, and point to different functional roles for scientific and religious cognition.

1. Introduction

> "The important thing is not to stop questioning. Curiosity has its own reason for existence. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery each day."
>
> **Albert Einstein** (quoted in (Miller, 1955), p.64)

> "Try and penetrate with our limited means the secrets of nature and you will find that, behind all the discernible concatenations, there remains something subtle, intangible and inexplicable. Veneration for this force beyond anything that we can comprehend is my religion. To that extent I am, in point of fact, religious."
>
> **Albert Einstein**, (quoted in (Kessler, 1971))

Recognizing that something is a mystery could highlight a gap in one's knowledge, prompting curiosity, questioning, and a search for information (e.g., Kidd & Hayden, 2015; Loewenstein, 1994). At the same time, declaring something a mystery suggests that the target of inquiry might be ineffable and inexplicable. The first perspective is reflected in Einstein's attitude to mystery as spurring on scientific advance (questioning), the second in his acknowledgement of limits to human understanding (resulting in a veneration that Einstein calls "religious").

In the current paper we investigate the role of questioning and mystery across the domains of science and religion. In particular, we ask: do questions about scientific matters (e.g., "Why is the center of the Earth so hot?") prompt a more strongly felt need for explanation than questions about religious matters (e.g., "Why did God create the world?")? Is declaring something a mystery more acceptable in the domain of religion (e.g., "Why does prayer work? It's a mystery") than in the domain of science (e.g., "Why does the Moon cause tides? It's a mystery")? We hypothesize that the answer to both questions is yes, and we argue that these domain differences in "need for explanation" and "mystery acceptability" can shed light on the different functional roles of scientific and religious beliefs in human cognition. In particular, whereas inquiry in science typically plays an epistemic role (helping us achieve a more veridical representation of the world that enables us to make predictions and take effective actions), inquiry (or its absence) in religion may often play non-epistemic roles (such as signaling individual and group identities and supporting narratives of meaning).

In the remainder of the Introduction, we first review prior work on

* Corresponding author.
E-mail addresses: eliquin@princeton.edu (E.G. Liquin), emlen.metz@berkeley.edu (S.E. Metz), lombrozo@princeton.edu (T. Lombrozo).

https://doi.org/10.1016/j.cognition.2020.104398
Received 30 January 2020; Received in revised form 3 July 2020; Accepted 7 July 2020

scientific and religious explanation, including (1) the similarities and differences in explanation across domains, and (2) when an unresolved mystery or "no explanation" is judged adequate. We then consider research on scientific and religious belief more generally, including important differences in their responsiveness to evidence, and in the (average) confidence with which they are held. This work motivates our proposal that the domains of science and religion are differently aligned with epistemic versus non-epistemic aims, and that this should be reflected in need for explanation and mystery acceptability, a hypothesis we go on to test in three studies.

1.1. Scientific and religious explanations

While prior work has not investigated need for explanation or mystery acceptability across domains, scholars have long emphasized the central role of explanation in both the domain of science and the domain of religion, where we can understand these domains as defined by their content (with "science" concerning natural entities like germs and mountains and "religion" concerning supernatural entities like gods and angels), their corresponding methods of inquiry, and the institutions that support them. Importantly, science and religion both offer explanations for the world and for human experience, and on many accounts, providing such explanations is a central aim of both—offering a sense of understanding for otherwise mysterious events, such as an unexpected power outage or a tragic loss. Harré (1985 p. 168) called scientific explanations "the crown of science"; similarly, some regard "the pursuit of explanation, prediction, and control [as] both necessary and sufficient for the flourishing of religious life" (Horton, 1997 p. 373).

These similarities in function are matched by similarities in cognitive processing. Prior research has suggested that shared cognitive processes underlie both religious and non-religious explanations for everyday events (e.g., Barrett, 2000, 2004; Hood et al., 2009; Lupfer et al., 1992). In particular, both scientific and religious beliefs are shaped by intuitive theories and by common cognitive biases, including tendencies to essentialize and to interpret behavior as purposeful (e.g., Shtulman, 2015; Willard & Norenzayan, 2013). Religious and scientific explanations can be integrated in individual minds (for a review, see Legare et al., 2012), and religious and scientific beliefs can co-exist just as distinct sets of factual beliefs co-exist (Harris & Giménez, 2005; Legare & Gelman, 2008; Shtulman & Lombrozo, 2016).

On the other hand, scientific and religious explanations are often distinct in their targets and forms. Adults are more likely to generate religious explanations for events that are life-altering and positive in valence as opposed to mundane or negative (Lupfer et al., 1996, see also Lupfer et al., 1992; Ransom & Alicke, 2012; Vonk & Pitzen, 2016; Weeks & Lupfer, 2000). Both children and adults are more likely to generate supernatural explanations for events that are positively valenced as opposed to negatively valenced (Woolley et al., 2011), and for things that are unlikely or impossible than for those that are possible (Woolley & Cornelius, 2017; see also Woolley & Dunham, 2017). There is also evidence that scientific explanations are thought to be more appropriate to explain aspects of human experience and behavior that are associated with the body and shared with other animals, such as headaches or discerning temperature through touch, than those that have a rich mental experience that is potentially unique to humans, such as falling in love or feeling transformed by a spiritual event (Gottlieb & Lombrozo, 2018). Beyond differing in their targets, religious and scientific explanations may differ in form. In particular, religious explanations often appeal to intentional agents and are teleological in nature (appealing to the function or purpose of the thing being explained; Kelemen, 2004), while scientific explanations instead emphasize physical causal mechanisms (Keil, 2019).

In sum, there is evidence that explanation is central to both science and religion, and that scientific and religious explanations make use of shared cognitive processes. On the other hand, scientific and religious explanations are distinct in their typical targets and forms. However, this prior work has focused largely on *what domain of explanation is demanded* given different circumstances (i.e., whether a scientific or religious explanation should be offered for a given phenomenon) and on how *explanations appealing to different domains* are similar or dissimilar (i.e., how an explanation appealing to scientific content is different from an explanation appealing to religious content). This work leaves open the question of *what demands an explanation across domains*. According to the philosophy literature on "need for explanation" (Grimm, 2008; Wong & Yudell, 2015), the sense that an event or phenomenon demands an explanation motivates explanatory inquiry, whether it takes the form of explanation generation, experimentation, or question-asking. We can therefore ask: when are explanations demanded for religious and scientific phenomena? And are explanations demanded with the same frequency and fervor across domains? Conversely, when is it appropriate to refuse explanatory inquiry, for example by stating that some event or phenomenon is (and perhaps even ought to be left as) a "mystery"?

1.2. Mystery in science and religion

Within science, declaring something a mystery is a sign that current scientific theory is inadequate or incomplete (for relevant discussion see Klein & Colombo, 2018). In contrast, some religious traditions seem to actively embrace (some) mysteries (Boudry & Coyne, 2016; Boudry & De Smedt, 2011; Boyer, 2001; Sperber, 1996). For instance, some Christian theologians maintain that the trinity is a mystery (Tuggy, 2016) and some mystical traditions consider mystical experiences to be "ineffable," meaning the experience or its object is beyond the possibility of explanation (Gellman, 2017). While empirical research on this topic has been sparse, there is evidence that the *absence* of a good scientific explanation can increase implicit endorsement of religious ideas, and perhaps lead to greater acceptance of the claim that "there is no way to explain" the premise of a question. We review these findings as evidence that science and religion may be differentially tolerant of weak or absent explanations.

In a study investigating the relationship between scientific and religious explanations, Preston and Epley (2009) manipulated whether science was presented as offering a strong explanation for the origins of the universe and of life, or as leaving many questions unanswered ("this was the best scientific theory on the subject to date, but it does not account for the other data and observations very well, and raises more questions than it answers"). They found that this manipulation affected associations between "science"/"God" and positive or negative words on a subsequent evaluation task, such that a weak scientific explanation boosted positive associations with God. In another study (Preston et al., 2013), participants read passages that offered weak or strong neuroscientific explanations for psychological phenomena, such as love. In the weak condition, the passages highlighted gaps in the neuroscientific explanations (e.g., "Neuroscientists note that these results only raise more questions than they answer. For instance it does not explain when or with whom we fall in love, and cannot explain what creates the feeling of love itself"). Participants who received the weak neuroscientific explanation were more likely than those who received the strong explanation to favor a title for the passage reflecting the explanatory gap (e.g., "the mystery of love" versus "the construction of love"), and also went on to show greater implicit belief in a soul (as reflected in the subjective value placed on the soul when given the opportunity to "sell" their soul—represented by a soul ID card—to the experimenter).

Woolley and Cornelius (2017) asked 5- to 9-year-olds and adults to evaluate the claim that there is "no way to explain" given mundane events (e.g., a dog waking up after falling asleep), improbable events (e.g., a dog waking up from a coma after two weeks at the vet), and extraordinary events (e.g., a dog waking up after being pronounced deceased by the vet). While endorsement of the "no way to explain"

claim was low across all event types, two interesting patterns emerged: first, only children younger than 8 years did not judge the "no way to explain" claim to be significantly worse than the neutral scale midpoint, suggesting a developmental trend in the belief that some things have no explanation or are unexplainable. Second, 8- and 9-year-olds and adults rated "no way to explain" as more acceptable for extraordinary events than for mundane events. While the extraordinary events used in this study were not explicitly religious or supernatural in nature, these findings are at least suggestive that "no way to explain" or mystery explanations may be more acceptable in the religious domain (where extraordinary events are often targets of explanation) than in the science domain (where mundane events tend to prevail).

Though the work reviewed in this section offers indirect support for the idea that declaring a mystery is more appropriate in answering questions about religion versus science, no research to our knowledge has directly tested whether mystery acceptability differs across domains. Additionally, the relationship between mystery acceptance and explanation remains an open question. Is declaring that something is a mystery equivalent to declaring that it does not demand an explanation? Or does appealing to mystery carry distinct or additional commitments? We address these questions in the present research.

1.3. Scientific and religious beliefs play different functional roles

Investigating need for explanation and mystery acceptability across the domains of science and religion is valuable because it sheds light on how inquiry is perceived and satisfied within each domain. However, these judgments also offer a window onto a much bigger and more elusive set of questions: Do scientific and religious beliefs differ only in their subject matter? Or is the nature of belief itself importantly different across domains? Prior theoretical and empirical work has identified distinct kinds of belief that may be differentially prevalent across the domains of science and religion (Buchak, 2012; Buckwalter et al., 2015; Heiphetz et al., 2013, 2014, 2018; Metz et al., 2018; Van Leeuwen, 2014), and a subset of this work further suggests that these kinds play different functional roles in human cognition. While both kinds of belief play a role in sensemaking (Chater & Loewenstein, 2016) and can offer a sense of control (Laurin & Kay, 2017; Rutjens et al., 2010; Rutjens et al., 2013), we suggest that scientific beliefs play a more central epistemic role (helping us form veridical representations of the world). This proposal is supported by past research investigating epistemic similarities and differences between scientific and religious beliefs.

Within psychology, research suggests that scientific and religious beliefs differ in how they are justified and in the (average) confidence with which they are held. Shtulman (2013) found that college students in the United States more often justified beliefs in scientific entities (such electrons or genes) vs. supernatural entities (such as souls or heaven) by appeal to evidence. Beliefs in supernatural entities were primarily justified by appeal to authority or with subjective justifications (such as appeal to intuition or volition), while beliefs in scientific entities were primarily justified by appeal to authority or evidence. Shtulman also found a notable difference in the confidence with which participants held these beliefs: while they were generally highly confident in the existence of both scientific and supernatural entities (selecting the highest possible rating for 46% of all selections), the average confidence with which they endorsed the existence of scientific entities was higher than that for supernatural entities. This confidence difference has been replicated cross-culturally (Clegg et al., 2019; Davoodi et al., 2018) and emerges in childhood (Harris et al., 2006): 5- and 6-year-old children are on average less confident in the existence of unobserved religious/supernatural entities (e.g., God, Santa Claus) than unobserved scientific entities (e.g., germs, oxygen), even though both are frequently endorsed by adults.

Additional evidence for differences in how scientific and religious beliefs are held and justified comes from Metz et al. (2018), who compared creationists and those who endorse evolutionary explanations for human origins, finding differences across these groups in which justifications for belief were judged generally acceptable. Across several studies, those who endorsed evolutionary explanations for human origins tended to endorse empirical evidence and consensus among scientists as strong justifications for belief, while those who embraced creationist explanations also endorsed religious authority and knowledge of the heart as strong justifications for beliefs, not only for creationism but in general. Additionally, endorsement of affiliative justifications for belief (e.g., the people I love believe it is true) predicted rejection of scientific beliefs like evolution and climate change (Metz et al., 2018).

While these findings reveal that there are often differences between scientific and religious beliefs, one possibility is that these beliefs differ only in their accessibility to investigation—we merely have more abundant and clear evidence about scientific phenomena (at least at the level of the scientific community), and this is reflected in both the confidence with which scientific beliefs are held and in the justifications that are given. However, another possibility is that there are multiple kinds of beliefs that play different functional roles in human cognition. Some recent proposals advocate for the latter position, arguing that religious beliefs can involve a faith-like dissociation from evidence that differs from a typical belief. One relevant proposal comes from Buchak (2012), who develops a formal analysis of faith as an epistemic attitude distinct from generic belief (see also Buchak, 2017). On her view, having faith in some claim involves a commitment to acting as if the claim is true *without first needing to examine additional evidence that could potentially bear on it*. Using formal arguments from decision theory and rational choice theory, Buchak identifies the conditions under which faith (in this sense) can be rational. Thus relative to generic belief, a proposition that is believed "on faith" has a distinct epistemic profile, whether its content is religious or scientific (for instance, non-experts could hold scientific beliefs "on faith").

