A new methodological tool for research on supernatural concepts

Joseph Sommer[1,3] · Chrystal Spencer[2] · Julien Musolino[1,3] · Pernille Hemmer[1,3]

Abstract

We introduce the first set of stimuli designed to resolve methodological and theoretical issues that have muddled the interpretation of results on the memorability of supernatural concepts (e.g., ghosts, souls, spirits), an important line of research in the cognitive science of religion (Barrett, 2007). We focus here on Boyer's (1994), Boyer, 2001) pioneering minimally counterintuitive (MCI) hypothesis according to which supernatural concepts tap a special memory-enhancing mechanism linked to violations of default intuitive inferences. Empirical tests of the MCI account have given rise to a vexed picture that renders meaningful interpretation difficult. The lack of a common standard of comparison among different studies, coupled with the presence of uncontrolled variables independently known to affect memorability, lie at the heart of these problems. We show that our new stimuli offer the hope of resolving these issues, thereby establishing a more secure foundation for the study of the memorability of supernatural concepts.

Keywords Memory · Supernatural concepts · Cognitive science of religion · Minimally counterintuive concepts · Von Restorff effect

Introduction

In this article, we introduce a stimulus set and a new standard for designing stimuli intended to resolve methodological shortcomings that have muddled the interpretation of results in the literature on the memorability of supernatural concepts (e.g., ghosts, souls, spirits), itself part of a growing body of work in the cognitive science of religion (Barrett, 2007). Specifically, we present a set of stimuli (216 new items in total) developed to establish a common standard of comparison across different studies and control for the effect of key variables independently known to affect memorability. By making our new dataset publicly available, we hope to contribute to the development of more robust scientific standards, and, ultimately, to a deeper level of theoretical understanding in the area of research under consideration.

A leading hypothesis in the cognitive science of religion is that supernatural concepts are ordinary concepts that have been modified to give rise to their otherworldly qualities. More precisely, Boyer (1998, 2001); Boyer & Ramble, 2001) proposed that supernatural concepts involve violations of intuitive ontological assumptions that lead to their enhanced memorability. On this view, a ghost, for example, is drawn from the ontological category PERSON. However, unlike real persons, ghosts can pass through walls (a violation of intuitive physics), and they are immortal (a violation of intuitive biology). The crux of Boyer's hypothesis is that in order to be optimally memorable, supernatural concepts must involve only a small number of such violations (say, one or two). A small number of violations will increase memorability, compared to concepts that do not contain such violations. However, too many violations (say, three or more) will lead to decreased memorability because the resulting concepts become too cognitively complex. Thus, on this view, optimal supernatural concepts are *minimally counterintuitive* (MCI).

Initial empirical tests of the MCI hypothesis lent support to the main predictions of Boyer's account in Western adults (Barrett & Nyhof, 2001; Johnson et al., 2010), children (Banerjee et al., 2013), and non-Western populations as well (Boyer & Ramble, 2001). In these experiments, memorability for supernatural concepts is assessed relative to intuitive (INT)

✉ Joseph Sommer
  Joseph.Sommer@Rutgers.edu

1 Department of Psychology, Rutgers, The State University of New Jersey, 152 Freylinghuysen Road, Piscataway, NJ 08854-8020, USA

2 Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260, USA

3 Rutgers University, New Brunswick, NJ, USA

control concepts with equivalent numbers of characteristics. MCI concepts with a single ontological violation are compared to INT concepts with a single natural characteristic; memory for maximally counterintuitive (MXCI) concepts with three violations is compared to recall of INT concepts with three characteristics. These studies found that MCI concepts have improved memorability relative to INT concepts, and that this mnemonic advantage decreases as the number or complexity of the violations grows. However, a number of studies have also failed to replicate the original MCI effect. These studies include cases where MCI items were remembered less frequently than INT items (Gregory & Barrett, 2009; Norenzayan & Atran, 2004; Porubanova-Norquist et al., 2014), and cases where MXCI concepts were remembered at a similar rate as MCI items (Harmon-Vukić & Slone, 2009).

To begin making sense of this fractured picture, we discuss in Background section important background regarding the MCI hypothesis itself, the nature of the stimuli used in empirical tests of the hypothesis, and the potential role of key theoretical variables. We conclude that the heterogeneity of the stimuli used across different studies, coupled with core assumptions in the MCI literature itself, is likely to have led to the vexed empirical picture found in that literature. To address these shortcomings, Methods and materials section presents the results of two studies in which we generated and rated a new set of stimuli along theoretically motivated dimensions. Results and discussion section discusses the implications of our results for future studies of the memorability of (super)natural concepts. Finally, Conclusion section offers some concluding remarks.

Background

The cognitive science of religion (CSR) is a rapidly growing field of research that emerged on the cognitive science scene during the 1990s. CSR views religious thought and behavior as natural products of the human mind amenable to scientific investigation (Barrett, 2007). By bringing to bear on the study of religious cognition what is known about the structure and functioning of the human mind, researchers in CSR have been able to illuminate a range of questions heretofore not fully integrated within cognitive science. In this section, we discuss a leading hypothesis in CSR, namely Boyer's minimally counterintuitive (MCI) account of the memorability of supernatural concepts. After briefly introducing the main ideas, we point out that empirical investigations of the MCI hypothesis have given rise to a seemingly contradictory set of conclusions. We then show that this muddled picture stems from three main features of the relevant literature: (a) the use of a heterogeneous set of stimuli, (b) implicit reliance on untested assumptions, and (c) a lack of control of key variables independently known to affect memorability.

The MCI hypothesis

Pascal Boyer's (1994, 1998, 2001) MCI account proposes to explain the ubiquity of supernatural concepts across cultures (e.g., gods, souls, spirits) as a byproduct of the human concept formation capacity and memory systems. The main idea is that supernatural concepts share an underlying structure that allows them to take advantage of cognitive mechanisms that were adapted to represent and recall natural concepts. This structure can be preserved across societies even if the semantic properties of these concepts vary.

According to the MCI account, supernatural concepts possess characteristics that violate intuitive ontological theories, i.e., a set of domain-specific, near-universal, and early-developing inferences (Gelman & Wellman, 1991; Gopnik & Schulz, 2004; Pinker, 1997, 2003; Shtulman, 2017; Spelke et al., 1992; Spelke & Kinzler, 2007). For example, "*a person who can walk through walls*" violates intuitive physics, since our intuitions tell us that solid objects cannot pass through each other (Baillargeon, 1994, 1998; Carey & Spelke, 1994; Leslie, 1982). Violating these intuitive theories increases a concept's salience, making it more memorable and arguably more transmissible.

Boyer (2001) also pointed out that extant supernatural concepts seem to possess limits on the number and type of violations that they can acquire. To account for these limits, he proposed that as the number or complexity of ontological violations increases, the resulting concepts become more difficult to represent and reason about. At the same time, additional violations may achieve diminishing returns of salience. For example, "a coconut tree that would blink at least five times every minute, could disappear and reappear in a different spot in the garden, and knew everything that had ever happened in the history of the world" (Banerjee et al., 2013, p. 1275) is quite difficult to represent. Boyer thus suggests that the optimal template for a supernatural concept is *minimally* counterintuitive, with concepts that contain few violations of intuitive theories achieving greater memorability compared to those with no such violations (i.e., intuitive, natural concepts) or those with too many violations.

Since the advent of Boyer's pioneering ideas, the empirical predictions of the MCI account have been supported by findings in Western adults (Barrett & Nyhof, 2001; Johnson et al., 2010), children (Banerjee et al., 2013), and non-Western populations (Boyer & Ramble, 2001). These studies find that MCI concepts have improved memorability relative to intuitive concepts, and that this mnemonic advantage decreases as the number or complexity of violations grows. Additionally, evidence from corpus analyses have found that concepts with the MCI template are common in folktales from around the world (Burdett et al., 2009), ancient Roman prodigies (short stories about portentous events) (Lisdorf, 2004), superhero comics (Carney & Mac Carron, 2017), and across multiple versions of an urban legend (Stubbersfield & Tehrani, 2013). Still other

evidence for MCI comes from findings that suggest that while people often report belief in very complicated religious entities in accordance with their theological doctrines, when they are placed under time pressure, they deviate from these doctrines toward concepts that are more minimally counterintuitive. For example, people's conceptions of God as nonphysical, formless, and omnipresent may shift toward a view of God as an old man located in the sky when their cognitive resources are limited (Barrett, 1999; Barrett & Keil, 1996).

However, a number of studies have failed to replicate the original MCI effect. These studies include cases where MCI items were remembered less frequently than INT items (Gregory & Barrett, 2009; Norenzayan & Atran, 2004; Porubanova-Norquist et al., 2013), and cases where maximally counterintuitive items were remembered at a similar rate as MCI items (Harmon-Vukić & Slone, 2009). This muddled picture has led some investigators to conclude that "MCI theory's fate remains as unclear as its defining features" (Purzycki & Willard, 2016, p. 29). In the next section, we highlight three important factors that have likely led to this confusing empirical picture.