Another proposal comes from Van Leeuwen (2014), who argues that religious "credences" differ from factual beliefs[1] in several ways. In particular, religious credences are less responsive to evidence than factual beliefs (for relevant evidence see Friesen et al., 2015). Moreover, religious credences are context dependent (i.e., only guide action in certain contexts), do not govern factual beliefs, have normative force, and are susceptible to special authority (authority appointed on non-evidential grounds, like many spiritual leaders). Van Leeuwen warns against taking the distinction too rigidly, suggesting that the two types characterize different "attractor positions" in an interdimensional space of epistemic attitudes, which may also include beliefs with mixed or intermediate features. Later, we will examine the possibility of finding these in folk conceptions of domains such as psychology or philosophy.

There is also some empirical evidence corroborating a conceptual distinction between factual belief and religious credence: adults are more likely to use the word "think" when referring to factual belief and "believe" when referring to religious credence (Heiphetz et al., 2018). Additionally, religious beliefs and scientific beliefs differ in ways that cannot readily be explained by higher levels of confidence in science relative to religion. For example, McPhetres and Zuckerman (2017) found that religious participants accepted weaker evidence for a

[1] On Van Leeuwen's account, "factual belief is an attitude we typically take toward contents so mundane as to be not worth mentioning, like *dogs have noses*, *silver is a metal*, or *the faucet spouts water*" (Van Leeuwen, 2014, p. 699). The distinction between factual belief and religious credence lies in the attitude, not in the content. However, it is plausible that the scientific contents that we consider are typically held with the attitude associated with factual belief, and that the religious contents that we consider are held with the attitude of religious credence. We therefore use factual and scientific interchangeably within this manuscript, but there may be interesting circumstances under which factual and scientific attitudes can diverge, and/or a belief with scientific content can be held as a religious credence. We revisit this in Section 5.

religious claim than for a scientific claim—a difference more consistent with greater prior confidence in the domain of religion. In another set of studies (Heiphetz et al., 2013), 5- to 10-year-old children and adults were asked to judge whether two people with differing beliefs on matters of fact, preference, and religion could both be right, or whether only one person could be right. Even for novel beliefs (for which prior confidence is undefined), children as young as 5 years old differentiated between religious beliefs and factual beliefs in their degree of objectivity, with factual beliefs being rated more objective. Religious beliefs and factual beliefs were both seen as more objective than preferences (for related research, see Flavell et al., 1992; Goodwin & Darley, 2008; Heiphetz & Young, 2017; Kuhn et al., 2000).

Why do scientific and religious beliefs have these distinct profiles? One possibility is that these profiles are uniquely suited to satisfy different functional roles (Tetlock, 2002). If a belief plays a predominantly epistemic role (such as supporting a maximally veridical representation of the world), one might expect this belief to be open to revision when counterevidence arises. Moreover, as this belief is taken to be accurate, it will be allowed to guide action and govern other beliefs. By contrast, if beliefs play an important role in individual identity or in signaling group membership (see also Kahan et al., 2012; Norenzayan, 2013; Sosis & Alcorta, 2003), one might expect a different profile. Specifically, one might expect this belief to be tied to personal identity or intuition, and to be sensitive to social factors (such as deference to authority or in-group cohesion). One might also expect weaker sensitivity to evidence: "If my credence that our god exists can be banished by something so trifling as mere evidence," Van Leeuwen writes, "how can you be sure that I am really committed to our group, which defines itself by allegiance to our god?" (Van Leeuwen, 2017, p. S56). Notably, the first (epistemic) role for belief aligns with empirical evidence on scientific belief, while the other (non-epistemic) role for belief aligns with empirical evidence on religious belief.

In sum, there are many documented differences between religious and scientific belief, from their perceived objectivity, to their responsiveness to evidence, to the ways they are justified. These differences could reflect something about the confidence with which beliefs in these domains are held—scientific beliefs tend to be endorsed with higher confidence than religious beliefs (Clegg et al., 2019; Davoodi et al., 2018; Harris et al., 2006; Shtulman, 2013). However, a more provocative proposal is that the difference between scientific and religious belief is more profound, with differences in objectivity, sensitivity to evidence, and justifications instead reflecting types of belief that play distinct functional roles in human cognition. While modest empirical evidence supports this latter proposal, few of the differences found in beliefs across domains have been dissociated from differences in confidence. Furthermore, these questions have not been explored through the lens of need for explanation and mystery acceptability, two judgments that may be particularly relevant given their potential ties to epistemic inquiry.

1.4. The present research

We predict that need for explanation and mystery acceptability will vary across scientific and religious domains. If scientific belief is largely tied to epistemic aims (forming and maintaining a veridical representation of the world), demanding an explanation for phenomena in the domain of science is appropriate because it may lead to explanatory inquiry, new evidence, and information gain. Conversely, declaring something a mystery is inappropriate for epistemically-aimed beliefs, because mystery suggests that the relevant belief cannot or should not be explained and thus terminates inquiry.[2] If, on the other hand, epistemic goals are less crucial for religious beliefs (which may instead be aimed towards maintaining personal identity or social cohesion), explanatory practice may differ as well, with lower demands for explanation and greater tolerance for mystery.

In the present research, we explore three questions regarding judgments of need for explanation (NFE) and mystery acceptability (MA). First, do judgments of NFE and MA differ across the domains of science and religion? Second, given that we do find such differences, what explains them? And third, what is the relationship between mystery and need for explanation?

An important hypothesis to consider (and potentially rule out) is that differences in NFE and MA are just a function of differential levels of confidence: We experience a higher level of NFE, and a lower level of MA, the stronger our belief in the truth or reality of the explanandum, regardless of the domain of the explanandum. After all, if a claim isn't true, it doesn't need explanation. For example, an individual could be more confident that rust is red (a belief in the science domain) than they are that God is good (a belief in the religion domain). This asymmetry in strength of belief could lead them to judge that the question "Why is God good?" is less in need of explanation than the question "Why is rust red?" and that declaring the former a mystery is more acceptable, solely because of the different levels of confidence with which these beliefs are held. We call this the *confidence confound* hypothesis, and we test it in Study 1 by investigating whether domain differences exist even for the most religious and supernaturally-inclined participants (i.e., those with high confidence concerning religious or supernatural explananda). In Study 2, we go beyond overall religiosity by measuring and statistically controlling for strength of belief in the premises of individual religion and science questions. Finally, in Study 3, we select only participants who hold both scientific and religious beliefs with an equal level of reported confidence, and control for strength of belief at an item-level and within-subjects. Across these increasingly stringent tests, we fail to find support for the confidence confound hypothesis.

An alternative hypothesis, the *domain-based difference* hypothesis, predicts that differences will persist even after equating strength of belief. Instead, differences in NFE and MA across domains reflect domain-based differences in the nature of belief: as suggested above, beliefs in different domains can play different functional roles.[3] For example, beliefs may serve a truth-tracking (epistemic) function in the domain of science, but a more affiliative function in the domain of religion (Van Leeuwen, 2014, 2017), or religious faith may have a distinct relationship with epistemic behaviors, such as evidence-gathering (Buchak, 2012, 2017). If this is the case, need for explanation—a signal that inquiry should be pursued—may be felt more strongly in the domain of science, where inquiry is a useful tool for gathering evidence, than in the domain of religion, where evidence is less relevant. Likewise, an explanation appealing to mystery—a signal that inquiry has been or will be futile—may be more acceptable in the domain of religion. Under this hypothesis, domain differences in NFE and MA can be mapped onto different functional roles that belief may play in different domains. In particular, whether belief serves an epistemic vs. a non-

[2] Whether inquiry is actually pursued on a given topic by a given individual depends on many factors beyond judgments of need for explanation or mystery acceptability: an individual could perceive direct inquiry to be too costly in time or effort to be worth the expected information gain, for example. Even in these cases, however, the judgment that inquiry is necessary (e.g., that something demands an explanation) or unnecessary (e.g., that something is a mystery) is likely to be related to whether the belief is geared towards epistemic aims. For example, an individual who holds a particular epistemically-aimed belief may not themselves pursue inquiry on that belief but may instead place their trust in others (e.g., domain experts) to pursue the relevant inquiry.

[3] Of course, there are many other domain differences beyond the functional role of belief that may explain any domain differences in NFE/MA. For example, individuals may have distinct ideas about the institutional frameworks of science and religion, their historical and present-day significance, etc. We revisit these additional features of science and religion in Section 5.

epistemic role in a given domain may determine judgments of NFE and MA in that domain. In Study 3, we test this account, finding evidence that domain differences can be partially explained by domain-specific properties of belief, which may map onto different functional roles.

Finally, we investigate a third question: what is the relationship between NFE and MA? Is mystery simply an abdication from explaining? Or do need for explanation and mystery track distinct features of a domain? In Study 2, we investigate this question by correlating NFE ratings and MA ratings, finding little overlap in NFE and MA. In Study 3, we measure how people interpret explanations that appeal to mystery and whether this differs across domains. Additionally, we find that the properties of belief that explain domain differences in NFE are partially distinct from those that explain domain differences in MA. These results suggest that NFE may persist whether one or multiple explanations are recognized as legitimate, while an appeal to mystery is seen as less appropriate in cases where only one explanation is believed to be legitimate.

Though our theoretical interest is in general differences between scientific and religious beliefs, we constrain our investigation of these research questions to the difference between scientific beliefs and Judeo-Christian beliefs, in a predominantly Judeo-Christian population in the United States. This may limit the generalizability of our findings (a point to which we return in Section 5), but allows us to constrain the scope of our investigation and is in keeping with much of the previous research we have reviewed. Additionally, we only explore beliefs that can be uncontroversially classified as "scientific" or "religious." Individuals can hold beliefs that combine both supernatural and natural elements (e.g., that prayer can cure disease) and can treat the same entities (e.g., disease) as caused by either natural or supernatural forces (e.g., germs or a curse). These instances of cross-domain beliefs (Legare et al., 2012) raise interesting questions for any account of belief, and we return to some of these questions in Section 5, but these cases are beyond the scope of the present research.

2. Study 1

In Study 1, we examine potential domain differences in NFE and MA. To do so, we ask participants to rate either NFE or MA in response to questions from various domains, including religion and science. Beyond testing for the presence of domain differences, we also conduct a preliminary test of the confidence confound hypothesis by measuring religious belief and paranormal belief, then testing whether domain differences remain even at the highest levels of religious and paranormal belief.

2.1. Method

2.1.1. Participants

Participants in Study 1 were 208 adults ranging from 20 to 69 years of age recruited from Amazon Mechanical Turk (124 identified as male, 84 as female, mean age 36, range 20 to 69). On average, participants were moderate in religious belief, *M (SD)* = 0.44 (0.31), and somewhat low in religious engagement, *M (SD)* = 0.27 (0.29), though scores on both measures spanned the full range from 0 to 1. Participants were also fairly low in paranormal belief, *M (SD)* = 2.45 (1.33), though again scores spanned the full range of 1 to 7. Participation was restricted to MTurk workers in the United States who had completed at least 50 prior tasks with a minimum approval rating of 95%. All participants provided informed consent before beginning the study. Eight additional participants were excluded for failing an attention check (described in Section 2.1.2), and one was excluded for failing to provide responses to more than half of the items.

2.1.2. Materials & procedure

Seventy questions were selected from Answers.com (http://www.answers.com), a website on which users can post questions for other users to answer. The questions were selected from seven domains, based on the website's classification. From each domain, 10 questions that contained the word "why" were extracted from the first 50 pages of questions, and edited lightly for grammar and readability. The domains used were science (primarily containing questions from the natural sciences, e.g., "Why don't objects pass through one another since atoms are mostly empty space?"), math (e.g., "Why do the angles of a quadrilateral add up to 360 degrees?"), health (e.g., "Why is calcium helpful in treating osteoporosis?"), religion and spirituality (e.g., "Why did God want Pharaoh to release the Israelites?"), supernatural and the occult (e.g., "Why are demons so powerful?"), psychology (e.g., "Why do people get addicted to computer games?"), and philosophy (e.g., "Why is truth valuable?").

Participants were randomly assigned to rate either NFE (N = 106) or MA (N = 102) for each question. Those who rated NFE read the instructions, "Some questions seem to demand an explanation in a way that others do not—they prompt us to wonder 'why?'. On the following pages, you will see a list of questions. For each one, we would like you to indicate to what extent the question demands explanation." They then responded on a 7-point scale to the prompt "To what extent does this question demand explanation?" for each question, with the scale endpoints labeled "does not demand explanation" and "demands explanation." Those who rated mystery acceptability read the instructions, "On the following pages, you will see a series of questions along with an answer to each question. We would like you to evaluate how good the answer is by indicating how well it explains what is being asked." They were also instructed to attempt to use the entire rating scale, even if they generally thought that all of the explanations were very good or very bad. Then, they were shown each question and the answer, "It's a mystery," and responded on a 7-point scale to the prompt "How good is this explanation?", with the scale endpoints labeled "very bad" and "very good." The order of questions was randomized, as was the position of an attention check, which instructed participants to select the sixth scale point. After rating NFE or MA for all 70 questions (randomized across domains), participants completed three measures in a random order: a religiosity inventory, a paranormal belief scale, and an epistemic preference scale. These three measures were drawn from previous research, as explained below.

To measure religious belief and religious engagement within a presumably predominantly Christian MTurk sample (see Berinsky et al., 2012), we used the religiosity inventory from Pennycook et al. (2012). This scale (see Appendix A) included three items measuring religious engagement (e.g., "Outside of attending religious services, how often do you pray?", α = 0.89) and six measuring religious belief (e.g., agreement with "There is a heaven where people who have led good lives are eternally rewarded," α = 0.94). To measure supernatural beliefs, we used a subset of the paranormal belief scale (Tobacyk, 2004). From this scale, we selected seven items most closely related to the Answers.com "supernatural and occult" items (e.g., agreement with "Psychokinesis, the movement of objects through psychic powers, does exist."), hoping to capture relevant differences in belief related to these items (α = 0.87). To measure epistemic preferences, we used the criteria for belief scale from Metz et al. (2018). This measure includes five subscales, which describe the extent to which people take scientific considerations (α = 0.82), religious considerations (α = 0.77), heart considerations (e.g., "It feels true in my heart," α = 0.90), affiliative considerations (e.g., "My parents believe it is true", α = 0.76), and explanatory considerations (α = 0.68) to be legitimate criteria for believing that a given premise is true. The criteria for belief ratings were collected for exploratory purposes and are not analyzed here.

Finally, at the conclusion of the study, participants reported their age and gender (male, female, or other/prefer not to specify).

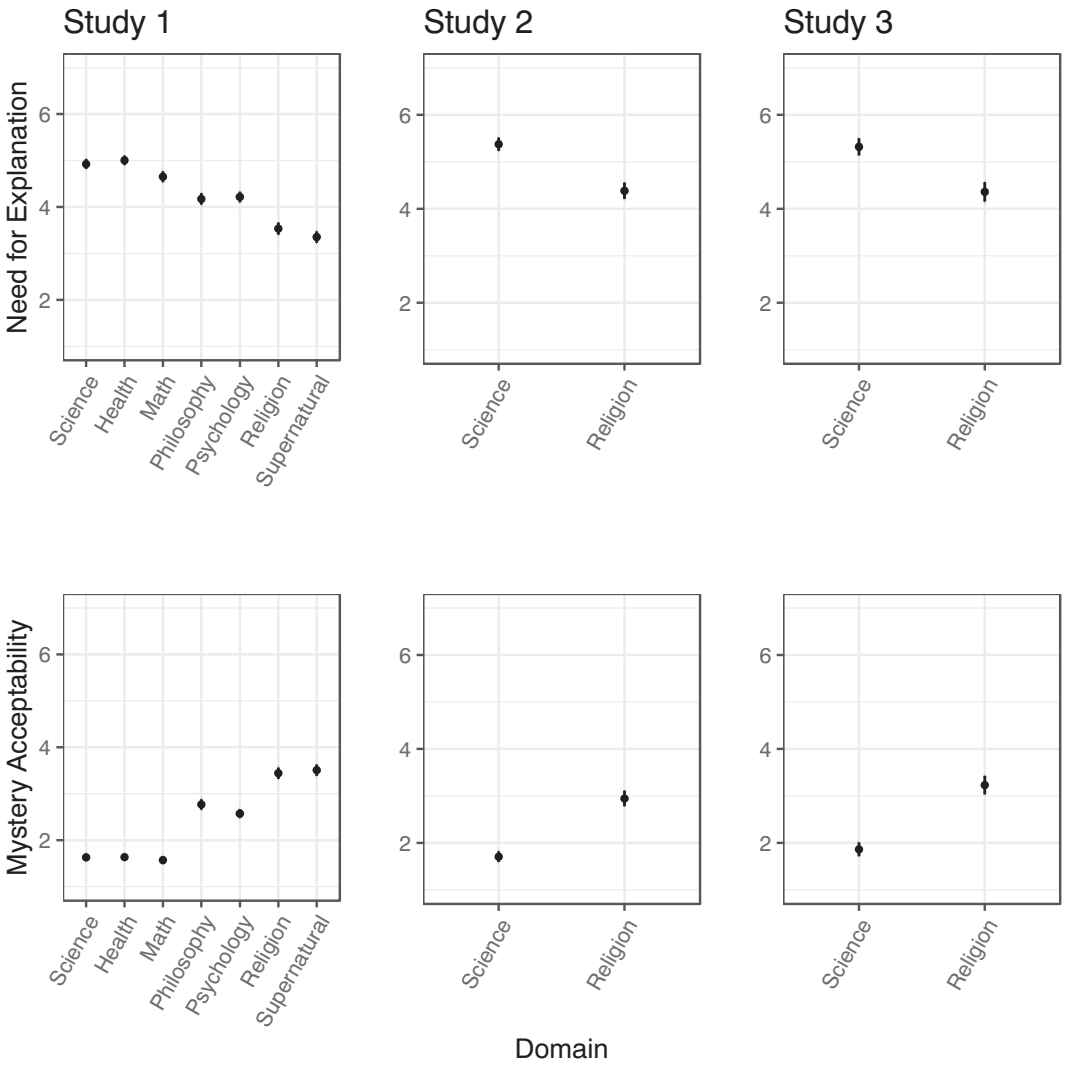

Fig. 1. Studies 1–3 mean ratings of NFE and MA for questions in all tested domains. Error bars = 95% CI. Where error bars are not visible, the confidence intervals are very narrow (limits within approximately 0.15 of the mean value).

2.2. Results

2.2.1. Domain differences in NFE and MA

First, we investigated domain differences in NFE and MA (see Fig. 1). For each measure, we fit a mixed-effects model predicting participant responses to each item. Domain was included in the model as a fixed effect (dummy coded, with science as the reference group). Random intercepts were included for participant and item, with items nested within domains. For NFE, there was a significant overall effect of domain, $\chi^2(6) = 87.13$, $p < .001$. Model coefficients revealed significantly higher ratings of NFE for questions about science than those about religion, $b = -1.39$, 95% CI [−1.73, −1.05], the supernatural, $b = -1.57$, 95% CI [−1.91, −1.24], psychology, $b = -0.71$, 95% CI [−1.05, −0.37], and philosophy, $b = -0.75$, 95% CI [−1.09, −0.41]. The ratings in the health domain, $b = 0.08$, 95% CI [−0.26, 0.42], and math domain, $b = -0.27$, 95% CI [−0.61, 0.06], were not significantly different from those in science.

For mystery acceptability, a complementary pattern of results emerged. There was an overall effect of domain, $\chi^2(6) = 126.73$, $p < .001$, with significantly lower ratings for science than religion, $b = 1.81$, 95% CI [1.50, 2.12], the supernatural, $b = 1.88$, 95% CI [1.57, 2.18], psychology, $b = 0.94$, 95% CI [0.64, 1.25], and philosophy, $b = 1.14$, 95% CI [0.83, 1.45]. Again, there was no difference between ratings in the health domain, $b = 0.004$, 95% CI [−0.30, 0.31], and math domain, $b = -0.06$, 95% CI [−0.37, 0.25], relative to ratings in the science domain.

These results provide support for a domain difference in need for explanation and mystery acceptability. Further, they suggest that domains outside of science and religion also reflect variability in NFE and MA: the domains of science, math, and health received similar ratings (high ratings for NFE, low ratings for MA), the domains of religion and the supernatural received similar ratings (lower ratings for NFE, higher ratings for MA), and the domains of philosophy and psychology received similar ratings (in each case falling between the other two sets).[4]

[4] It may be concerning to a reader of this journal that questions in the domain of psychology are separate from questions in the domain of science and that participants rate psychology questions distinctly from those in the science domain (we thank an anonymous reviewer for raising this point). We emphasize that these domain classifications are based on Answers.com's classification of questions. Questions under the "science" topic primarily concern the life sciences and physical sciences, while questions under the "psychology" topic concern human behavior. Prior work shows that people believe certain psychological phenomena to be outside of the scope of (scientific) explanation (Gottlieb & Lombrozo, 2018), and that even young children think of psychology as "easier" and distinct from natural science (Keil et al., 2010), potentially explaining asymmetries in participants' judgments of natural science vs. psychology questions.

2.2.2. Domain differences moderated by belief

Next, we conducted a preliminary test of the confidence confound hypothesis by testing whether domain differences in NFE and MA remained even when restricting analysis to the most religious/paranormally-inclined participants. First, we tested whether strength of religious belief (measured by the religious belief subscale of the religiosity inventory; Pennycook et al., 2012) moderated the differences in NFE and MA ratings between the domains of science and religion, and whether strength of paranormal belief moderated the differences in NFE and MA ratings between the domains of science and the supernatural. For each, a multilevel model was fit predicting NFE or MA, with domain and religious or paranormal belief (mean-centered over the entire sample) as regressors. Random intercepts were included for participant and for item, with the latter nested within domains. We tested for moderation by comparing a model including the interaction between domain and religious or paranormal belief with a model excluding this term, using likelihood ratio tests.

For NFE, religious belief significantly moderated the domain difference between science and religion, $\chi^2(1) = 52.51$, $p < .001$, and paranormal belief significantly moderated the domain difference between science and the supernatural, $\chi^2(1) = 24.94$, $p < .001$ (see Fig. 2). In both cases, the domain difference was smaller at higher levels of religious/paranormal belief. For MA, the same moderation effects were significant: religious belief moderated the domain difference between science and religion, $\chi^2(1) = 19.45$, $p < .001$, and paranormal belief moderated the domain difference between science and the supernatural, $\chi^2(1) = 39.54$, $p < .001$. These results suggest that belief in the content of the relevant domain is a moderator of domain differences in NFE and MA. In all cases, increasing levels of religious/paranormal belief *decreased* the domain difference in NFE/MA. Specifically, people with stronger belief in religion and the paranormal were more likely to express a need for explanation in those domains in addition to science, and they were more likely to accept "it's a mystery" as an adequate explanation in science as well as in religion/the supernatural. This suggests that different levels of confidence might indeed be a factor in determining NFE and MA.

Next, we tested whether NFE/MA domain differences still emerged among participants who most strongly endorsed religious/paranormal beliefs. To test this, we restricted analysis of domain differences to participants who scored in the upper quartile of the religious belief scale (to test domain differences between science and religion) and those who scored in the upper quartile of the paranormal belief scale (to test domain differences between science and the supernatural). Among those who were most religious, there were significant domain differences between science and religion in NFE, $b = -0.73$, 95% CI $[-1.24, -0.23]$, $\chi^2(1) = 7.42$, $p = .006$, and MA, $b = 1.37$, 95% CI $[1.03, 1.70]$, $\chi^2(1) = 29.42$, $p < .001$. Likewise, the difference between the domains of science and the supernatural was significant for both NFE, $b = -1.34$, 95% CI $[-1.72, -0.97]$, $\chi^2(1) = 25.94$, $p < .001$, and MA, $b = 1.39$, 95% CI $[1.05, 1.73]$, $\chi^2(1) = 29.88$, $p < .001$, among those who were highly paranormally inclined. These findings provide preliminary evidence against the confidence confound hypothesis—even among those who hold religious beliefs confidently (and who presumably hold the non-contentious science beliefs used in this study confidently, as well), NFE and MA are judged differently for questions premised on scientific beliefs compared to questions premised on religious beliefs.

2.3. Discussion

Study 1 offers evidence that NFE and MA differ across domains, and that these differences extend beyond the domains of science and religion. In particular, ratings of NFE were relatively high in the science, health, and math domains; moderate in philosophy and psychology; and low in religion and the supernatural. Ratings of MA were relatively low in science, health, and math; moderate in philosophy and

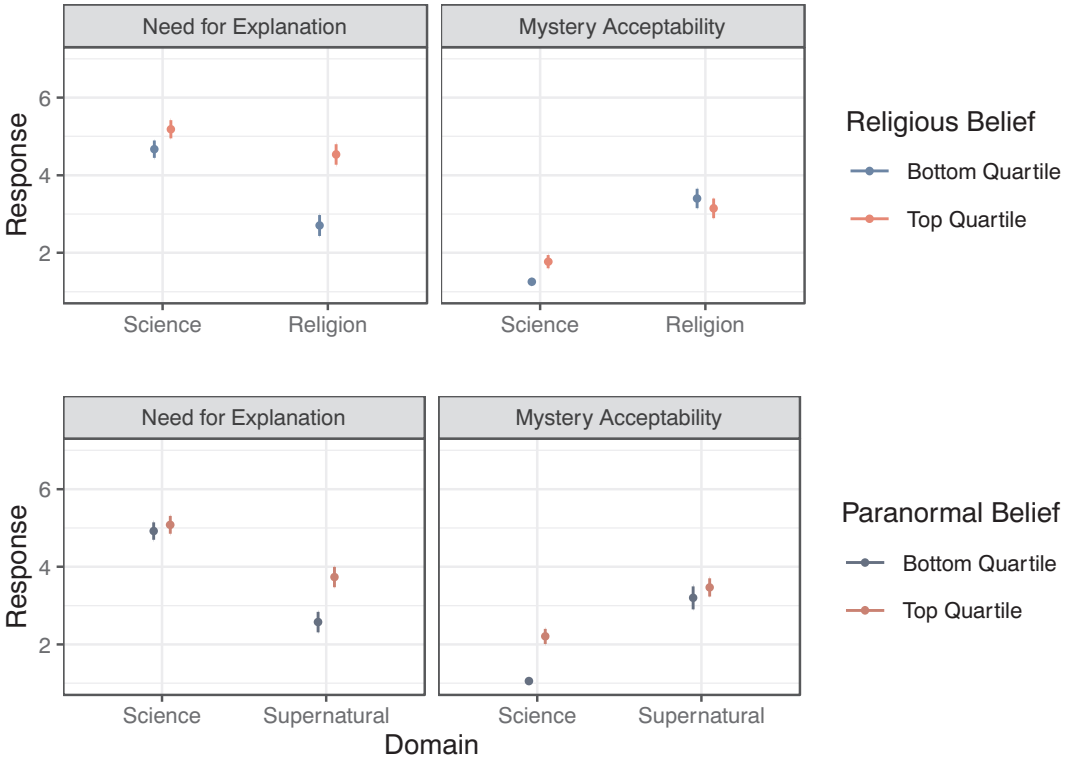

Fig. 2. Study 1 ratings of NFE and MA for questions in the science domain compared to the religion/supernatural domains, with a moderating effect of religious belief/paranormal belief (showing data from the top quartile and bottom quartile of religious belief in the top panel, and paranormal belief in the bottom panel). Error bars = 95% CI.

psychology; and higher in religion and the supernatural. Additionally, our results suggest that strength of belief (religious or paranormal) moderates the domain differences in NFE and MA. Most importantly, however, the differences in NFE and MA could not be fully accounted for by strength of belief; even restricting our analyses to participants in the highest quartile of religious/paranormal belief, domain differences remained. These findings provide preliminary evidence against the confidence confound hypothesis, which rests on any domain differences being explained by differing levels of confidence or strength of belief. Additionally, these results provide preliminary support for the hypothesis that NFE and MA are partially distinct (and thus that mystery cannot be interpreted as merely an abdication from the need to explain). In particular, while religious/paranormal belief moderated domain differences in both NFE and MA, increasing religious/paranormal belief was associated with a difference in NFE in the religion and supernatural domains (but not the science domain), and a difference in MA in the science domain (but not in the religion and supernatural domains; see Fig. 2). These asymmetries tentatively suggest differences in NFE and MA.