Concerns with the MCI literature

To begin making sense of the mixed empirical results described above, it is informative to consider the kind of stimuli that have been used in tests of the MCI account. An examination of the literature reveals a concerning amount of variation among the MCI concepts used in different studies. Stimuli range from concepts of the form noun + characteristic, as in "*A lizard that could never die no matter how old it was*" (Banerjee et al., 2013), to more elaborate descriptions such as "*A being that can see or hear things no matter where they are. For example, it could make out the letters on a page in a book hundreds of miles away and the line of sight is completely obstructed*" (Barrett & Nyhof, 2001), to items like "*closing cat*" and "*thirsty door*" (Norenzayan & Atran, 2004).

This last set of stimuli, which was used in a few studies that failed to find a memorability advantage for MCI concepts (e.g., Norenzayan & Atran, 2004), has been criticized for its potential metaphorical interpretation. For example, "*thirsty door*" may be understood as a wooden door that has dried out rather than as a concept with a salient ontological violation (i.e., an artifact with physiological needs) (Barrett, 2008a). In addition, Barrett (2004) distinguishes between counterintuitive (CI) concepts and "category mistakes," noting that:

> A category mistake involves modifying a thing with a predicate that does not and may not meaningfully apply to its ontology. For example, a "god that happened yesterday" would be a category mistake but is not counterintuitive (in the technical sense Boyer has coined). Such a notion generates no inferences and does not seem to garner any special attention or enjoy any mnemonic advantages. (p. 732)

Barrett suggests that some items used in previous studies were closer to these category mistakes than to true MCI concepts.

These stimuli have also been criticized for failing to strictly violate ontological "deep inferences" as Boyer originally intended. Instead, characteristics often violate merely "shallow" inferences and the resulting concepts are therefore better regarded as merely bizarre or even intuitive concepts rather than MCI concepts (at least as originally intended by Boyer). For instance, Purzycki and Willard (2016) write:

> Some studies designate concepts as counterintuitive that… are counterschematic or intuitive concepts. For instance, these studies consider "swimming cow," "admiring frog," and "melting lady" (or "grandfather") to be just as counterintuitive as "giggling seaweed," "arguing car," and "limping newspaper." However, cows are able swimmers, white phosphorus melts ladies and grandfathers, and picturing frogs admiring each other is cognitively effortless. Which intuitive processes do these items violate? (p. 20)

In sum, the highly variable nature of the stimuli used in empirical tests of the MCI account complicates any straightforward interpretation of the overall pattern of results emerging from these different studies.

The heterogeneous stimuli used in these studies might have arisen from two implicit assumptions that have guided work in the MCI literature without any serious vetting. The first is the assumption that as far as memorability is concerned, all MCI items are created equal (i.e., have the same enhanced memorability profile). The second is that violations of intuitive ontologies produce a unique kind of memorability that overwhelms the effect of other factors which might contribute to recall and thus that any MCI item will always be better remembered than any bizarre (BIZ) or INT item (see Sommer et al., 2022, for further discussion).[1] In part because of the implicit adoption of these assumptions, studies that have failed to find a comparative memorability advantage for MCI items have been viewed as failures to replicate the expected effect. However, if these assumptions are incorrect, as there are good reasons to believe, the apparent failures to replicate the MCI effect found in the literature may have a different explanation. If all MCI items are not created equal in terms of memorability, and if one accounts for the effect of other (uncontrolled) variables known to affect memorability, there may be no reason to expect a strict hierarchy in which MCI items are always remembered better than INT or BIZ items. A similar position is advocated by Bendixen and Purzycki

---

[1] INT concepts possess characteristics that are mundane properties of entities in the real world. For example, an INT dog concept might possess the characteristic of barking loudly. Bizarre (BIZ) concepts have characteristics that are unusual, but which do not violate an intuitive ontological theory. For example, a BIZ dog concept might be a dog that has been sprayed painted bright pink.

(2021), who suggest that the memorability of MCI concepts may be due to a confluence of factors, including ontological violations, inferential potential (see below), and social learning heuristics, which are expected to vary across contexts. The operative question is not which single variable is responsible for the entire MCI effect, but how and under what conditions a confluence of factors jointly determines the cultural success of MCI concepts.

This brings us to an additional concern about the MCI literature regarding the potential memory effect of variables other than violations of intuitive ontologies. A key phenomenon of interest in this regard with a long history within the MCI literature itself is the Von Restorff effect (VR) (1933). VR describes improved memorability for items in a list which are "isolated" or outliers. For example, in a list of fish, a mammal, such as a lion, would stand out and be disproportionately remembered. The resemblance between the salience of a surprising VR item and that of an MCI item's surprising characteristics was noted early in the MCI literature (Barrett & Nyhof, 2001; Boyer & Ramble, 2001).[2]

Due to the obvious similarity between the VR and MCI effects, a number of studies attempted to tease them apart

(Atran & Norenzayan, 2005; Barrett & Nyhof, 2001; Boyer & Ramble, 2001; Gregory & Greenway, 2017). However, experiments attempting to find empirical differences in memorability between MCI items and merely bizarre (BIZ) (i.e., VR) items have led to mixed results. When concepts are rated for unusualness, a proxy for bizarreness, these ratings are sometimes correlated with recall, but on other occasions, they are not (Barrett & Nyhof, 2001; Boyer & Ramble, 2001). Thus, though the connection between VR and MCI was noted early on, the two effects have yet to be conclusively disentangled. As discussed above, MCI studies are still criticized for inadvertently creating shallow VR violations instead of the intended deeper ontological violations that Boyer had in mind for MCI concepts (Purzycki & Willard, 2016).

In addition to the VR effect and other factors such as imageability (Paivio, 1986), which have been controlled for in only a handful of MCI studies (e.g., Gonce et al., 2006), there is another long known but rarely measured variable that is of significance for MCI concepts. The notion of "inferential potential" (IP) has been discussed since the field's inception (Boyer, 1996) and was believed to be an important component of the differential success of MCI concepts. However, despite early recognition, IP has rarely been rigorously defined, operationalized, or controlled for. On the rare occasions where IP has been manipulated or controlled for, results have suggested that it may play an important role in the memorability of MCI concepts. One such study used ratings of thought-provokingness and imageability as a proxy for IP (Gregory & Barrett, 2009). Interestingly, these authors found a correlation between IP and recall even within a rather narrow sliver of the IP spectrum, namely concepts with median scores between 2.5 and 3.5 on a five-point scale. In another study of IP, Beebe and Duffy (2020) hypothesized that characteristics with moral valence or existential anxieties, such as death, deception, or disease, might contribute to the memorability of MCI concepts. They found that morally relevant characteristics, such as a person knowing about the moral transgressions of others, and descriptions that provoked existential anxieties, such as a story about a near-death experience, both achieved higher recall than control stimuli. Moreover, these effects of moral valence and existential anxiety were stronger than that of MCI structure (studies 2 and 3).

One reason that IP may have failed to generate more rigorous scrutiny is that the concept has been interpreted in two different ways in the MCI literature. In Boyer's early writings (1996), IP was regarded as a feature of preserved inferences from unviolated intuitive ontological theories. When an intuitive theory is violated, it was thought to be "blocked" in the sense that inferences about a concept could no longer be drawn from that theory. For example, if a concept gains the ability to *walk through walls*, intuitive physics has been

---

[2] Note that the interpretation of the VR effect as a result of salience has been challenged. VR effects are found when the isolated item is presented first on the list, when there is no prior context for it to stand out against (Kelley & Nairne, 2001). Additionally, the VR effect persists when the item is not noticeably isolated at encoding but only differentiated at retrieval. Chee and Goh (2018) presented the homophone "kiwi" in a list of either birds or fruit and gave participants a hint that one of the items they had studied was a member of the alternate category (e.g., participants who studied a list of birds were informed that one of the items had been a fruit). They found a VR effect when this hint was withheld until retrieval, suggesting that the effect is not due to differential encoding of salient items. On this interpretation of the VR effect, the memorability advantage gained by counterintuitive characteristics would not be a result of increased processing of surprising properties, but because these characteristics serve as more effective retrieval cues. This effectiveness might be due to retrieval relying on a *relative match* with encoded material, rather than an *absolute match* (Nairne, 2002). In other words, the most effective retrieval cues may be those that uniquely specify a single item for recall. For example, trying to recall a specific memory of a tree may be difficult as there are many competitor memories of other trees that are very similar to the target memory. A supernatural tree that grants wishes may be better recalled because its counterintuitive property disambiguates it from other trees in memory.

However, note as well that in the context of MCI concepts, Swan and Halberstadt (2021) found that inducing anxiety at encoding, but not at retrieval, led to a greater bias toward remembering MCI concepts. This discrepancy may be evidence for a nuanced view of the MCI effect that implicates multiple processes in the memorability of MCI concepts. Part of the improved memorability of MCI concepts may be a VR or bizarreness effect at retrieval, while factors like moral content or emotional state may make contributions during encoding (cf. Sommer, Musolino, & Hemmer, preprint; Bendixen & Purzycki, 2021). Future research is needed to ascertain which processes are involved in memory for MCI concepts as well as the time course of their operation.

blocked and can no longer be relied upon for easy inferences about the concept. However, as long as other intuitive theories like intuitive biology or psychology remain unviolated and thus operative, the concept retains some inferential potential. This "preserved" view sees this remaining inferential potential as the relevant factor for MCI concepts.