However, several open questions must be addressed. First, the ten questions we used from each domain, though sampled from a well-known online question-asking forum, may not be representative of "why" questions from each domain. Thus, in Study 2, we ask whether our findings replicate with a new set of questions sampled from a different source. Second, though Study 1 finds evidence against the confidence confound hypothesis by controlling for belief at a domain level, it remains possible that differing levels of belief in the premises of particular questions in fact explain the domain difference (e.g., someone could have a high level of religious belief overall, but doubt the specific claim that God had angels guarding the Garden of Eden). Moreover, we did not measure or control for strength of belief in science (in general) or in the specific premises of science questions. In Study 2, we conduct a stronger test by controlling for belief at an item-level in both domains of interest.

3. Study 2

In Study 2, we replicate the results of Study 1, using a more fine-grained measure of belief that allows us to investigate agreement with the premise of *each question* (in both science and religion) as a potential mediator of domain differences in NFE and MA. While general religious belief may be a rough proxy for item-level belief in the domain of religion, it cannot fully capture item-level differences. For example, if a person does not believe that Jesus healed the sick on the Sabbath, this person is unlikely to judge this event to be in need of explanation. However, this does not preclude their belief in *other* religious claims, and, perhaps, their judgment that questions about these claims demand explanation. Furthermore, an item-level measure of belief allows us to account for differences in agreement with scientific claims as well as religious claims. We also use new stimuli, restricted to the science domain and the religion domain, for which belief in the general US population is roughly matched. By matching general levels of belief across domains and using a fine-grained measure of belief for each question, we can conduct a more stringent test of the confidence confound hypothesis.

Study 2 had two additional aims. First, we also test an additional potential mediator of the domain difference in NFE and MA: the extent to which people believe they already know the answer to the question. One might expect that a person who thinks they know the answer to the question “Why do continents move?”, for example, might not feel the need for an explanation (since they already have one) and would not accept mystery as an acceptable answer. If participants' perceived knowledge differs across domains, then knowledge may also mediate domain differences in NFE or MA. Second, we test the relationship between NFE and MA more directly than in Study 1 by eliciting within-subjects ratings of both measures, allowing us to estimate the correlation between NFE and MA.

3.1. Method

3.1.1. Participants

Participants in Study 2 were 112 adults recruited from Amazon Mechanical Turk (70 identified as male, 42 as female, mean age 33, range 18 to 71). On average, participants were moderate in religious belief, *M (SD)* = 0.46 (0.29), and somewhat low in religious engagement, *M (SD)* = 0.28 (0.31), scales 0–1. Participants represented a range of education levels: 16% reported that they had graduated high school, 21% had completed some college, 56% had received a two- or four-year college degree, and 6% had completed a professional or doctoral degree. Participation was restricted to MTurk workers recruited as in Study 1. All participants provided informed consent before beginning the study. Seven additional participants who failed to pass two attention checks (described in Section 3.1.2) were excluded.

3.1.2. Materials & procedure

Five claims or questions about science and five claims or questions about religion were selected from several large-scale, representative national polls (Funk & Goo, 2015; Funk & Kennedy, 2016; National Science Board, 2014; Pew Research Center, 2015). These claims were selected so as to match general acceptance across domains, based on reported acceptance rates. Averaged over the five claims, general acceptance was 77% in the religion domain (*SD* = 0.14) and 77% in the science domain (*SD* = 0.07). Each claim was rewritten as a “why” question. For example, the claim “The center of the Earth is very hot” was rewritten as “Why is the center of the Earth so hot?” and the question “Do you believe in hell, where people who have lived bad lives and die without being sorry are eternally punished?” was rewritten as “Why is there a Hell?” (See Appendix A for all items).

All participants rated NFE, MA, belief, and knowledge for each question. NFE and MA were rated as in Study 1. For the belief measure, participants indicated on a seven-point scale how much they agreed with the question's premise (e.g., “Please rate your agreement with the following: that there is a Hell”). For the knowledge measure, participants were presented with the "why" question (e.g., “Why is there a Hell?”) and responded on a seven-point scale to the prompt “How confident are you that you know the answer to this question?”. The belief and knowledge measures were completed together in a single block, with the order of these items randomized. NFE and MA were completed separately in two additional blocks. The order of the three blocks was randomized.

Next, in a random order, participants completed the religiosity inventory from Pennycook et al. (2012), providing measures of religious engagement (α = 0.89) and religious belief (α = 0.95), and the criteria for belief scale from Metz et al. (2018). Again, the criteria for belief scale ratings were collected for exploratory purposes and are not analyzed here.

Within each scale, an attention check instructed participants to select a given option (“I disagree” for the religiosity inventory and “bad reason” for the criteria for belief scale). Finally, participants reported their age, gender, and highest level of education.

3.2. Results

3.2.1. Domain differences and religious belief

First, to test whether domain differences in NFE and MA persisted even using questions that were matched for agreement in the general population, we investigated general domain differences in NFE and MA. For each dependent variable, a mixed-effects model was fit to the data, with domain as a fixed effect (dummy coded with science as the reference group) and with random intercepts for participant. We initially also fit random intercepts for item (with items nested within domains), but these models resulted in singular fit (with estimated variance for the

random intercept equal to zero) for all models predicting NFE and MA. As a result, this random effect was excluded from all analyses of the Study 2 data unless otherwise noted (though fixed effect estimates for all analyses remained unchanged when the random effect was included).

There was a significant difference between domains for both NFE, $b = -0.99$, 95% CI [−1.18, −0.80], $\chi^2(1) = 100.27$, $p < .001$, and MA, $b = 1.24$, 95% CI [1.10, 1.38], $\chi^2(1) = 265.49$, $p < .001$, replicating the patterns from Study 1 (see Fig. 1): NFE was higher for science (M = 5.37, SD = 1.73) than for religion (M = 4.38, SD = 2.13), while MA was higher for religion (M = 2.95, SD = 2.00) than for science (M = 1.71, SD = 1.39).

Next, we tested whether religious belief moderated the domain difference, and whether the domain difference remained significant in the upper quartile of religious belief. Replicating Study 1, religious belief was a significant moderator of the domain difference in NFE, $\chi^2(1) = 29.13$, $p < .001$, with a narrowing domain difference as belief increased. However, religious belief was *not* a significant moderator of the domain difference in MA, $\chi^2(1) = 0.01$, $p = .912$. Furthermore, the domain difference in both NFE, $b = -0.51$, 95% CI [−0.94, −0.07], $\chi^2(1) = 5.27$, $p = .022$, and MA, $b = 0.93$, 95% CI [0.60, 1.26], $\chi^2(1) = 29.32$, $p < .001$, remained significant when restricting analyses to participants who scored in the top quartile of the religious belief scale.

3.2.2. Item-level belief and knowledge

Next, we tested whether the domain effect persisted even accounting for item-level belief and item-level knowledge across domains. Participants were more likely to agree with question premises in the domain of science than the domain of religion, $b = -2.75$, 95% CI [−3.03, −2.47], $\chi^2(1) = 38.14$, $p < .001$, and were more likely to report they knew the answers to questions in the domain of science than the domain of religion, $b = -1.33$, 95% CI [−1.60, −1.05], $\chi^2(1) = 24.36$, $p < .001$ (controlling for participant and item with random intercepts). Therefore, it was possible that the domain differences in NFE and/or MA could be fully explained by the corresponding domain differences in item-level belief or knowledge.

First, we tested whether item-level belief *moderated* the effect of domain on NFE or MA—that is, is the size of the domain difference in NFE/MA different at distinct levels of item-level belief? We found no significant interaction between domain and item-level belief in a mixed-effects regression model predicting NFE, $\chi^2(1) = 1.14$, $p = .285$. Thus, there is no evidence that strength of belief moderates the magnitude of the domain difference in NFE. In a mixed-effects regression model predicting MA, the interaction between domain and item-level belief approached significance, $\chi^2(1) = 3.83$, $p = .050$. For the domain of science, every one-point increase in item-level agreement decreased MA judgments by 0.11 points, 95% CI [−0.19, −0.02]. However, the interaction effect reduced the effect of item-level agreement in the domain of religion to virtually zero, interaction: $b = 0.10$, 95% CI [0.00, 0.21]. Thus, for stronger beliefs relative to weaker beliefs, the domain difference in MA was larger.

We additionally planned to test whether item-level belief *mediated* the relation between domain and NFE/MA—that is, can the domain difference in NFE/MA be explained in part by a domain difference in the confidence with which the relevant beliefs are held? Mediation analyses are potentially problematic, in that they make some critical assumptions that are rarely, if ever, met in typical datasets (Bullock et al., 2010; Green et al., 2010). For example, it is assumed the variables are temporally and causally ordered (e.g., that domain causes confidence and NFE/MA and that confidence causes NFE/MA) and that there are no omitted variables that influence the relations in the mediation model (MacKinnon, 2008), assumptions that may not be met in the present correlational, cross-sectional dataset. Additionally, mediation analyses should not be conducted if there is an interaction between the independent variable (in our case, domain) and the mediating variable (Jacoby & Sassenberg, 2014; MacKinnon, 2008), which is the case for MA in our dataset. In the remainder of this paper, we report mediation results for cases where interactions between the independent variable and mediating variable were non-significant. However, we urge caution in interpreting these results as definitive evidence in favor of the mediation model, due to the high probability that our data fail to meet the assumptions outlined above—instead, the results offer one possible pathway relating the variables we measured, which must be further tested and compared against competing models in future research.

To test whether item-level belief mediated the relation between domain and NFE, we used the "mediation" R package (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014), which computes the magnitude of the direct and indirect effects using a quasi-Bayesian Monte Carlo simulation. According to this analysis (with 1000 random draws), item-level belief was a significant mediator of the relationship between domain and NFE, with an estimated indirect effect of −0.40, 95% CI [−0.57, −0.26], $p < .001$. The remaining direct effect, however, remained significantly different from zero, $b = -0.59$, 95% CI [−0.82, −0.35], $p < .001$. Thus, while item-level belief differed across domains and was positively associated with NFE judgments, the domain difference in NFE was not fully explained by item-level belief.

For MA, we do not test for mediation because the interaction between domain and item-level belief was marginally significant. Instead, to test whether the domain difference held at all levels of belief, we restricted analyses to the bottom quartile of item-level belief (as this is the quartile in which the domain difference is smallest, given the analysis reported above). Even for items rated within the bottom quartile of item-level belief, the domain difference in MA between science and religion remained significant, $b = 1.04$, 95% CI [0.34, 1.73], $\chi^2(1) = 8.35$, $p = .004$. Thus, though low levels of confidence in the premise of a question were associated with increased acceptance of mystery as an explanation in the domain of science, domain differences remained even when strength of belief was lowest (and thus, when MA ratings for science were closest to those for religion).

We repeated the above analyses to test whether item-level knowledge (perceived knowledge of the answer to a given question) could explain domain differences in NFE and MA. For NFE, item-level knowledge was a significant moderator of the domain difference: there was a significant interaction between domain and item-level knowledge in predicting NFE, $\chi^2(1) = 18.78$, $p < .001$. As one's perceived knowledge of the answer to a question increased, the size of the domain difference also increased. However, even in the bottom quartile of item-level knowledge (where the domain difference was smallest), the domain difference remained significant, $b = -0.56$, 95% CI [−0.98, −0.15], $\chi^2(1) = 7.13$, $p = .008$. For MA, there was also a significant moderating effect of item-level knowledge, $\chi^2(1) = 4.79$, $p = .029$. As item-level knowledge increased, the domain difference in mystery acceptability decreased. However, among the highest ratings of item-level knowledge[5] (where the domain difference was smallest), the domain difference in MA was still significant, $b = 0.79$, 95% CI [0.52, 1.06], $\chi^2(1) = 32.04$, $p < .001$. Because item-level knowledge was a significant moderator of domain differences for both NFE and MA, we did not test for mediation.

3.2.3. Relationship between NFE and MA

Finally, we investigated the relationship between NFE and MA by estimating the correlation between these variables, independently within the science domain and the religion domain. For questions in the science domain, NFE and MA were significantly though weakly negatively correlated, $r = -0.11$, $t(558) = -2.57$, $p = .010$. For questions

[5] We were unable to use quartiles for this analysis because just over 30% of the data used the highest scale point of item-level knowledge. Instead, we restricted this analysis to this top 30%.

in the religion domain, NFE and MA were not significantly correlated, $r = 0.01$, $t(558) = 0.22$, $p = .824$. These results suggest that NFE and MA are distinct constructs.

3.3. Discussion

Study 2 replicates several key findings from Study 1 while matching confidence in belief at a population level and an item level. While item-level belief differed across domains, it did not fully explain the different attitudes towards NFE and MA between questions of science vs. religion. Item-level belief did not fully mediate the association between domain and NFE, and analyses restricted to the subset of participants/items for which the domain difference in MA was smallest (on the basis of item-level belief ratings) still revealed a significant effect of domain. This was also the case for item-level knowledge.