A second sense of the term stems from Boyer's (2001) observation that some concepts are better candidates for supernatural items than others. For example, "*a person who can read minds*" seems somehow better than "*a statue that vanishes whenever someone thinks about it*" (p. 38). This alternative understanding of IP has been loosely defined as a concept's capacity to "readily generate inferences, explanations, and predictions with little effort" (Burdett et al., 2009) or "the potential a particular concept has to generate thoughts, predictions, memories, mental imagery and other personal inferences in the mind representing it" (Gregory & Barrett, 2009). In contrast with the former view, this notion of IP is construed as a semantic feature of the created concept, and not a property of the intuitive ontologies. On this "tangential" view, IP is orthogonal to CI characteristics and represents an additional dimension whereby certain concepts might be better candidates for recall and/or transmission.

These two interpretations of IP are compatible with the assumption that violations of intuitive ontologies are the most important factor in the MCI effect. Preserved IP is directly related to violations, in that IP drops as violations increase and block additional intuitive theories, which means all MCI concepts with the same number of violations should draw on the same amount of IP from their preserved intuitive theories. Tangential IP is unrelated to violations; however, because it is tangential, it should vary randomly across items and have no relation to whether an item is MCI or INT. On both views then, IP should not be systematically different between categories of concepts, such as MCI, BIZ, or INT items. Thus, any effect of IP on memorability should be swamped by the effect of CI violations, which do give one category of concepts, namely MCI concepts, an advantage.

There is, however, a third possible view of IP which would give IP a much larger role in the memorability of MCI concepts. It assumes IP is systematically related to violations, like the preserved account, but like the tangential account, it locates IP in the semantics of those violations. This account of IP, which we call the "created" view, suggests that generation of CI violations is likely to (but critically does not have to) result in high-IP concepts. Intuitively, supernatural characteristics often produce concepts with high IP. *A person who can read minds* or *can fly* or *can turn items they touch into gold* is powerful and interesting in ways matched by few intuitive concepts. However, supernatural characteristics are no guarantee of high IP, as in Boyer's "*dishwasher that gives birth to offspring but*

they are telephones, not little dishwashers. (p. 62)." Thus, if the created view of IP is correct, MCI concepts should often possess higher IP than other concepts, but low-IP MCI concepts, which may have particularly poor memorability, remain possible.

Moving forward

As the preceding discussion indicates, controlling for the effects of IP and other memory-related variables in the creation and comparison of INT, BIZ, MCI, and MXCI concepts would have a number of desirable consequences. First, a more tightly controlled stimuli set would allow us to better understand the relationship between the MCI and VR effects. It might explain the mixed results found in comparisons of MCI and BIZ items, including the occasional correlation between recall and unusualness ratings. If other dimensions like IP are important for memorability, these findings might have resulted from variation in IP. As, to our knowledge, no study has simultaneously assessed the degree of unusualness of MCI items along with IP, it remains possible that the MCI effect merely is the VR effect with additional "created" IP. (Incidentally, if it is the case that the VR and MCI effects can be reduced to the same mechanism, this will also address more recent theoretical criticisms, mentioned above, that argue that some MCI stimuli are only BIZ).

Second, this approach may allow a more nuanced understanding of where the MCI effect applies and further restrict the domain of concepts that are likely to become successful supernatural entities to those with high IP. Intriguingly, it may also broaden the scope of the MCI effect to concepts that are not supernatural. As we have argued elsewhere (Sommer et al., 2022), if the MCI effect can indeed be reduced to a VR effect of salience due to surprising characteristics + high IP, this might explain the cultural success of many non-supernatural entities that nonetheless possess salient bizarre characteristics + high IP, such as extreme intelligence, strength, or physical capabilities.

Third, the mixed empirical support for the MCI effect might be explained as a function of different semantic features of the stimuli. For example, studies which used stimuli like "*closing cat,*" which Barrett (2004) dubbed category mistakes, and which failed to support the MCI effect (e.g., Norenzayan & Atran, 2004) may have resulted from these concepts possessing extremely low IP.

In fact, a similar case might also be made for some of Boyer's early discussion of concepts that make poor candidates for religious items. Though it is now common to think of so-called maximally counterintuitive (MXCI) items as those which have a greater number of ontological violations (Barrett, 2008a), this was not always the case. Boyer (1996), Boyer, 2001; Boyer & Ramble, 2001) originally had

a different conception of MXCI concepts which involved especially complex violations. Thus, when Boyer and Ramble (2001) assess recall of complex counterintuitive concepts, they use concepts which possess a just a single complex violation with two parts.[3] Take the example mentioned above of a "*dishwasher that gives birth to offspring but they are telephones, not little dishwashers*" or "*an amulet that can hear what people will say in the future*" (both items from Boyer, 2001, p. 62). Each of these takes an ontological violation, giving birth to a different "kind" and hearing things in the future, but then additionally applies it to an ontological category to which it is not naturally suited, i.e., appliances do not give birth, nor can jewelry hear. Here too, results finding poor memorability may be due at least in part to what appears to be low IP.

In short, we believe that much of the confusion surrounding the MCI effect can be reduced by the creation of a standardized set of stimuli that will establish a common standard of comparison and also allow researchers to control for the effect of variables independently known to affect memorability. We now turn to a description and analysis of a new set of stimuli designed with these important goals in mind.

Method and materials

Stimuli generation

Stimuli were designed in accordance with Barrett's (2008a) recommendations for constructing CI stimuli. The stimulus set contains intuitive (INT), counterintuitive (CI), and bizarre (BIZ) items. The latter possess characteristics that are unusual, but which do not violate intuitive ontological theories. The inclusion of both BIZ and CI items will allow us to determine whether differences on dimensions relevant to IP can tease apart the VR and MCI effects, as the "created" account of IP predicts.

We created new concepts of the form *N + C*, where *N* is a noun (e.g., lizard) and *C* is a characteristic of that noun

Table 1 Stimulus nouns

Ontological category (Barrett, 2008a, 2008b)	Nouns
Spatial entities	Cloud
	Puddle
Solid object (man-made)	Lamp
	Rocking chair
Solid object (natural)	Stone
	Icicle
Living things	Tree
	Cactus
Animates	Rabbit
	Bird
Persons	Person
	Child

that takes the form of a relative clause (e.g., that has rough skin). Each noun was modified with one, two, or three characteristics that are either CI, BIZ, or INT, yielding items of the form $N + C_1$; $N + C_1 + C_2$; and $N + C_1 + C_2 + C_3$. For example, a noun might yield the following three (CI) concepts: *an icicle that knows the future; an icicle that knows the future and has a mouth that can speak; an icicle that knows the future and has a mouth that can speak and tells fortunes if you bring it sand from the desert.*

Nouns were drawn from Barrett's (2008a) set of ontological categories: spatial entities, solid objects (here divided into natural and artificial), living things, animates, and persons (see Table 1 for the full list of nouns used). Two nouns were selected from each category, yielding a total of 12 nouns in the full set. Every noun is represented in all three categories (CI, BIZ, and INT) to avoid different nouns influencing ratings across categories.

Additionally, concepts were designed to be either high or low in inferential potential, with the goal of validating this difference in the rating studies, discussed below. Each noun received two sets of characteristics, one for the low IP level and one for high IP. In all, the set comprises 12 nouns (2 from each ontological category) × 3 categories (CI, BIZ, INT) × 3 characteristic numbers (1, 2, 3) × 2 IP levels (high, low), yielding 216 concepts.

Counterintuitive characteristics were designed to violate the standard three intuitive ontological categories of physics, biology, and psychology used in most studies of the MCI effect. In keeping with the general practice in the field, when a concept possessed multiple characteristics, these characteristics were usually drawn from different ontological categories (Barrett, 2008a). However, because we suspect that this constraint may artificially induce mnemonic difficulty (see Sommer et al., 2022), some concepts were allowed to possess multiple violations from the same intuitive theory. For BIZ items, there are no comparable set of overarching

---

[3] In the MCI literature, a *breach* refers to a characteristic that no extant creature or object can possess, e.g., the ability to see the future. A *transfer,* on the other hand, applies a characteristic that is a feature of a natural concept to a one that cannot possess that property under normal circumstances, e.g., a tree that can speak. Boyer and Ramble combined these two types of violations to create a *breached transfer,* which transfers a breached concept to an item that does not normally have the original characteristic. Boyer notes that these combinations are particularly rare in the anthropological record. For example, many Catholics believe that certain statues can hear prayers and that God can hear people anywhere in the world. Here, the ability to hear has been transferred to the statue, while God has a breached ability to hear from any distance. However, these violations are not combined to create a statue that listens to prayers from anywhere in the world (Boyer & Ramble, 2001).

categories for weird characteristics. To guide the construction of these concepts, the following categories were "violated": color, orientation, location in a scene, value, prestige, emotional valence, and behavior.

Rating study 1

The stimuli were rated in two studies on dimensions related to IP, as well as on unusualness. Study 1 rated the stimuli on the dimensions of usefulness, imageability, thought-provokingness, and unusualness. Usefulness was selected based on the intuition that IP might be related to the literature on adaptive memory (e.g., Nairne et al., 2013; Nairne et al., 2017; VanArsdall et al., 2015) which might understand IP to be a mnemonic advantage for items with evolutionary utility. The dimensions of imageability and thought-provokingness were based on Gregory and Barrett's (2009) method of controlling for IP. In their study, which sought to compare MCI items with epistemically incongruous concepts (e.g., a circular triangle), the proxy for inferential potential used was a combination of the extent to which a concept was thought-provoking and how easily it brought mental images to mind. Additionally, imagery makes well-known contributions to memory (e.g., Paivio, 1986). Finally, items were rated for unusualness to assess differences on this dimension within items in the CI condition as well as to compare CI items to their BIZ counterparts.