These findings challenge the confidence confound hypothesis. However, while strength of belief was well controlled across participants, it is possible that NFE for religion questions with high levels of belief and NFE for science questions with high levels of belief were rated by *different participants*. That is, those participants that agreed strongly with the premises of religious questions may not have agreed equally strongly with the premises of scientific questions, and vice versa. It is possible that these patterns reflect two different types of participant, one that believes more strongly in science and has a high need for explanation and low acceptance of mystery, another that believes more strongly in religion and has a lower need for explanation and higher acceptance of mystery. Therefore, while strength of belief does not seem to account for the domain differences in NFE and MA at a population level, it remains possible that strength of belief fully explains the domain differences in NFE and MA *within an individual*. In Study 3, we conduct a maximally-stringent test of the confidence confound hypothesis, by matching strength of belief at an item-level within participants.

The findings from Study 2 also suggest that mystery acceptability is not just a consequence of low NFE, or vice versa. The correlations between MA and NFE were weak (in the domain of science) or non-significant (in the domain of religion), and MA and NFE were differentially influenced by strength of belief. This suggests that MA and NFE may reflect different aspects of scientific and religious belief—an issue we also revisit in Study 3.

4. Study 3

In Study 3, we replicate the domain differences found in Studies 1 and 2, controlling for item-level strength of belief *within individuals*. That is, we can ask: if an individual believes equally strongly in a religious claim and a scientific claim, is there still a domain difference in their judgments of NFE and MA for the corresponding questions? Or does item-level belief fully explain the domain difference when matched within individuals?

Additionally, we test five potential mediators of these domain differences, which map onto potential functional roles for belief within the domains of science and religion. In particular, a belief could serve an epistemic functional role—helping us achieve a more veridical representation of the world—or a non-epistemic functional role—such as helping us signal individual or group identities (among other possibilities). We measure four epistemic judgments—the perceived limits of human understanding relative to the target belief's explanation ("epistemic limits"), perceptions of whether there is a "fact of the matter" about the target belief's explanation ("subjectivism"), the salience of counterfactual alternatives to the target belief[6] ("counterfactual salience"), and participants' perceived knowledge of the target belief's explanation ("knowledge")—and one non-epistemic judgment—social norms about whether the target belief *should or should not* be explained ("explanatory norms").

If beliefs serve different functional roles across the domains of science and religion, we might expect judgments of epistemic limits, subjectivism, counterfactual salience, knowledge, and explanatory norms to also differ across domains. And, to the extent that these judgments do differ across domains, we can test whether these judgments mediate domain differences in NFE and MA. That is, in Study 3, we ask: Do judgments of the epistemic and non-epistemic functional roles for belief mediate the domain differences in NFE and MA? A positive answer to this question would indicate that domain differences arise at least in part because of different epistemic or non-epistemic features of beliefs in each domain, providing support for the domain-based difference hypothesis.

Finally, we explore how people interpret the declaration that something is a mystery. The small correlations between NFE and MA in Study 2 suggest that acceptance of mystery is not directly parallel to low levels of NFE. An explanation appealing to mystery may not simply indicate that an explanation is currently or in principle unavailable, raising the questions of whether mystery is itself taken to be an explanation (e.g., as an explanatory concept similar to a miracle: see Woolley & Dunham, 2017), and whether this is something that differs across domains. To explore these questions, we test two alternative interpretations of the phrase "it's a mystery" in response to a "why" question. First, it is possible that appeal to mystery signals the respondent's belief that *there is no explanation* that can be given in response to the question. Second, it is possible that appeal to mystery is taken to be a genuine explanation in response to a question. In Study 3, we measure which of these possible interpretations participants endorse, as well as whether the endorsed interpretation differs across domains, as a window onto the relationship between NFE and MA.

4.1. Method

4.1.1. Participants

Participants in Study 3 were 324 adults recruited from Amazon Mechanical Turk (128 identified as male, 196 as female, mean age 37, range 19 to 79). Participation was restricted to MTurk workers recruited as in previous studies. However, following Metz et al. (2018), participation was restricted to workers in the following states: Alabama, Arkansas, Georgia, Kentucky, Louisiana, Mississippi, North Dakota, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, and West Virginia. These states have a relatively high proportion of Evangelical Christians, and thus increase the proportion of our sample that is likely to endorse both religious and scientific phenomena with equal confidence, counteracting the tendency of MTurk samples to be less religious than the general population (Berinsky et al., 2012). In fact, participants were on average reasonably high in religious belief, *M (SD)* = 0.72 (0.21), and moderate in religious engagement, *M (SD)* = 0.53 (0.26), scales 0–1. Participants represented a range of education levels: 12% reported that they had graduated high school, 25% had completed some college, 53% had received a two- or four-year college degree, and 10% had completed a professional or doctoral degree. All participants provided informed consent before beginning the study. Eleven additional participants who failed to pass an attention check (described in Section 4.1.2) were excluded. An additional 169 participants passed the attention check but were not included in analyses because they did not agree or strongly agree with any item from either the science or religion domains (see Section 4.1.2 for further

[6] According to Grimm (2008), an event or phenomenon stands in need of explanation to the extent that there is some salient alternative way the world could have been. For instance, the fact that continents move may demand explanation because we can easily imagine a world in which continents remain stationary.

details).

4.1.2. Materials & procedure

We again used five science questions and five religion questions, matched for agreement in the general population. These were modified slightly from the questions used in Study 2, detailed in Appendix A.

Participants first rated their agreement with the premise of each question ("item-level belief") on a seven-point scale, using a similar measure to that used in Study 2. Based on these ratings, one science question and one religion question were identified for additional ratings. These questions were chosen as follows: If a participant responded "strongly agree" to one or more items from each domain, one strongly agreed-to item per domain was randomly selected. If the participant did not respond "strongly agree" to at least one item from each domain, but did respond "agree" to at least one item from each domain, one item per domain was randomly selected that received this "agree" rating. If the participant fell into neither of these categories, the participant would be (if possible) assigned an item from one domain that was rated "agree" and an item from the other domain that was rated "strongly agree." However, if no items within a domain were rated "agree" or "strongly agree" (6 or 7 out of 7), the participant did not complete ratings for any questions within that domain (and these participants were excluded). This procedure resulted in a set of participants who rated a science item and a religion item matched (within one point) for high levels of item-level belief: 206 participants who strongly agreed with an item from both domains, 79 participants who agreed with an item from both domains, and 39 participants who strongly agreed with an item from one domain and agreed with an item from the other domain.

With a target science and religion question thus identified, participants rated NFE, MA, and knowledge as in Study 2. Participants also completed five additional measures on seven-point scales (epistemic limits, subjectivism, counterfactual salience, explanatory norms, and mystery interpretation), shown in Table 1.

Next, participants completed the religiosity inventory from Pennycook et al. (2012), again providing measures of religious engagement ($\alpha = 0.78$) and religious belief ($\alpha = 0.87$). An attention check instructed participants to select a given option ("I disagree"), and participants who did not select the correct option were excluded. Finally, participants reported their age, gender, and highest level of education.

4.2. Results

4.2.1. Domain differences in NFE and MA

First, we tested whether domain differences in NFE and MA were still significant, even having matched high levels of belief within one point within individual participants. As in previous studies, a mixed-effects model was fit to the data, with domain as a fixed effect and with random intercepts for participant. As in Study 2, we initially fit all models with random intercepts for item (nested within domain), but this resulted in singular fit for several models. This random effect was thus excluded for all analyses predicting NFE and MA in Study 3, but fixed effect estimates remained similar when it was included. For models with other dependent measures, random effects for item were included.

The effect of domain was significant for both NFE, $b = -0.96$, 95% CI [−1.20, −0.71], $\chi^2(1) = 54.47$, $p < .001$, and for MA, $b = 1.37$, 95% CI [1.15, 1.58], $\chi^2(1) = 128.93$, $p < .001$ (see Fig. 1). As in Studies 1 and 2, NFE was higher for science questions (M = 5.32, SD = 1.70) than for religion questions (M = 4.36, SD = 1.91), whereas MA was higher for religion questions (M = 3.23, SD = 1.84) than for science questions (M = 1.86, SD = 1.39).

Despite having matched belief within one point (agree or strongly agree), a regression model predicting item-level belief with domain as a fixed effect (and random intercepts for participant and item) revealed that a domain difference in item-level belief persisted, $b = 0.11$, 95% CI [0.03, 0.18], $\chi^2(1) = 6.38$, $p = .011$, with beliefs in the domain of science (M = 6.76, SD = 0.43) being held slightly more strongly than beliefs in the domain of religion (M = 6.64, SD = 0.48). Thus, we evaluated whether differences in strength of belief could account for the domain difference in NFE and MA using the same analyses as in Study 2. The interaction between domain and item-level belief was not significant in predicting NFE, $\chi^2(1) = 0.40$, $p = .529$, nor in predicting MA, $\chi^2(1) = 0.02$, $p = .892$. Thus, we tested for mediation for both variables. For NFE, item-level belief was not a significant mediator: the indirect effect was equal to −0.02, 95% CI [−0.06, 0.01], $p = .250$. For MA, item-level belief was a partial mediator of the domain difference, though only 4% of the total effect of domain on MA was accounted for by the indirect effect. The estimated indirect effect was equal to 0.06, 95% CI [0.02, 0.10], $p < .001$, while the remaining direct effect was equal to 1.30, 95% CI [1.09, 1.53], $p < .001$. These results suggest that the domain difference in NFE and MA held even controlling for belief at an item level within participants—a maximally stringent test of the confidence confound hypothesis.

4.2.2. Functional roles across domains

Having established that strength of belief does not explain domain differences in NFE and MA, we next turned to exploring the functional role measures. For the five additional measures that capture a set of potential epistemic and non-epistemic functional roles (knowledge, subjectivism, epistemic limits, explanatory norms, and counterfactual salience), we first investigated differences across the domains of science and religion. Predicting each measure, we fit a regression model with domain as a fixed effect (dummy coded, with science as the reference group) and with random intercepts for participant and item, with the latter nested within domain. Judgments of subjectivism, epistemic

Table 1
New measures assessing functional roles of belief and mystery interpretation in Study 3.

Measure	Item text	Response scale
Epistemic limits	Do you think the answer to this question is beyond human comprehension?	Definitely within human comprehension – Definitely beyond human comprehension
Subjectivism	Suppose two people provide two different answers to this question. Do you think both people can be right?	Only one can be right – Both can be right
Counterfactual salience	Do you think it's possible that the world could have been such that [there *is not* a God]?	Definitely not possible – Definitely possible
Explanatory norms	Do you think people *shouldn't* try to answer this question?	Definitely shouldn't – Definitely should
Mystery interpretation	Question: Why is there a God? Anne's response: It's a mystery. Which of the following is a better characterization of Anne's response? (A) Anne is saying that this cannot be explained. (B) Anne is offering an explanation: the explanation that it is a mystery.	(A) is definitely better – (B) is definitely better

Table 2
Mean (SD) rating of each functional role measure for questions in each domain, and the results of likelihood ratio tests assessing domain differences.

Measure	Science mean (SD)	Religion mean (SD)	Likelihood ratio test
Knowledge	4.56 (2.02)	4.13 (2.24)	$\chi^2(1) = 0.50, p = .478$
Epistemic limits	2.10 (1.60)	4.64 (2.00)	$\chi^2(1) = 29.19, p < .001$
Subjectivism	3.45 (2.08)	4.90 (2.01)	$\chi^2(1) = 19.69, p < .001$
Counterfactual salience	3.21 (2.01)	3.44 (2.11)	$\chi^2(1) = 1.48, p = .224$
Explanatory norms	5.96 (1.57)	4.94 (1.73)	$\chi^2(1) = 18.24, p < .001$

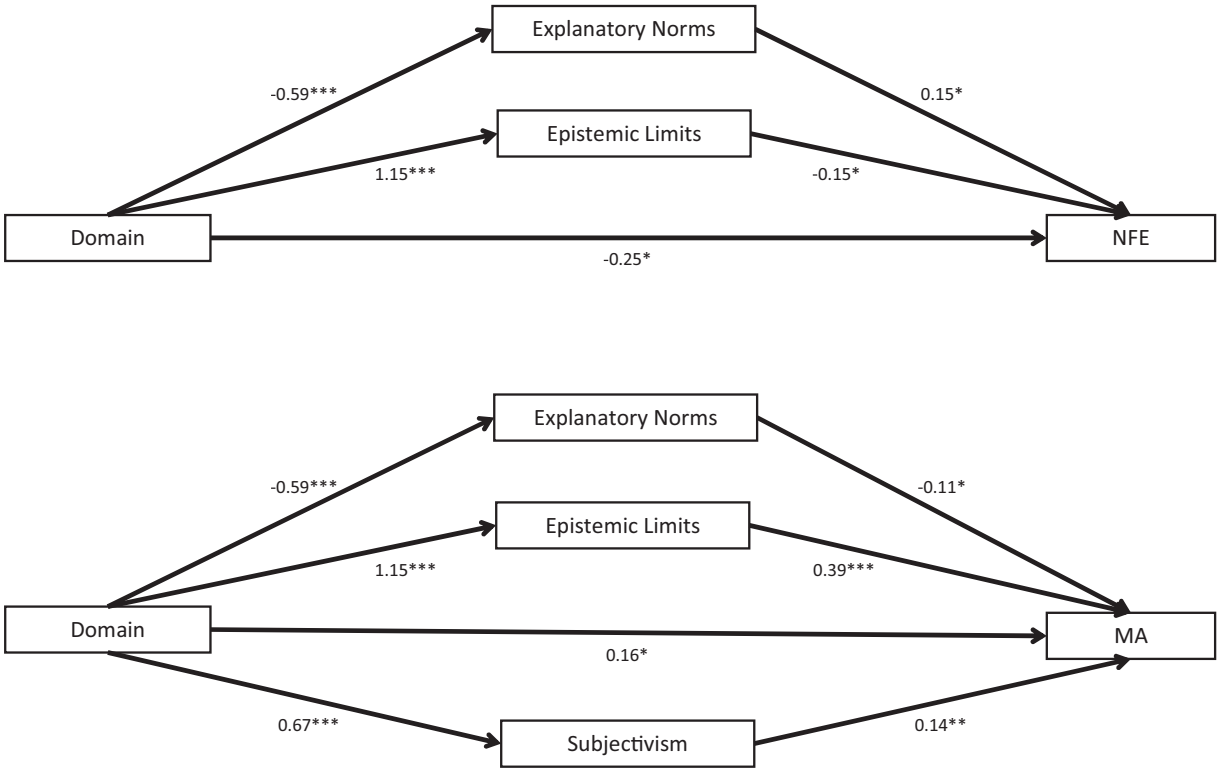

Fig. 3. Path diagrams displaying mediators of the domain difference in NFE and MA in Study 3, in which item-level belief was matched for strong agreement within one point. ***$p < .001$, **$p < .01$, *$p < .05$.

limits, and explanatory norms differed across domains (see Table 2), with participants judging that questions in the domain of science tended to have only one right answer, were within human comprehension, and should be explained. In contrast, participants tended to judge that questions in the domain of religion could have multiple answers and were beyond human comprehension. As with science, participants tended to judge that religious questions should be explained—but to a lesser degree than science questions. Knowledge, counterfactual salience, and interpretation of mystery as explanatory or non-explanatory did not differ across domains.