Participants

All experiments were conducted in accordance with protocols approved by the Rutgers University Institutional Review Board (IRB). Participants were 181 adult Amazon Mechanical Turk (M-TURK) workers (75 female, 3 unspecified), ranging in age from 19 to 71 years, with a mean age of 33.6 years (SD = 10.4). All participants were US residents. Participants were recruited from the 90th percentile of M-Turk workers and were paid at a rate of $2.50 per 15 minutes of participation. We excluded from analysis participants who only used two or fewer values on the five-point Likert scale (this usually meant that the participant chose the same value, e.g., 4, for every rating). After exclusion, the final sample contained 157 participants. None of the analyses presented below changed substantively after removing data from the sample (see the Supplementary Material for this article for more information).

Procedure

Using the Qualtrics survey and questionnaire platform, participants first completed a brief demographic survey and were then presented with 27 items drawn at random from the set. Randomization was automatically balanced

by Qualtrics to present each stimulus item approximately equally across participants. In total, each of the 216 stimuli was rated between 19 and 24 times, with a mean of 21.5 ratings per item. Participants viewed one item at a time and were asked to provide ratings on a five-point Likert scale (1 = low, 5 = high) on each of the four dimensions for each concept that were presented to them. Ratings were elicited with the following questions: (1) "How useful is the item?" (2) "How easily do mental images about the item come to mind?" (3) "How thought-provoking is the item?" (4) "How unusual is the item?"

Rating study 2

Study 2 repeated the procedure from study 1 with the same set of stimuli. A second group of M-Turk workers were recruited and rated the stimuli on a second set of dimensions. The dimensions used in study 2 were adapted from Barrett's (2008b) suggestions for properties that successful religious concepts tend to possess in addition to an MCI structure. These properties are intentional agency, strategic knowledge, the ability to act in the world, and the propensity for reinforcing motivating behaviors or rituals.

Many of Barrett's properties seem closely related to agency. Recently, we proposed (Sommer et al., 2022) that agency may be a component of IP and might be an important factor in the memorability of MCI items. Agency has been found to improve memorability (Nairne et al., 2013; Nairne et al., 2017; VanArsdall et al., 2015) and this dimension might also vary more in MCI concepts than in natural concepts. MCI concepts can easily achieve improved agency, such as *a dog that can talk*, or have their agency supernaturally reduced, such as a *person "who only sees what does not happen behind them"* (Boyer, 2001, p. 62). Indeed, a few studies have found agents to be particularly memorable in MCI experiments (Porubanova et al., 2014; Porubanova-Norquist et al., 2013), however, this dimension is still largely overlooked in the literature.

Participants

Participants were 172 adult Amazon M-Turk workers (59 female, 1 unspecified), ranging in age from 19 to 69 years, with a mean age of 34.2 years (SD = 9.8). All participants were US residents. Participants were again recruited from the 90th percentile of M-Turk workers and received $2.50 per 15 minutes of participation. As in study 1, participants who repeatedly assigned ratings to a single value were excluded from analysis. After exclusion, 101 participants remained in study 2's sample. Analyses did not differ substantively before and after exclusion. All data from both experiments and the stimulus set are available at https://osf.io/4xsc8/.

Procedure

As in the previous study, participants were presented with 27 concepts randomly drawn from the stimulus set and asked to rate each item on all four dimensions. The following questions were used to operationalize Barrett's dimensions and were rated on a five-point scale (1 = low, 5 = high): (1) "To what extent does this item have goals and desires?" (2) "To what extent does this item have the ability to learn information that will help it achieve goals?" (3) "To what extent would this item perform actions that would be noticed by others?" (4) "To what extent would people be likely to try to win the support of this item?"

Results and discussion

The stimuli and ratings are particularly useful to future research in the field if they can assist in operationalizing the vague notion of IP and if they demonstrate that concepts do indeed vary along the rated dimensions in ways that might have influenced previous studies. The rating data suggests that these criteria were met by our stimuli. In analyzing the data, the primary questions were whether and how these dimensions vary across the categories of CI and INT concepts. These questions are critical to understanding which factors may influence memorability for MCI items, which dimensions may be components of IP, and whether heterogeneous findings in the field might be explained by stimuli that vary along these dimensions. For example, it might be the case that INT and CI items are nearly identical on all dimensions ostensibly related to IP and only differ on unusualness. This might be the pattern most consistent with the traditional focus on ontological violations and its concomitant neglect of other factors. However, if INT and CI items differ on other dimensions as well as unusualness, these differences might explain results where INT concepts are better recalled than CI concepts and these dimensions could be the targets of future research.

A second question was whether the rating data might permit more nuanced comparisons of CI and BIZ items and a deeper understanding of some of the ambiguous research comparing these items in the past. Again, one might imagine that the only difference between CI and BIZ items appears on the dimension of unusualness and even there, the difference between violation types might be one of kind, rather than of degree, with BIZ and CI items presenting as equally unusual but in different ways. On this view, CI and BIZ items would look identical on all dimensions discussed above. On the other hand, these two types of items might differ from each other on several dimensions. This possibility is intriguing because, as in the question of variability between CI and INT concepts,

if BIZ and CI concepts differ, this might explain why some studies find that MCI items outperform BIZ items, while others find the opposite pattern (Barrett & Nyhof, 2001).

Third, data were analyzed to ascertain whether and how concepts vary on the rated dimensions within each category (i.e., CI, BIZ, and INT). These questions have direct bearing on whether failed replications are likely to be due to variation on factors like IP. While the stimuli used in the present study do not vary as widely as some used in the literature, if concepts do not vary within the CI category, it is difficult to argue that IP plays an important role in the differential memorability of CI items across studies.

Finally, in light of previous findings about the benefits of agentic MCI concepts (Porubanova et al., 2014; Porubanova-Norquist et al., 2013) as well as arguments that MCI concepts are likely to have an advantage in gaining agency over other types of concepts and that this advantage might be an important component of IP (Sommer et al., 2022), ratings were analyzed to find out whether CI items disproportionately gain agency.

Which dimensions matter for IP?

Figure 1 displays the mean ratings for INT and CI concepts, collapsed over high/low IP and number of characteristics.[4] Even at this coarse level of analysis, results indicate that there may be large differences between CI and INT concepts on multiple dimensions which may influence memorability.

Usefulness, at least abstracting over IP level (see "Rating study 2" section below), is the only dimension that shows little difference between CI and INT items ($M_{CI} = 2.60$; $M_{INT} = 2.59$). A two-sample t-test fails to find a significant difference for usefulness when collapsed over IP level, $t(2829) = 0.28$, $p = 0.78$. For imageability, INT items were rated higher than CI items ($M_{CI} = 2.86$; $M_{INT} = 4.16$), replicating the findings of Gonce, et al. (2007), who found lower imagery ratings for CI items, though they also found

---

[4] For almost all concepts, regardless of category, number of characteristics did not have a large effect and is not discussed further. Note that this is highly unlikely to be the case for memorability, where number of characteristics would be expected to have an effect. According to MCI theory, MXCI concepts are difficult to represent and recall due to their multiple violations of intuitive ontological theories. We have suggested that lower recall for MXCI concepts might be explained by a lower "coherability" of their supernatural characteristics (Sommer, Musolino, & Hemmer, preprint). INT concepts tend to have coherable characteristics, e.g., "a bee that is yellow and black, likes flowers, and can sting," whereas supernatural characteristics may have nothing in common with each other or with the noun they are appended to, e.g., "a bee that breathes fire, can fly through walls, and can read minds." On both of these theories, the reduced memorability of MXCI concepts is not caused by properties like IP, which makes the lack of an effect of characteristic number somewhat unsurprising for IP ratings.
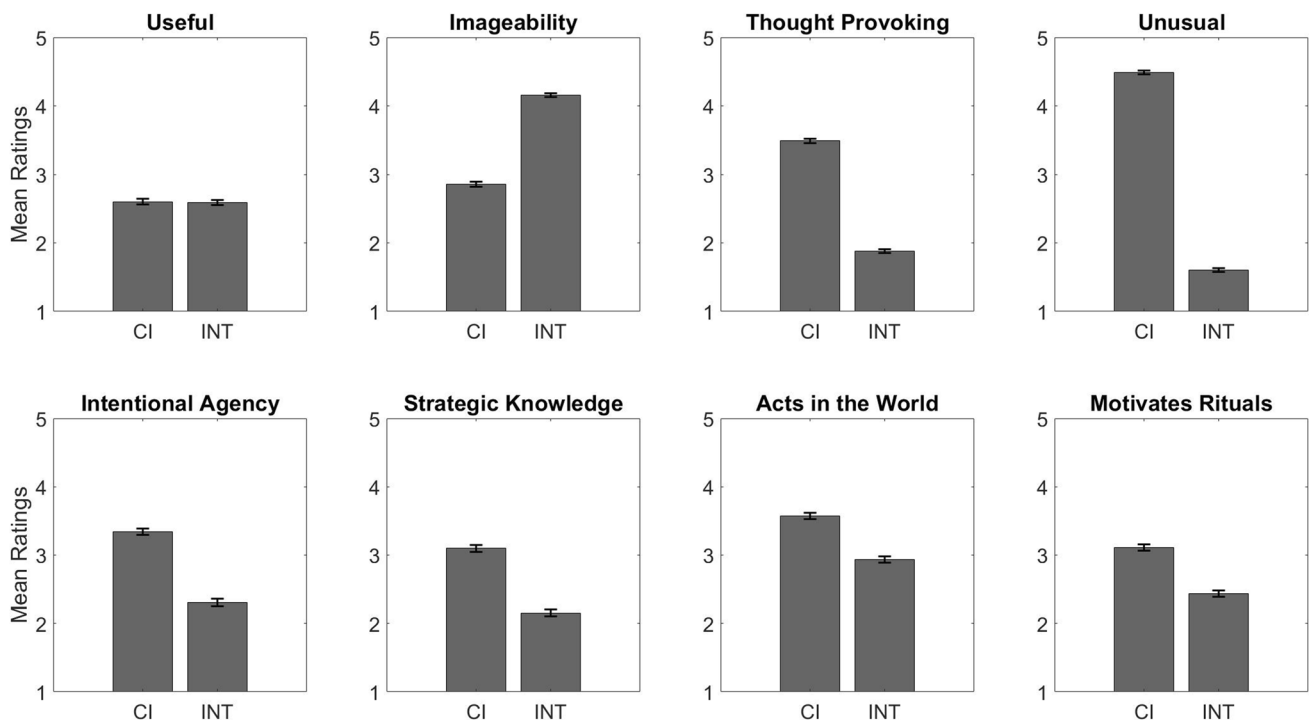
Fig. 1 Mean ratings for CI and INT stim items for all dimensions rated in studies 1 and 2. Means are collapsed over IP level and number of violations. Error bars represent the standard error of the mean

that this did not reduce memorability for MCI items. This difference is significant, $t(2829) = -29.02$, $p < 0.0001$.[5] CI concepts were rated as significantly more thought-provoking ($M_{CI} = 3.49$; $M_{INT} = 1.88$), $t(2829) = 36.49$, $p < 0.0001$; and unusual ($M_{CI} = 4.49$; $M_{INT} = 1.60$), $t(2829) = 75.34$, $p < 0.0001$ than INT concepts.

Interestingly, results from study 2 on Barrett's (2008a) characteristics of successful religious concepts show that the CI items, at least in the target stimuli, seem to have an advantage over INT items on all of these dimensions, which include intentional agency, ($M_{CI} = 3.34$; $M_{INT} = 2.31$), $t(1837) = 14.75$, $p < 0.0001$; possession of strategic knowledge ($M_{CI} = 3.1$; $M_{INT} = 2.15$), $t(1837) = 13.79$, $p < 0.0001$; the ability to act in the world ($M_{CI} = 3.57$; $M_{INT} = 2.93$), $t(1837) = 9.9$, $p < 0.0001$; and the tendency to promote reinforcing behaviors or rituals ($M_{CI} = 3.11$; $M_{INT} = 2.43$), $t(1837) = 10.08$, $p < 0.0001$. See Agency section, below, for a more detailed discussion of these findings.

It appears that CI and INT concepts systematically differ on multiple dimensions which might affect their relative memorability. These dimensions include those suggested to be components of IP, such as how thought-provoking the concepts are, those suggested to be components of successful religious concepts, such as the possession of strategic knowledge, and unsurprisingly, unusualness.

CI versus BIZ

If CI and INT items systematically differ on these dimensions, this raises questions about whether and how BIZ concepts differ from both INT and CI concepts. If BIZ concepts bear little resemblance to CI concepts, naïve comparisons of BIZ and CI items might be influenced by factors other than the type of unusualness they possess.

Figure 2 adds mean ratings for BIZ items to the graphs from Fig. 1. Perhaps the most obvious but important finding is that BIZ concepts are less unusual than CI concepts ($M_{CI} = 4.49$; $M_{BIZ} = 3.68$), $t(2821) = -18.26$, $p < 0.0001$. This might imply that the difference between CI violations and BIZ characteristics is one of degree, rather than of kind, which may diminish the importance of violations of intuitive theories. It also argues against comparing CI and BIZ items without at least equating for unusualness. Apart from this fact, however, BIZ and CI items differ on every other dimension, as well. In addition to being rated as less unusual
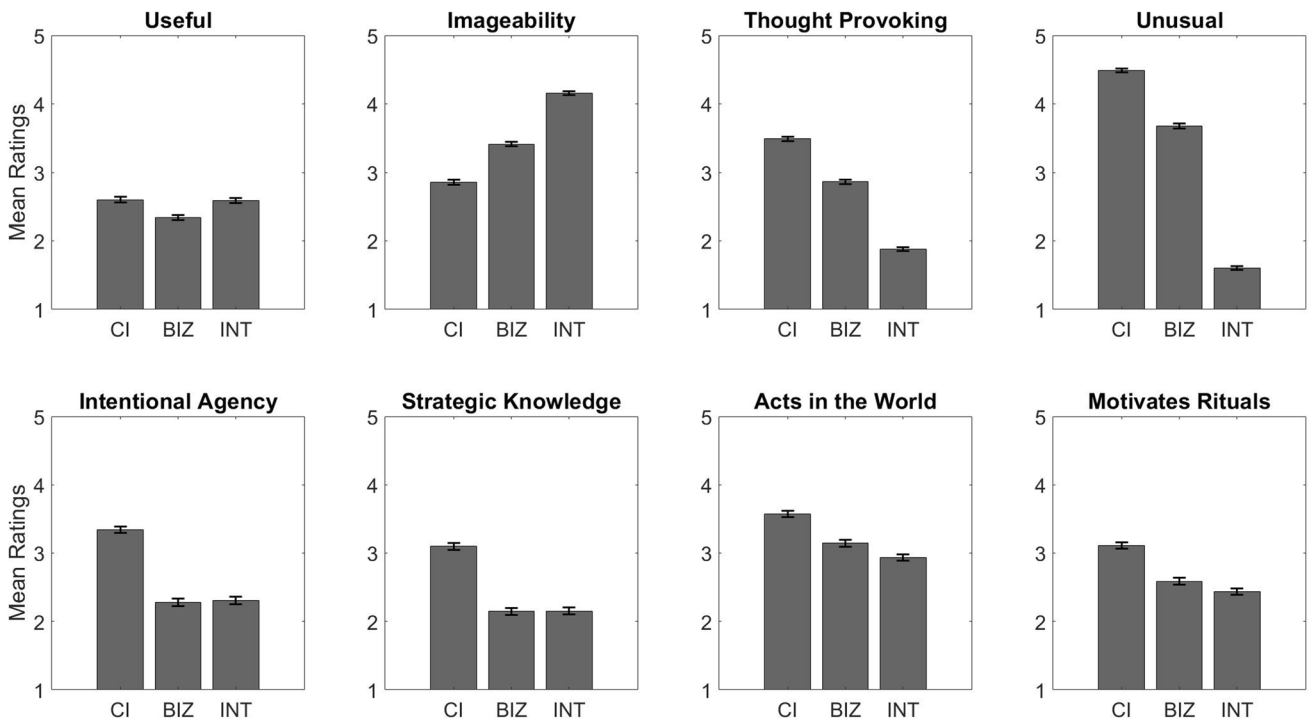
---

[5] It should be noted that, collapsed over the many stimuli concepts as well as number of characteristics and IP level, each condition contains between ~900 and ~1400 ratings, which lowers the p-values in these comparisons. At this level of analysis, the experiments are extremely high powered and will detect even small differences between concept types. Power analyses for the sample sizes in the present study (collapsed over number of violations and IP level) indicate that differences ranging from 0.12 to 0.18 are detectable with 90% power.

Fig. 2 Mean ratings for CI, BIZ, and INT stim items for all dimensions rated in studies 1 and 2. Means are collapsed over IP level and number of violations. Error bars represent the standard error of the mean

than CI concepts, BIZ concepts were rated as less thought-provoking ($M_{CI} = 3.49$; $M_{BIZ} = 2.86$), $t(2821) = -13.25$, $p < 0.0001$; more imageable ($M_{CI} = 2.86$; $M_{BIZ} = 3.4$), $t(2821) = 11.64$, $p < 0.0001$; and less useful ($M_{CI} = 2.60$; $M_{BIZ} = 2.34$), $t(2821) = -4.84$, $p < 0.0001$.

Additionally, BIZ concepts were rated lower than CI items on intentional agency, ($M_{CI} = 3.34$; $M_{BIZ} = 2.28$), $t(1780) = -14.78$, $p < 0.0001$; possession of strategic knowledge ($M_{CI} = 3.1$; $M_{BIZ} = 2.15$), $t(1780) = -13.42$, $p < 0.0001$; the ability to act in the world ($M_{CI} = 3.57$; $M_{BIZ} = 3.15$), $t(1780) = -6.27$, $p < 0.0001$; and motivation of reinforcing behaviors or rituals ($M_{CI} = 3.11$; $M_{BIZ} = 2.59$), $t(1780) = -7.6$, $p < 0.0001$.

Variability

Beyond mean differences, there is also the question of how variable concepts are on the dimensions under consideration. If all CI concepts are roughly the same, then beyond group-level comparisons, such as between CI and BIZ items, there is no need to carefully select stimuli within a category. However, if concepts can vary within a category, an unlucky set of stimuli may lead to results that appear to contradict the MCI effect's predictions.