For the measures that did not differ significantly across domains, we fit an additional mixed-effects model predicting each measure (recoded so that the midpoint value was equal to zero) with only an intercept and random effects for participant and item. Using the "lmerTest" package in R (Kuznetsova et al., 2017), we estimated the statistical significance of the intercept to determine whether average ratings were different from the scale midpoint. Participant ratings of knowledge were marginally higher than the scale midpoint, $b = 0.44$, 95% CI [0.01, 0.88], $t(9.21) = 2.10$, $p = .065$. Ratings of counterfactual salience were significantly lower than the midpoint, $b = -0.70$, 95% CI [−0.92, −0.49], $t(10.24) = -6.67$, $p < .001$, suggesting that across domains, participants judged that counterfactual alternatives to these particular religious and scientific propositions were not very plausible.

4.2.3. Potential mediators of NFE and MA

Next, we tested whether the functional role measures that differed across domains (subjectivism, epistemic limits, and explanatory norms) could explain the effect of domain on NFE and MA. We first tested whether any of the functional role measures moderated the effect of domain on NFE/MA. In all cases, the interaction between domain and a given functional role measure was non-significant. We then used structural equation modeling (SEM) to analyze whether these measures mediated the effect of domain on NFE and MA (Gunzler et al., 2013). By using SEM rather than a traditional mediation analysis, we were able to estimate the mediating effects of all variables in a single model, though similarly to standard mediation analyses, SEM suffers from many limitations and has been criticized (see Bullock et al., 1994; Tomarken & Waller, 2005). Again, the SEM results should not be taken as definitive evidence in support of the tested model, but rather as a suggestion for an empirically plausible model that must be tested (ideally experimentally) in future research. For each dependent measure (NFE/MA), a multilevel SEM model was specified in the R package lavaan (Rosseel, 2012), with participant as the clustering variable. All variables entered into the SEM model were z-scored. At level one, paths were estimated from domain (dummy coded, with science as the reference group) to NFE/MA, from domain to each potential mediator, and from each potential mediator to NFE/MA. At level two, we specified a saturated model, containing the variances and covariances of all endogenous

variables (Wu & Kwok, 2012). The parameter values were estimated using maximum likelihood estimation, and the significance of these parameters was computed using the Wald statistic.

For NFE, the indirect effect of explanatory norms was significant, $\beta = -0.08$, 95% CI [−0.15, −0.02], $z = -2.39$, $p = .017$, as was the indirect effect of epistemic limits, $\beta = -0.15$, 95% CI [−0.29, −0.002], $z = -1.99$, $p = .047$, but the indirect effect of subjectivism was not significant, $\beta = -0.06$, 95% CI [−0.14, 0.02], $z = -1.52$, $p = .130$. Thus, we refit a model for NFE including only explanatory norms and epistemic limits as potential mediators. In this revised model (see Fig. 3), both explanatory norms, $\beta = -0.09$, 95% CI [−0.16, −0.02], $z = -2.52$, $p = .012$, and epistemic limits, $\beta = -0.18$, 95% CI [−0.32, −0.04], $z = -2.46$, $p = .014$, had significant indirect effects. However, the SEM model only explained 13% of the variance in NFE ratings. For MA, the indirect effects of explanatory norms, $\beta = 0.06$, 95% CI [0.01, 0.12], $z = 2.23$, $p = .025$, epistemic limits, $\beta = 0.45$, 95% CI [0.32, 0.58], $z = 6.85$, $p < .001$, and subjectivism, $\beta = 0.10$, 95% CI [0.03, 0.16], $z = 2.89$, $p = .004$ were significant. As all mediators were significant, we retained this full model for MA (see Fig. 3). The SEM model explained 38% of the variance in MA.

4.2.4. What is a mystery?

Finally, we tested how participants interpreted an appeal to mystery in response to a "why" question, as well as whether the dominant interpretation of mystery differed across domains. The mean rating for the mystery interpretation measure was 3.24 (SD = 2.10) in the domain of science and 3.35 (SD = 2.20) in the domain of religion. Collapsing across domains, the average rating was significantly below the scale midpoint, $b = -0.70$, 95% CI [−0.98, −0.41], $t(10.52) = -4.95$, $p < .001$. This suggests that participants tended towards interpreting mystery as an abdication from explaining, rather than as an explanation itself. A regression model predicting mystery interpretation with domain as a fixed effect and with random intercepts for participant and item revealed no significant effect of domain, $\chi^2(1) = 0.36$, $p = .550$. While these results suggest that mystery may in fact be taken to be an abdication from explaining, we again found only a weak correlation between NFE and MA for both science questions, $r = -0.14$, $t(322) = -2.58$, $p = .010$, and religion questions, $r = -0.18$, $t(322) = -3.27$, $p = .001$.

4.3. Discussion

While Study 2 provided evidence against the confidence confound hypothesis at a group level, Study 3 provides evidence against this hypothesis at an individual level. Even within individuals, holding strength of belief fixed, domain differences in NFE and MA persisted. After establishing this, we carried out a test of the domain-based difference hypothesis, finding that epistemic limits and explanatory norms partially mediated the domain difference in NFE and that epistemic limits, explanatory norms, and subjectivism partially mediated the domain difference in MA. While these individual measures explained limited variance in NFE and MA, we take these findings as preliminary evidence for the domain-based difference hypothesis. They are suggestive that a suite of domain-varying factors—related to the epistemic and non-epistemic roles for belief—account for domain differences in NFE and MA. Additionally, they suggest that explanation and mystery may have somewhat distinct functions: subjectivism (the judgment that two people could have different conflicting beliefs and both be correct) mediated the domain difference in MA but *not* NFE. One interpretation is that NFE guides the pursuit of inquiry towards questions that *should* be explained and that are within human comprehension, regardless of whether there is one vs. multiple "true" explanation(s). Declaring something a mystery, in contrast, could halt inquiry (because an explanation should not or cannot be pursued), but additionally signals that typical criteria for explanatory evaluation should be suspended, because there is no single, objective answer. Thus, while mystery may in fact be seen as an abdication from explaining, the asymmetrical role of subjectivism in driving MA relative to NFE could partially explain why these constructs diverge.

5. General discussion

In the present research, we investigated domain differences in "need for explanation" and "mystery acceptability," two judgments that plausibly reflect epistemic commitments and play a role in regulating inquiry. Taken together, these studies suggest the existence of genuine domain differences in judgments of both kinds. In the domain of science, "why" questions were generally judged to be strongly in need of explanation, and appeal to mystery was not judged to fulfill this need. However, in the domain of religion, "why" questions were judged to be somewhat less in need of explanation, and explanations that appeal to mystery were judged to be more acceptable (though still not very good).

Though we replicated previous findings that beliefs in the domain of science were held with higher confidence than beliefs in the domain of religion (Clegg et al., 2019; Davoodi et al., 2018; Harris et al., 2006; Shtulman, 2013), this domain difference in confidence could not fully explain the domain differences in NFE and MA judgments. In Study 1, we found domain differences in NFE and MA across all levels of religious belief. In Study 2, we found domain differences in NFE and MA controlling for confidence at an item-level. And in Study 3, we found domain differences in NFE and MA controlling for confidence at an item-level within participants. These results are the first (to our knowledge) to demonstrate differences between real-world scientific and religious beliefs while controlling for differences in confidence or strength of belief.

Additionally, we found preliminary support for the idea that there are domain differences in the function of belief—that is, that scientific beliefs are aligned with primarily epistemic aims, while religious beliefs are (at least partially) aligned with non-epistemic aims. Controlling for confidence, questions in the domain of science were mostly judged to have only one "right" answer, to be within human comprehension, and to be normatively appropriate to explain, while questions in the domain of religion were judged to have several possible "right" answers (see also Heiphetz et al., 2013), to be beyond human comprehension, and to be less normatively appropriate to explain—differences that plausibly map onto whether epistemic or non-epistemic aims are privileged. Furthermore, these domain differences mediated domain differences in NFE and MA: domain differences in NFE were partially explained by domain differences in explanatory norms and epistemic limits, while domain differences in MA were partially explained by domain differences in explanatory norms, epistemic limits, and the subjective nature of belief.

Our findings also shed light on the relationship between explanation and mystery. In particular, we found evidence that NFE and MA are at least partially distinct constructs; accepting a mystery is not equivalent to denying a need for explanation. Studies 2 and 3 found little or no correlation between NFE and MA, and Study 3 revealed that domain differences in MA were partially mediated by beliefs about the subjective nature of belief, while domain differences in NFE were not. In addition to raising new questions about the relationship between mystery and NFE, these findings raise questions about the relationship between mystery (the idea that an explanation is perhaps unknowable) and ignorance (the idea that the explanation is simply unknown). By age nine, children recognize that some facts, such as the number of leaves in the world, are "unknowable," and they favor experts who acknowledge ignorance about them (Kominsky et al., 2016). There is also evidence that children and adults care that informants can accurately report their uncertainty (Bridgers et al., 2016; Tenney et al., 2007). So while declaring a question a mystery may be unacceptable in science, recognizing one's own ignorance could be a virtue. It remains an open question for future research how ignorance is treated in the domain of religion, and how and why mystery is functionally distinct.

5.1. Implications of this research

Domain differences in explanation and mystery have the potential to shed light on the psychological roles of scientific and religious cognition. Our finding that scientific beliefs are judged more in need of explanation than religious beliefs is consistent with prior claims that scientific or factual beliefs are more closely tied to evidence than are religious or "faith-like" beliefs (e.g., Buchak, 2012; Heiphetz et al., 2018; Van Leeuwen, 2014). Both sets of results could share a common explanation in terms of the non-epistemic functions of religious belief. For example, it could be that religious beliefs are less susceptible to evidence and less pressing targets of inquiry because this makes them more compelling social signals: they are more robust in the face of evidence (Boudry & Coyne, 2016; Boudry & De Smedt, 2011; Friesen et al., 2015; Van Leeuwen, 2017), and more costly and difficult to fake (Wilkins, 2018). Preliminary evidence supports the hypothesis that abdicating from evidence-seeking or explanation-seeking can be a signal of religious (but not scientific) commitment (Gill & Lombrozo, 2019), though additional work is necessary to fully articulate the non-epistemic aims of (religious) belief and to identify how epistemic and non-epistemic aims might be expressed in actual inquiry in a given domain (e.g., in seeking more information before settling on a belief or in revising belief in light of evidence).

A related question is whether scientific and religious beliefs involve different cognitive attitudes (e.g., Van Leeuwen, 2014), or merely different contents. That is, are scientific and religious beliefs both the same *kind* of belief, or do (some) religious beliefs involve a more faith-like attitude fundamentally different from (some) scientific beliefs? If there are such differences across scientific and religious beliefs—-whether in degree or in kind—can the associations between content domain and functional role reported here come apart? For instance, can people hold scientific beliefs with the functional profile suggested here for religion, or religious beliefs with the functional profile of science? Some evidence suggests that politicized scientific beliefs, such as belief in anthropogenic climate change, can be relatively dissociated from evidential considerations, and instead more closely tied to affiliation and social signaling (e.g., Kahan et al., 2012). On the flip side, it's possible that some religious beliefs are more closely linked to epistemic aims and inquiry, and that they would correspondingly show the profile found here for science.

Of course, science and religion could play additional psychological roles, and the roles identified above do not exhaust the possibilities for other domains. Recall that in Study 1, we found that the domains of health and math patterned with the domain of science, while the supernatural domain patterned with the domain of religion. However, the domains of psychology and philosophy fell between these extremes. Do laypeople's beliefs about these domains serve *both* science-like and religion-like functional roles, or are there additional functional roles for belief that have not been considered here? The fact that people generated systematic judgments about a range of domains about which they're likely to have little expertise raises the possibility that judgments of need for explanation and mystery acceptability reflect high-level "meta-beliefs" about different domains of inquiry, rather than more fine-grained beliefs tied to particular contents. If this is the case, what shapes these meta-beliefs, and how do they manifest in other judgments and behaviors?