Figures 3 and 4 plot histograms of the proportions of ratings for each dimension by category (again, abstracting over IP level and number of characteristics). For most dimensions

and for most concept types, there is a wide range of ratings which suggests that items sampled from these distributions should be expected to vary. For example, in a set of CI items, thought provokingness ratings might range from 1–5. This variability could lead to a set of stimuli that all have quite low thought provokingness and which might, as a result, underperform expectations for CI concepts or even be remembered less than INT concepts. Though these results are highly suggestive and, in many cases, match intuitive pre-experimental expectations (e.g., CI concepts are more unusual than INT concepts), it is possible that they are driven by outliers. In the Supplementary Material for this article, we present a more detailed concept-level analysis that rules out this possibility and supports the present interpretation of Figs. 3 and 4.

While Figs. 3 and 4 indicate that items of all types receive a significant proportion of ratings spanning the full scale on most dimensions, more detailed breakdowns show that this variability is not random. Figures 5, 6 and 7 split the concepts by whether they were designed to be high or low in IP. Differences in ratings based on IP level suggest that there is underlying structure to the variation among concepts. These figures plot the frequency distributions of means for each concept and compare concepts that were designed (based on intuition) to be high IP with those intended to be low IP. For example, in Fig. 5, unusualness ratings for CI concepts are clustered on the right side of the graph. Just about every
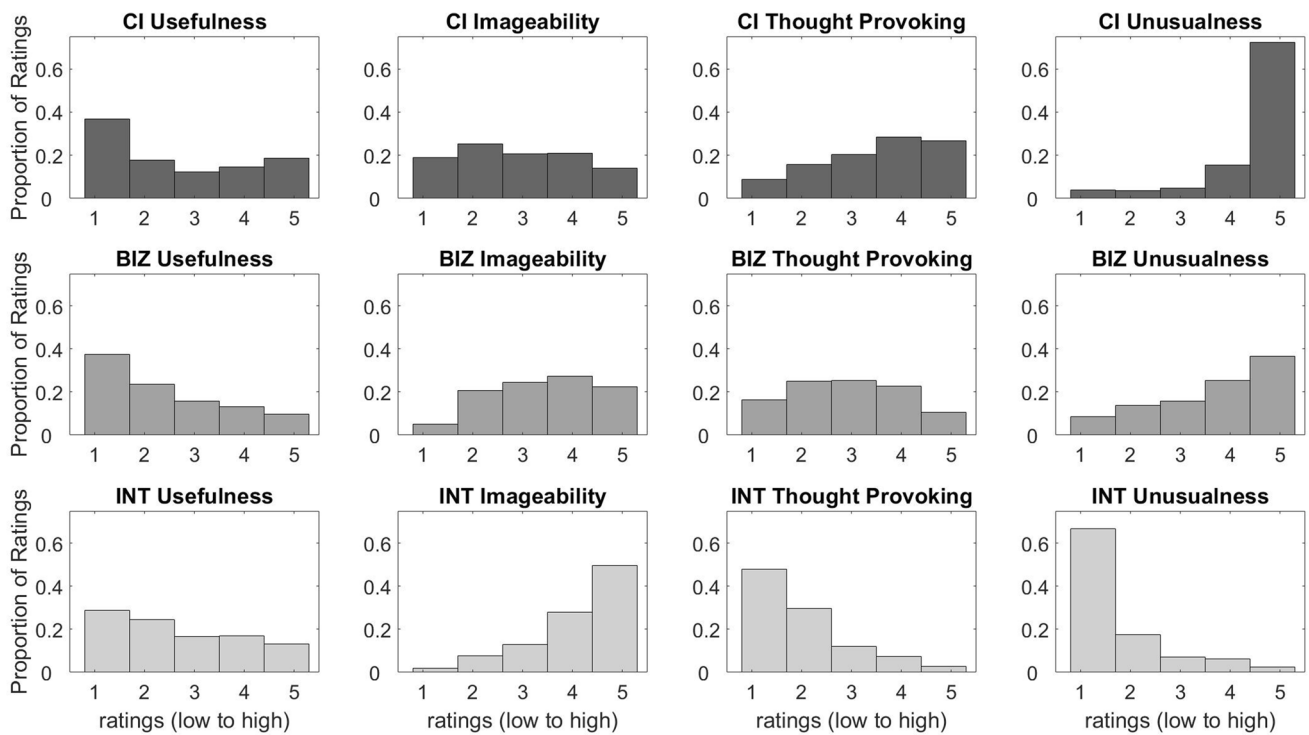
Fig. 3 Histograms of ratings for all items, split by concept category, for dimensions of usefulness, imageability, thought provokingness, and unusualness
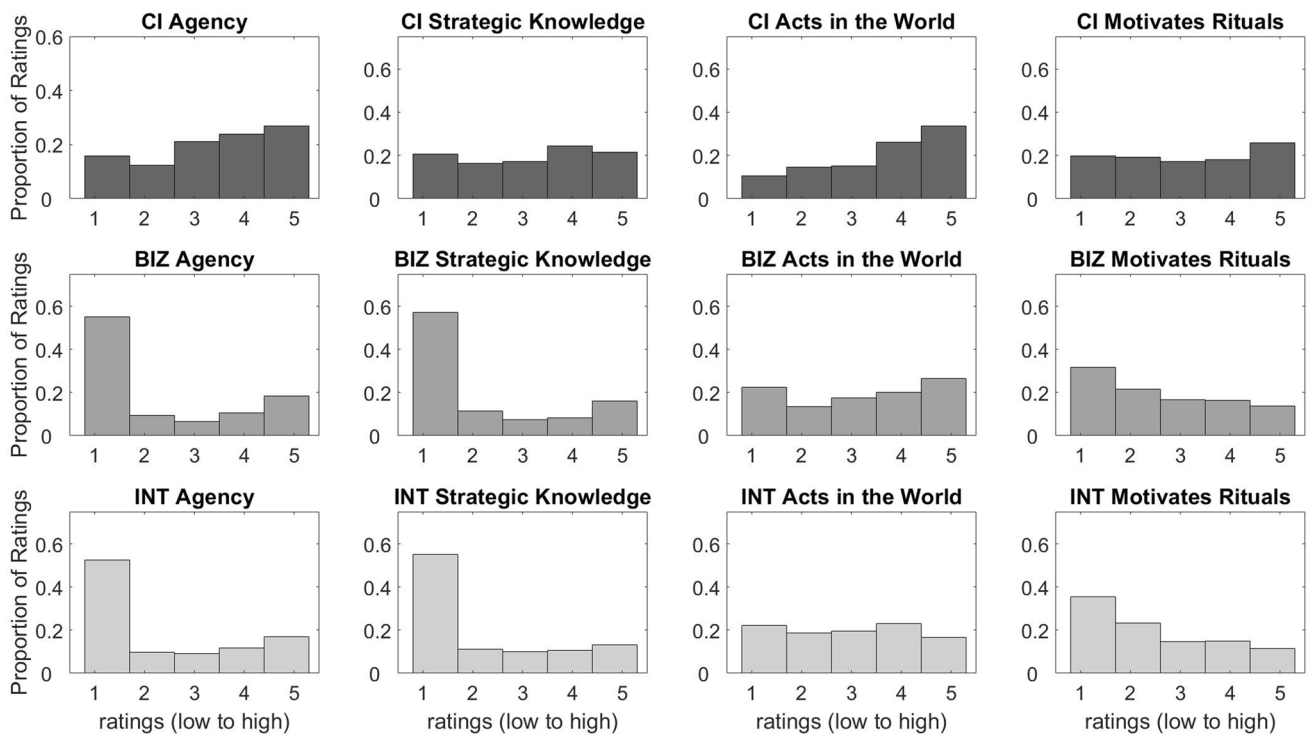

Fig. 4 Histograms of ratings for all items, split by concept category, for dimensions of intentional agency, possession of strategic knowledge, the ability to act in the world, and the tendency to motivate reinforcing behaviors or rituals (adapted from Barrett, 2008a)
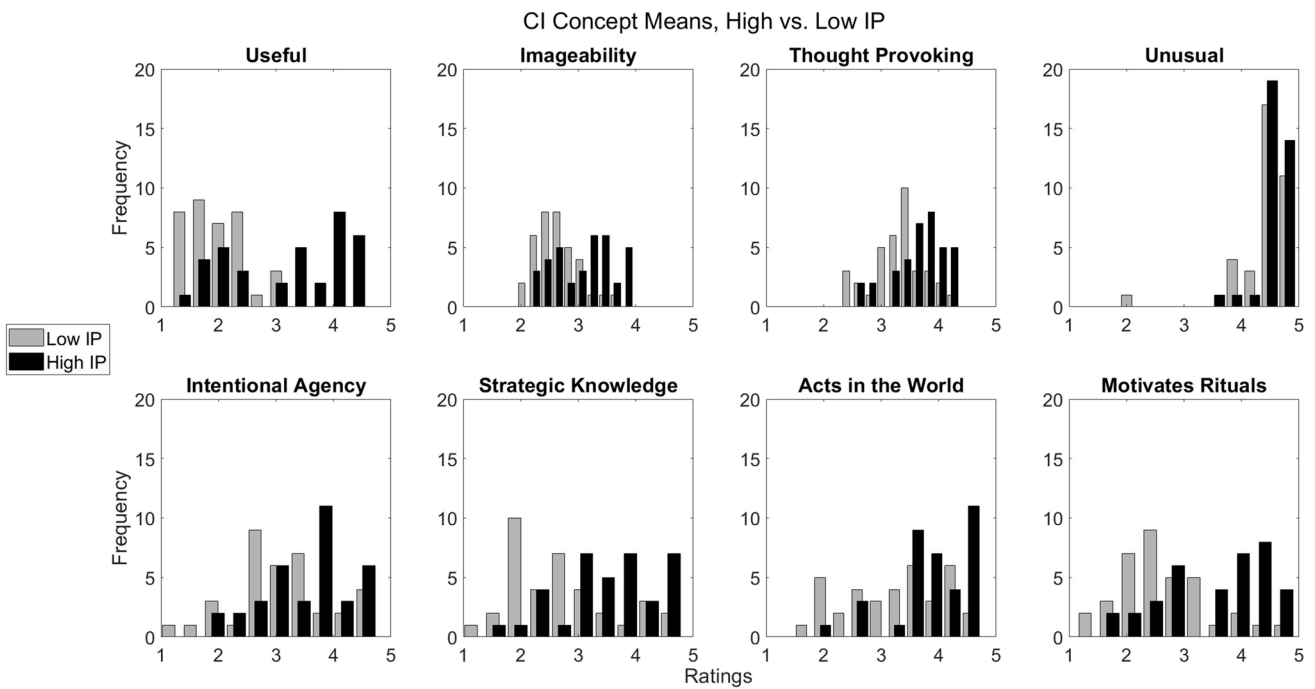
Fig. 5 Histograms of mean ratings for individual CI items in the stimulus set, separated by high and low IP

individual CI item's mean unusualness rating was between 3.5 and 5, and most were between 4.5 and 5. There was little difference between items that were high versus low IP. On the other hand, ratings of CI items on usefulness show that low-IP items were predominantly rated as not useful. In contrast, more than half of the high-IP items were rated as highly useful.

In addition to usefulness differences between high- and low-IP CI items, high-IP items appear to be rated slightly higher on imageability and thought provokingness as well as for all four dimensions in study 2. In contrast, for BIZ and INT items, there are much less pronounced differences between high- or low-IP items on all dimensions. This may be an artifact of some facet of our stimuli design or it might indicate that CI concepts may have a particularly broad range of IP available to them. There is some support for this possibility, especially on the latter four dimensions, where CI items appear to be represented in greater frequencies across the range of possible ratings. If this is the case, the probability that sets of CI concepts might be prone to eliciting extreme memorability results, including those where they are poorly recalled, is increased.

These results also provide further evidence that CI and BIZ items systematically differ. Differences between CI and BIZ ratings are revealed to be largely due to fewer ratings on the very high end of the scale for BIZ items compared to CI items (see Figs. 3, 4, 5, 6 and 7). BIZ items also show less

differentiation between high- and low-IP concepts (Figs. 5, 6 and 7). These differences might explain findings in the literature where degree of unusualness does not necessarily correlate with recall (Barrett & Nyhof, 2001) and results where BIZ items outperform MCI items (e.g., Norenzayan & Atran, 2004). If other factors influence memorability, differences on these dimensions might dominate the effect of unusualness and if a set of CI items is particularly low on these dimensions or a set of BIZ items happens to be particularly high, BIZ items might be better remembered than CI concepts.

Agency

The concept of agency is closely associated with the MCI literature, as many of the prominent entities the theory is meant to explain are supernatural agents. As we noted above, previous research has found a memorability advantage for agentic MCI concepts, which complements the literature demonstrating a general mnemonic advantage for agents (Nairne et al., 2013; Nairne et al., 2017; VanArsdall et al., 2015). Additionally, Barrett's (2008a) characteristics of successful religious concepts used in study 2 are closely related to agency.

We recently speculated that a potential contributing factor to the success of CI concepts is that they can achieve supernatural levels of agency (Sommer et al., 2022). This
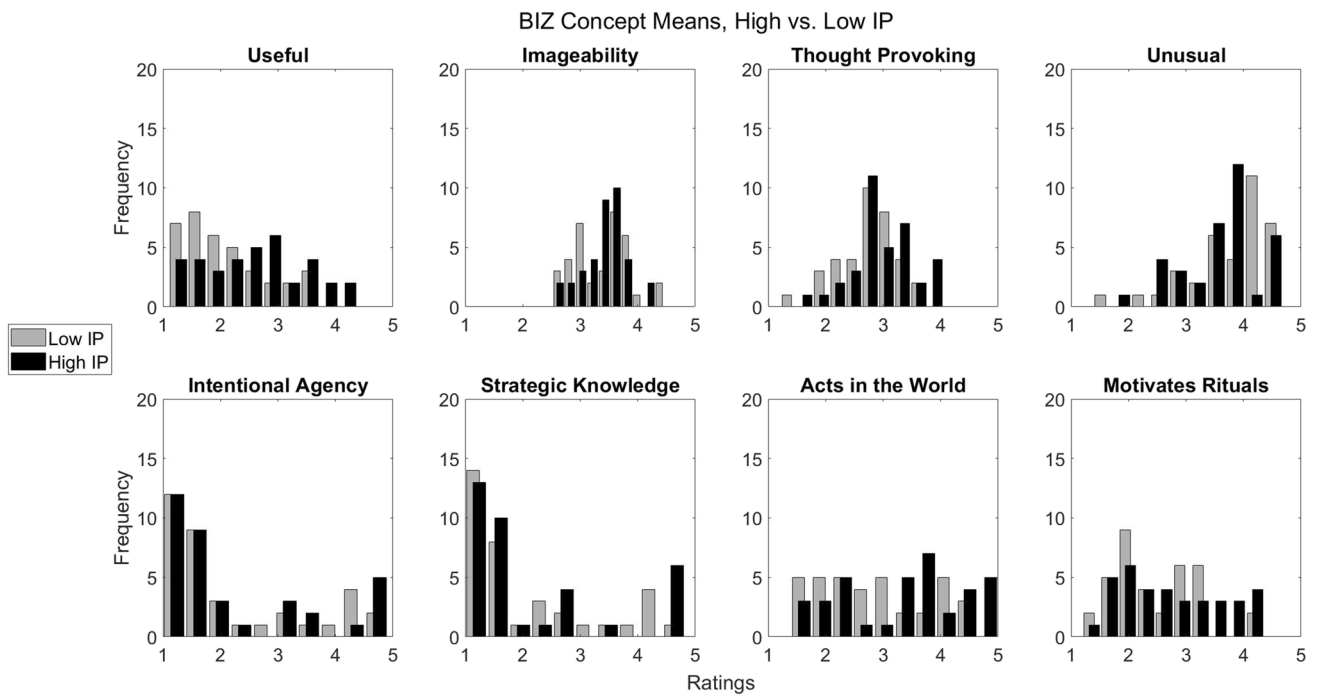
Fig. 6 Histograms of mean ratings for individual BIZ items in the stimulus set, separated by high and low IP
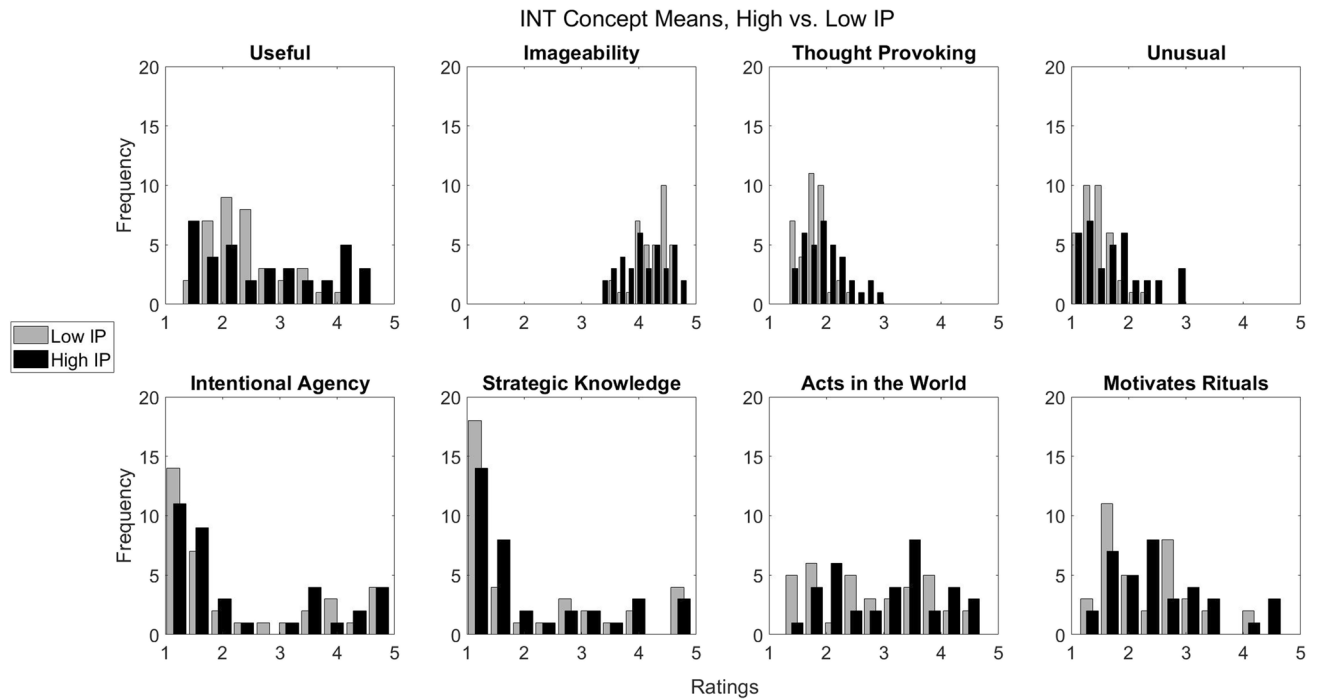