Beyond the contrast between science and religion, our findings shed light on the nature of explanatory inquiry more generally. Recent work suggests that curiosity about the answer to a "why" question is greater when one anticipates learning new and valuable information, and that this effect holds above and beyond the extent to which the target of explanation is novel or surprising (Liquin & Lombrozo, 2020; see also Dubey & Griffiths, 2020). Thus, believing that an explanation is unavailable (e.g., because it is beyond human comprehension) could depress curiosity, and perhaps a corresponding demand for explanation, even if the target of explanation is counterintuitive or otherwise surprising. Similarly, anticipating that an answer won't supply actionable information could decrease perceptions of value, and perhaps of NFE. Given that curiosity and NFE are likely to be closely related (see Liquin & Lombrozo, 2018, for both similarities and differences), our findings also raise new questions about whether curiosity is moderated by explanatory norms and perceived subjectivity—factors that, to our knowledge, have not been explored in prior empirical work on curiosity.

Within philosophy, accounts of "need for explanation" (Grimm, 2008; Wong & Yudell, 2015) specify the conditions under which we *ought* to seek an explanation. For example, Grimm (2008) suggests that a need for explanation is prompted by the recognition of a salient alternative way that things could have been. In the absence of such a counterfactual "foil," it doesn't occur to us to ask why the fact (as opposed to the foil) came about. In Study 3, we considered the hypothesis that NFE could be lower in the domain of religion because counterfactual alternatives are less salient (e.g., we don't feel the need to explain why "God is good" because the counterfactual of God *not* being good is not salient or even considered plausible). While we failed to find support for this particular hypothesis, our finding of domain differences in NFE introduces a new constraint on theories of need for explanation. At a descriptive level, we can ask whether a single account of NFE generates different judgments across domains because of relevant ways in which those domains differ (such as the salience of counterfactuals), or whether different accounts of NFE are in fact required in distinct domains. At a normative level, we can ask when one *ought* to seek explanations given different epistemic and non-epistemic aims.

The sense that a phenomenon demands an explanation could also have implications beyond explanatory inquiry. For example, researchers have documented a "reductive allure," whereby scientific explanations are judged to be more satisfying to the extent that they contain reductive information, even if this information is explanatorily irrelevant (Hopkins et al., 2016). This may be driven by the judgment that individual components of an explanation themselves demand an explanation and must be explained at a more reductive level. Given that religious phenomena are judged less in need of explanation than scientific phenomena, the "reductive allure" may then be less pronounced in the domain of religion. On the other hand, the causal link might go the other direction; perhaps one reason we feel less need for explanation in religion is a reluctance to "reduce" religious beliefs to the kind of fine-grained material explanations that are satisfying in other contexts. More generally, we can ask how differences in NFE and MA might translate into different judgments and behaviors: even a small domain difference could have downstream consequences for how scientific and religious claims are communicated, evaluated, and used to guide action.

5.2. Limitations and further questions

Before concluding, it is important to acknowledge several limitations of these studies. First, our stimuli were almost entirely limited to beliefs consistent with Christian traditions. A focus on Christianity was convenient given our sample: while MTurk samples tend to be less religious than other samples, most religious MTurk workers are Protestant or Catholic (Berinsky et al., 2012). More importantly, however, many Christian traditions involve a notion of mystery, ineffability, or faith—some of the very notions that motivate our predictions concerning domain differences between science and religion when it comes to explanatory inquiry. Given our restriction to Christianity, it remains an open question whether NFE and MA are treated differently from science within other religious traditions, or even across different kinds of religious questions (e.g., those regarding ritual practice versus theological doctrine). We speculate that members of religious traditions that encourage question-asking and inquiry might indeed show different patterns of judgment, at least for the kinds of questions that are typically encouraged and addressed.

Given our sample, we are also unable to ask comparative questions that address the effects of cultural context. For instance, research in the United States has sometimes found conflict between scientific and religious explanations (Preston et al., 2013; Preston & Epley, 2009; but see Ecklund & Park, 2009; Scheitle, 2011), but there is evidence from other cultural contexts that religious and scientific beliefs can peacefully coexist in individual minds (Astuti & Harris, 2008; Legare & Gelman, 2008), with little or no conflict between these two domains (Davoodi et al., 2018; Payir et al., 2018). How might explicit or implicit beliefs about the relationship between science and religion moderate the effects we report? Relatedly, it is important to acknowledge that the domains of science and religion are not defined solely by the propositions in the minds of individuals; rather, science and religion are both well-established institutions with which individuals in particular cultural contexts engage in variegated ways. Our studies were not designed to assess the effects of institutional factors, nor of individual patterns of engagement with these institutions. Any differences between domains, however, are likely to be explained by a variety of complex and interacting factors, cognitive, social, cultural, and institutional.

Our studies are also limited by our reliance on explicit judgments. In some cases, explicit judgments about religion can diverge from more implicit measures (Järnefelt et al., 2015) or be complemented by behavioral measures (Preston et al., 2013). One possibility, for example, is that participant judgments of "confidence" themselves track different attitudes in science and religion (e.g., confidence in a scientific premise reflects epistemic confidence, while confidence in a religious premise reflects group allegiance). We think this possibility is unlikely to pose a threat to the interpretation of the studies here because confidence across domains was rated within-participants (and with all confidence items on the same screen in Study 3). However, the ideal measure of strength of belief might assess not only the explicitly reported confidence with which the belief is held, but also the belief's resistance to revision (perhaps even behaviorally, e.g., by presenting counterevidence and observing belief change). Relatedly, future research might seek converging evidence for our claims using different methods to assess inquiry attitudes/behaviors (e.g., by analyzing search engine data; Jansen et al., 2010) and epistemic commitments (e.g., by asking participants to explain or justify their own scientific and religious beliefs; Metz et al., 2018; Shtulman, 2013).

Finally, we examined judgments about a fairly limited sample of explanation-seeking questions. While these questions were naturalistic (Study 1) and well-matched for strength of belief in the general population (Studies 2–3), future research must sample the space of possible explanation-seeking questions more broadly. For example, do the same domain differences hold for "how" questions in addition to "why" questions, for questions seeking teleological versus mechanistic explanations, and for questions asking about properties of entities (e.g. "Why is God good?") versus the existence of entities (e.g., "Why does God exist?")? Moreover, are the effects limited to questions that seek explanations, or would they also arise for questions concerning facts (e.g., "Is the center of the earth hot?" vs. "Does hell exist?"), such that fact-seeking questions in the domain of science demand an answer more urgently than fact-seeking questions in the domain of religion?

5.3. Conclusion

Science and religion both aim to provide explanations. Yet the demand for explanations and the satisfaction of this demand appear to function differently across these two domains. Participants in our sample, even those with matched levels of confidence in the premise of a question, reported a greater need for explanation and had a lower tolerance of mystery for scientific questions than for religious ones. In science, unanswered questions were judged to stand in need of explanation, while in religion, participants were often content to leave unanswered questions as mysteries. Philosophical and psychological questions fell between science and religion. These differences are not a simple function of differences in confidence, but are instead at least partly explained by different beliefs concerning explanatory norms, epistemic limits, and perceived subjectivity. While many questions remain open, we are confident that the psychological community will continue to demand and search for explanations.

Supplementary material

All data, analysis scripts, and experimental materials are available at https://osf.io/cdy8q/.

CRediT authorship contribution statement

Emily G. Liquin:Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft, Visualization, Funding acquisition.**S. Emlen Metz**:Conceptualization, Methodology, Writing - review & editing.**Tania Lombrozo**:Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

We would like to thank members of the Concepts and Cognition Lab for their useful comments. Some of the results reported here were presented at the 2018 meeting of the Cognitive Science Society, and we are grateful to this audience for their discussion and feedback. This work was supported by a James S. McDonnell Foundation Scholar Award in Understanding Human Cognition and a Templeton Foundation Grant awarded to TL, as well as an NSF Graduate Research Fellowship to EL [grant numbers DGE-1752814 and DGE-1656466]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the McDonnell Foundation, the Templeton Foundation, or the National Science Foundation.

Declaration of competing interest

None.

Appendix A

Study 2 Items

Religion Questions

Why is there a God?
Why did God create the world?
Why is there an afterlife?
Why is there a hell?
Why does prayer work?

Science Questions

Why does burning fossil fuels create CO2?
Why do continents move?
Why is the center of the Earth so hot?
Why does the Earth go around the Sun?
Why does the Moon cause tides?

Study 3 Items

Religion Questions

Why is there a God?
Why do human beings have souls?
Why is there a life after death?
Why do miracles happen?
Why does prayer work?

Science Questions

Why does burning fossil fuels create CO2?
Why do continents move so slowly?
Why is the center of the Earth so hot?
Why does the Earth go around the Sun?
Why does the Moon cause tides?

References

Astuti, R., & Harris, P. L. (2008). Understanding mortality and the life of the ancestors in rural Madagascar. *Cognitive Science, 32*(4), 713–740. https://doi.org/10.1080/03640210802066907.

Barrett, J. L. (2000). Exploring the natural foundations of religion. *Trends in Cognitive Sciences, 4*(1), 29–34. https://doi.org/10.1016/s1364-6613(99)01419-9.

Barrett, J. L. (2004). *Why would anyone believe in God?* AltaMira Press.

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis, 20*(3), 351–368. https://doi.org/10.1093/pan/mpr057.

Boudry, M., & Coyne, J. (2016). Disbelief in belief: On the cognitive status of supernatural beliefs. *Philosophical Psychology, 29*(4), 601–615. https://doi.org/10.1080/09515089.2015.1110852.

Boudry, M., & De Smedt, J. (2011). In mysterious ways: On petitionary prayer and subtle forms of supernatural causation. *Religion, 41*(3), 449–469. https://doi.org/10.1080/0048721X.2011.600464.

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

Bridgers, S., Buchsbaum, D., Seiver, E., Griffiths, T. L., & Gopnik, A. (2016). Children's causal inferences from conflicting testimony and observations. *Developmental Psychology, 52*(1), 9–18. https://doi.org/10.1037/a0039830.

Buchak, L. (2012). Can it be rational to have faith? In J. Chandler, & V. Harrison (Vol. Eds.), *Probability in the philosophy of religion. Vol. 12. Probability in the philosophy of religion* (pp. 225–248). Oxford University Press.

Buchak, L. (2017). Reason and faith. In W. J. Abraham, & F. D. Aquino (Eds.). *The Oxford handbook of the epistemology of theology* (pp. 46–63). Oxford University Press.

Buckwalter, W., Rose, D., & Turri, J. (2015). Belief through thick and thin. *Noûs, 49*(4), 748–775. https://doi.org/10.1111/nous.12048.

Bullock, H. E., Harlow, L. L., & Mulaik, S. A. (1994). Causation issues in structural equation modeling research. *Structural Equation Modeling: A Multidisciplinary Journal, 1*(3), 253–267. https://doi.org/10.1080/10705519409539977.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). *Journal of Personality and Social Psychology, 98*(4), 550–558. https://doi.org/10.1037/a0018933.

Chater, N., & Loewenstein, G. (2016). The under-appreciated drive for sense-making. *Journal of Economic Behavior & Organization, 126*, 137–154. https://doi.org/10.1016/j.jebo.2015.10.016.

Clegg, J. M., Cui, Y. K., Harris, P. L., & Corriveau, K. H. (2019). God, germs, and evolution: Belief in unobservable religious and scientific entities in the US and China. *Integrative Psychological and Behavioral Science, 53*(1), 93–106. https://doi.org/10.1007/s12124-019-9471-0.

Davoodi, T., Jamshidi-Sianaki, M., Abedi, F., Payir, A., Cui, Y. K., Harris, P. L., & Corriveau, K. H. (2018). Beliefs about religious and scientific entities among parents and children in Iran. *Social Psychological and Personality Science, 10*(7), 847–855. https://doi.org/10.1177/1948550618806057.

Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review, 127*(3), 455–476. https://doi.org/10.1037/rev0000175.

Ecklund, E. H., & Park, J. Z. (2009). Conflict between religion and science among academic scientists? *Journal for the Scientific Study of Religion, 48*(2), 276–292. https://doi.org/10.1111/j.1468-5906.2009.01447.x.

Flavell, J. H., Mumme, D. L., Green, F. L., & Flavell, E. R. (1992). Young children's understanding of different types of beliefs. *Child Development, 63*(4), 960–977. https://doi.org/10.2307/1131247.

Friesen, J. P., Campbell, T. H., & Kay, A. C. (2015). The psychological advantage of unfalsifiability: The appeal of untestable religious and political ideologies. *Journal of Personality and Social Psychology, 108*(3), 515–529. https://doi.org/10.1037/pspp0000018.

Funk, C., & Goo, S. K. (2015). *A look at what the public knows and does not know about science.* Pew Research Centerhttps://www.pewresearch.org/science/2015/09/10/what-the-public-knows-and-does-not-know-about-science/.

Funk, C., & Kennedy, B. (2016). *The politics of climate.* Pew Research Centerhttps://www.pewresearch.org/science/2016/10/04/the-politics-of-climate/.

Gellman, J. (2017). Mysticism. In E. N. Zalta (Ed.). *The Stanford encyclopedia of philosophy (Spring 2017)*https://plato.stanford.edu/archives/spr2017/entries/mysticism/.

Gill, M., & Lombrozo, T. (2019). Social consequences of information search: Seeking evidence and explanation signals religious and scientific commitments. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.). *Proceedings of the 41st annual conference of the Cognitive Science Society* (pp. 1837–1843). Cognitive Science Society.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition, 106*(3), 1339–1366. https://doi.org/10.1016/j.cognition.2007.06.007.

Gottlieb, S., & Lombrozo, T. (2018). Can science explain the human mind? Intuitive judgments about the limits of science. *Psychological Science, 29*(1), 121–130. https://doi.org/10.1177/0956797617722609.

Green, D. P., Ha, S. E., & Bullock, J. G. (2010). Enough already about "black box" experiments: Studying mediation is more difficult than most scholars suppose. *The Annals of the American Academy of Political and Social Science, 628*(1), 200–208. https://doi.org/10.1177/0002716209351526.

Grimm, S. R. (2008). Explanatory inquiry and the need for explanation. *The British Journal for the Philosophy of Science, 59*(3), 481–497. https://doi.org/10.1093/bjps/axn021.

Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai Archives of Psychiatry, 25*(6), 390–394. https://doi.org/10.3969/j.issn.1002-0829.2013.06.009.

Harré, R. (1985). *The philosophies of science: An introductory survey* (2nd ed.). Oxford University Press.

Harris, P. L., & Giménez, M. (2005). Children's acceptance of conflicting testimony: The case of death. *Journal of Cognition and Culture, 5*(1), 143–164. https://doi.org/10.1163/1568537054068606.

Harris, P. L., Pasquini, E. S., Duke, S., Asscher, J. J., & Pons, F. (2006). Germs and angels: The role of testimony in young children's ontology. *Developmental Science, 9*(1), 76–96. https://doi.org/10.1111/j.1467-7687.2005.00465.x.

Heiphetz, L., Landers, C. L., & Van Leeuwen, N. (2018). Does think mean the same thing as believe? Linguistic insights into religious cognition. *Psychology of Religion and Spirituality*. https://doi.org/10.1037/rel0000238.

Heiphetz, L., Spelke, E. S., Harris, P. L., & Banaji, M. R. (2013). The development of reasoning about beliefs: Fact, preference, and ideology. *Journal of Experimental Social Psychology, 49*(3), 559–565. https://doi.org/10.1016/j.jesp.2012.09.005.

Heiphetz, L., Spelke, E. S., Harris, P. L., & Banaji, M. R. (2014). What do different beliefs tell us? An examination of factual, opinion-based, and religious beliefs. *Cognitive Development, 30*, 15–29. https://doi.org/10.1016/j.cogdev.2013.12.002.

Heiphetz, L., & Young, L. L. (2017). Can only one person be right? The development of objectivism and social preferences regarding widely shared and controversial moral beliefs. *Cognition, 167*, 78–90. https://doi.org/10.1016/j.cognition.2016.05.014.

Hood, R. W., Hill, P. C., & Spilka, B. (2009). *The psychology of religion: An empirical approach.* Guilford Press.

Hopkins, E. J., Weisberg, D. S., & Taylor, J. C. V. (2016). The seductive allure is a reductive allure: People prefer scientific explanations that contain logically irrelevant reductive information. *Cognition, 155*, 67–76. https://doi.org/10.1016/j.cognition.2016.06.011.

Horton, R. (1997). *Patterns of thought in Africa and the West: Essays on magic, religion and science.* Cambridge University Press.

Jacoby, J., & Sassenberg, K. (2014). *Why an interaction term in a three variable mediation model suggests that the model is problematic and how this might be solved.* (Unpublished Manuscript).

Jansen, B. J., Tapia, A., & Spink, A. (2010). Searching for salvation: An analysis of US religious searching on the World Wide Web. *Religion, 40*(1), 39–52. https://doi.org/10.1016/j.religion.2009.07.002.

Järnefelt, E., Canfield, C. F., & Kelemen, D. (2015). The divided mind of a disbeliever: Intuitive beliefs about nature as purposefully created among different groups of non-religious adults. *Cognition, 140*, 72–88. https://doi.org/10.1016/j.cognition.2015.02.005.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change, 2*(10), 732–735. https://doi.org/10.1038/NCLIMATE1547.

Keil, F. C. (2019). The challenges and benefits of mechanistic explanation in folk scientific understanding. In D. A. Wilkenfeld, & R. Samuels (Eds.). *Advances in experimental philosophy of science* (pp. 41–57). Bloomsbury Academic.

Keil, F. C., Lockhart, K. L., & Schlegel, E. (2010). A bump on a bump? Emerging intuitions concerning the relative difficulty of the sciences. *Journal of Experimental Psychology: General, 139*(1), 1–15. https://doi.org/10.1037/a0018319.

Kelemen, D. (2004). Are children "intuitive theists"? Reasoning about purpose and design in nature. *Psychological Science, 15*(5), 295–301. https://doi.org/10.1111/j.0956-7976.2004.00672.x.

Kessler, H. (1971). *The diaries of a cosmopolitan: Count Harry Kessler, 1918–1937.* Weidenfeld & Nicolson.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron, 88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010.

Klein, D., & Colombo, M. (2018). Mystery and the evidential impact of unexplainables. *Episteme, 15*(4), 463–475. https://doi.org/10.1017/epi.2017.13.

Kominsky, J. F., Langthorne, P., & Keil, F. C. (2016). The better part of not knowing: Virtuous ignorance. *Developmental Psychology, 52*(1), 31–45. https://doi.org/10.1037/dev0000065.

Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development, 15*(3), 309–328. https://doi.org/10.1016/S0885-2014(00)00030-7.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Laurin, K., & Kay, A. C. (2017). Chapter four—The motivational underpinnings of belief in god. In J. M. Olson (Vol. Ed.), *Advances in experimental social psychology. Vol. 56. Advances in experimental social psychology* (pp. 201–257). Academic Press. https://doi.org/10.1016/bs.aesp.2017.02.004.

Legare, C. H., Evans, E. M., Rosengren, K. S., & Harris, P. L. (2012). The coexistence of natural and supernatural explanations across cultures and development. *Child Development, 83*(3), 779–793. https://doi.org/10.1111/j.1467-8624.2012.01743.x.

Legare, C. H., & Gelman, S. A. (2008). Bewitchment, biology, or both: The co-existence of natural and supernatural explanatory frameworks across development. *Cognitive Science, 32*(4), 607–642. https://doi.org/10.1080/03640210802066766.

Liquin, E. G., & Lombrozo, T. (2018). Determinants and consequences of the need for explanation. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.). *Proceedings of the 40th annual conference of the Cognitive Science Society* (pp. 696–701). Cognitive Science Society.

Liquin, E. G., & Lombrozo, T. (2020). A functional approach to explanation-seeking curiosity. *Cognitive Psychology, 119*, 101276. https://doi.org/10.1016/j.cogpsych.2020.101276.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75.

Lupfer, M. B., Brock, K. F., & DePaola, S. J. (1992). The use of secular and religious attributions to explain everyday behavior. *Journal for the Scientific Study of Religion, 31*(4), 486–503. https://doi.org/10.2307/1386858.

Lupfer, M. B., Tolliver, D., & Jackson, M. (1996). Explaining life-altering occurrences: A test of the "god-of-the-gaps" hypothesis. *Journal for the Scientific Study of Religion, 35*(4), 379–391. https://doi.org/10.2307/1386413.

MacKinnon, D. (2008). *Introduction to statistical mediation analysis.* Taylor & Francis Group.

McPhetres, J., & Zuckerman, M. (2017). Religious people endorse different standards of evidence when evaluating religious versus scientific claims. *Social Psychological and Personality Science, 8*(7), 836–842. https://doi.org/10.1177/1948550617691098.

Metz, S. E., Weisberg, D. S., & Weisberg, M. (2018). Non-scientific criteria for belief sustain counter-scientific beliefs. *Cognitive Science, 42*, 1477–1503. https://doi.org/10.1111/cogs.12584.

Miller, W. (1955, May). Death of a genius. *Life Magazine, 38*, 61–64.

National Science Board (2014). *Science and engineering indicators 2014.* National Science Foundation.

Norenzayan, A. (2013). *Big gods: How religion transformed cooperation and conflict.* Princeton University Press.

Payir, A., Davoodi, T., Sianaki, M. J., Harris, P. L., & Corriveau, K. (2018). Coexisting religious and scientific beliefs among Iranian parents. *Peace and Conflict: Journal of Peace Psychology, 24*(2), 240–244. https://doi.org/10.1037/pac0000335.

Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition, 123*(3), 335–346. https://doi.org/10.1016/j.cognition.2012.03.003.

Pew Research Center (2015). *U.S. public becoming less religious*.

Preston, J. L., & Epley, N. (2009). Science and God: An automatic opposition between ultimate explanations. *Journal of Experimental Social Psychology, 45*(1), 238–241. https://doi.org/10.1016/j.jesp.2008.07.013.

Preston, J. L., Ritter, R. S., & Hepler, J. (2013). Neuroscience and the soul: Competing explanations for the human experience. *Cognition, 127*(1), 31–37. https://doi.org/10.1016/j.cognition.2012.12.003.

Ransom, M. R., & Alicke, M. D. (2012). It's a miracle: Separating the miraculous from the mundane. *Archive for the Psychology of Religion, 34*(2), 243–275. https://doi.org/10.1163/15736121-12341239.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software, 48*(2), 1–36.

Rutjens, B. T., van der Pligt, J., & van Harreveld, F. (2010). Deus or Darwin: Randomness and belief in theories about the origin of life. *Journal of Experimental Social Psychology, 46*(6), 1078–1080. https://doi.org/10.1016/j.jesp.2010.07.009.

Rutjens, B. T., Van Harreveld, F., Van der Pligt, J., Kreemers, L. M., & Noordewier, M. K. (2013). Steps, stages, and structure: Finding compensatory order in scientific theories. *Journal of Experimental Psychology: General, 142*(2), 313–318. https://doi.org/10.1037/a0028716.

Scheitle, C. P. (2011). U.S. college students' perception of religion and science: Conflict, collaboration, or independence? A research note. *Journal for the Scientific Study of Religion, 50*(1), 175–186. https://doi.org/10.1111/j.1468-5906.2010.01558.x.

Shtulman, A. (2013). Epistemic similarities between students' scientific and supernatural beliefs. *Journal of Educational Psychology, 105*(1), 199–212. https://doi.org/10.1037/a0030282.

Shtulman, A. (2015). How lay cognition constrains scientific cognition. *Philosophy Compass, 10*(11), 785–798. https://doi.org/10.1111/phc3.12260.

Shtulman, A., & Lombrozo, T. (2016). Bundles of contradiction: A coexistence view of conceptual change. In D. Barner, & A. S. Baron (Eds.). *Core knowledge and conceptual change* (pp. 53–71). Oxford University Press.

Sosis, R., & Alcorta, C. (2003). Signaling, solidarity, and the sacred: The evolution of religious behavior. *Evolutionary Anthropology: Issues, News, and Reviews, 12*(6), 264–274. https://doi.org/10.1002/evan.10120.

Sperber, D. (1996). *Explaining culture: A naturalistic approach. Vol. 323*. Oxford: Blackwell.

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18*(1), 46–50. https://doi.org/10.1111/j.1467-9280.2007.01847.x.

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109*(3), 451–471. https://doi.org/10.1037/0033-295X.109.3.451.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59*(5), 1–38. https://doi.org/10.18637/jss.v059.i05.

Tobacyk, J. J. (2004). A revised paranormal belief scale. *The International Journal of Transpersonal Studies, 23*(1), 94–98. https://doi.org/10.24972/ijts.2004.23.1.94.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1*, 31–65. https://doi.org/10.1146/annurev.clinpsy.1.102803.144239.

Tuggy, D. (2016). Trinity. In E. N. Zalta (Ed.). *The Stanford encyclopedia of philosophy (Winter 2016)*https://plato.stanford.edu/archives/win2016/entries/trinity/.

Van Leeuwen, N. (2014). Religious credence is not factual belief. *Cognition, 133*(3), 698–715. https://doi.org/10.1016/j.cognition.2014.08.015.

Van Leeuwen, N. (2017). Do religious "beliefs" respond to evidence? *Philosophical Explorations, 20*(sup1), 52–72. https://doi.org/10.1080/13869795.2017.1287294.

Vonk, J., & Pitzen, J. (2016). Religiosity and the formulation of causal attributions. *Thinking & Reasoning, 22*(2), 119–149. https://doi.org/10.1080/13546783.2015.1073623.

Weeks, M., & Lupfer, M. B. (2000). Religious attributions and proximity of influence: An investigation of direct interventions and distal explanations. *Journal for the Scientific Study of Religion, 39*(3), 348–362. https://doi.org/10.1111/0021-8294.00029.

Wilkins, J. S. (2018). Why do believers believe silly things? Costly signaling and the function of denialism. In H. van Eyghen, R. Peels, & G. van den Brink (Eds.). *New developments in the cognitive science of religion: The rationality of religious belief* (pp. 109–129). Springer International Publishing. https://doi.org/10.1007/978-3-319-90239-5_7.

Willard, A. K., & Norenzayan, A. (2013). Cognitive biases explain religious belief, paranormal belief, and belief in life's purpose. *Cognition, 129*(2), 379–391. https://doi.org/10.1016/j.cognition.2013.07.016.

Wong, W., & Yudell, Z. (2015). A normative account of the need for explanation. *Synthese, 192*(9), 2863–2885. https://doi.org/10.1007/s11229-015-0690-8.

Woolley, J. D., & Cornelius, C. A. (2017). Wondering how: Children's and adults' explanations for mundane, improbable, and extraordinary events. *Psychonomic Bulletin & Review, 24*(5), 1586–1596. https://doi.org/10.3758/s13423-016-1127-1.

Woolley, J. D., Cornelius, C. A., & Lacy, W. (2011). Developmental changes in the use of supernatural explanations for unusual events. *Journal of Cognition and Culture, 11*(3–4), 311–337. https://doi.org/10.1163/156853711X591279.

Woolley, J. D., & Dunham, J. A. (2017). Children's beliefs about miracles. *Journal of Cognition and Culture, 17*(1–2), 73–93. https://doi.org/10.1163/15685373-12342192.

Wu, J.-Y., & Kwok, O. (2012). Using SEM to analyze complex survey data: A comparison between design-based single-level and model-based multilevel approaches. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(1), 16–35. https://doi.org/10.1080/10705511.2012.634703.