Fig. 7 Histograms of mean ratings for individual INT items in the stimulus set, separated by high and low IP
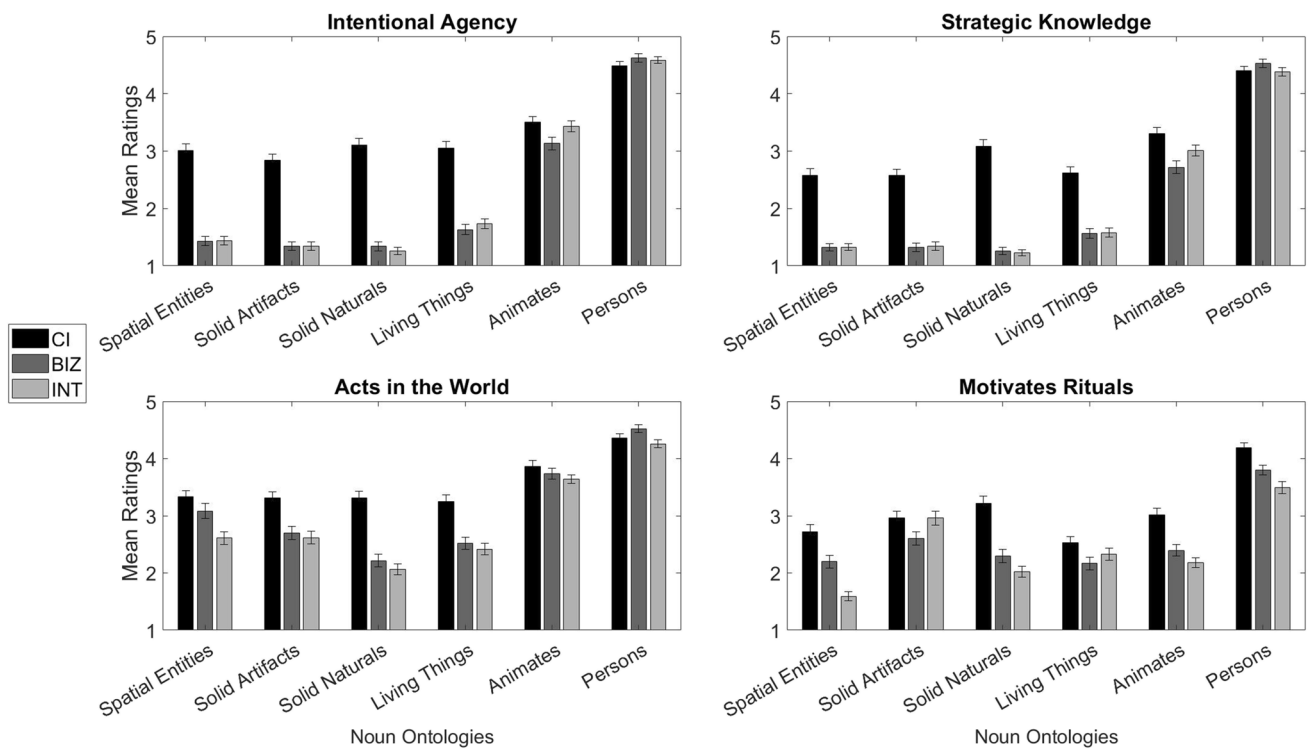
Fig. 8 Mean ratings for CI, BIZ, and INT items by ontological category of each concepts' nouns for the dimensions of intentional agency, possession of strategic knowledge, the ability to act in the world, and the tendency to motivate reinforcing behaviors or rituals (Barrett, 2008a). Error bars represent standard error of the means

can occur in two ways. First and most obviously, supernatural properties can offer agentic abilities that are not possible for natural items to match. A second possibility is that supernatural concepts gain agency by adding agentic characteristics to nouns drawn from inanimate ontological categories. For example, a CI person can gain agency by supernaturally increasing their awareness, as in *a person who can read minds*. In this way, CI items might come to have more agency than BIZ or INT items, on average, due to increased agency in nouns that already had some agency. However, an alternative route would be a CI non-agent gaining any agency at all, such as *a hammer that can speak*.[6]

To assess whether this might be partially responsible for CI concepts' advantage on the dimensions adapted from Barrett (2008a), we split the ratings data based on the ontological category of the nouns in the stimulus set. Recall that concepts began with nouns drawn from the ontological categories of spatial entities, solid objects (both artifacts and natural), living things, animates, and persons (see Barrett, 2008a for more detail on these categories). The results displayed in Fig. 8 show that at least for intentional agency and strategic knowledge, the two dimensions with the greatest CI advantage, BIZ and INT items from the first four ontological categories are at floor (see Table 2 for descriptive statistics). Only by gaining CI characteristics can items in these non-agentic ontologies achieve higher ratings. These findings suggest an additional mechanism by which CI items might be better remembered than other concepts.

Conclusion

We have argued that several puzzling empirical and theoretical questions that have plagued the MCI literature can be best addressed by a standardized set of stimuli that put the field on an even footing and allow for rigorous control of variables long theorized to be relevant, but rarely defined

---

[6] This perspective assumes that agency is a continuum rather than a binary where an entity is either an agent or a non-agent. If agents are characterized by abilities like self-propelled movement and goal-directed behavior, a graded notion of agency might be appropriate. An ant could be said to be less agentic than a mouse which is, in turn, less of an agent than a person. Likewise, a person who is inebriated has less agency than someone who is sober, but both remain agents. Supernatural characteristics can enhance (or reduce, see "Concerns with the MCI literature" section) these properties of agents. For example, a person who can move at the speed of light has enhanced self-propelled motion and a person who can read minds can more easily achieve certain potentially nefarious goals. See Beebe and Duffy (2020), pp. 5–6) for similar considerations and Gray et al. (2007) for empirical evidence that people's intuitions of agency are graded.

Table 2 Mean ratings by ontology

Rating dimension	Characteristics type	Mean rating (SD)					
		Spatial entities	Solid artifacts	Solid naturals	Living things	Animates	Persons
Intentional agency	CI	3.01 (1.43)	2.84 (1.38)	3.11 (1.31)	3.05 (1.42)	3.51 (1.12)	4.49 (0.93)
	BIZ	1.43 (0.97)	1.34 (0.9)	1.34 (0.91)	1.63 (1.12)	3.13 (1.35)	4.63 (0.86)
	INT	1.44 (0.94)	1.34 (0.88)	1.26 (0.76)	1.74 (1.12)	3.43 (1.22)	4.59 (0.74)
Strategic knowledge	CI	2.58 (1.47)	2.58 (1.36)	3.08 (1.38)	2.62 (1.34)	3.31 (1.12)	4.4 (0.93)
	BIZ	1.32 (0.8)	1.32 (0.87)	1.26 (0.76)	1.56 (1.01)	2.72 (1.34)	4.53 (0.91)
	INT	1.32 (0.74)	1.34 (0.88)	1.23 (0.65)	1.58 (1.02)	3.01 (1.17)	4.38 (0.95)
Acts in the world	CI	3.34 (1.3)	3.31 (1.41)	3.31 (1.45)	3.25 (1.37)	3.86 (1.12)	4.36 (0.91)
	BIZ	3.08 (1.57)	2.7 (1.4)	2.22 (1.36)	2.52 (1.28)	3.74 (1.16)	4.52 (0.85)
	INT	2.61 (1.4)	2.62 (1.38)	2.06 (1.15)	2.42 (1.25)	3.64 (0.91)	4.26 (0.89)
Motivates rituals	CI	2.72 (1.5)	3 (1.44)	3.21 (1.55)	2.53 (1.31)	3.01 (1.12)	4.19 (1.09)
	BIZ	2.2 (1.32)	2.6 (1.41)	2.3 (1.42)	2.17 (1.31)	2.39 (1.25)	3.8 (1.1)
	INT	1.6 (0.99)	2.96 (1.49)	2.03 (1.18)	2.33 (1.38)	2.18 (1.07)	3.49 (1.31)

or controlled for in empirical studies. Failure to consider variables like the VR effect and IP, perhaps because they were assumed to be swamped by the effect of violations of intuitive ontological theories, has led to the creation and use of heterogeneous stimuli across studies, complicating the analysis of their experimental findings.

Here, we developed and rated a set of stimuli containing 216 CI, BIZ, and INT control concepts. Rating results reveal important IP, agency, and unusualness differences between and within concept categories, which strongly suggests that controlling for these variables is critical to understanding the memorability of supernatural concepts. This procedure offers a common method for creating additional stimuli and for controlling for IP in future experiments. This method creates stimuli which allow direct comparisons of CI and INT items at different levels of IP to ascertain the role of ontological violations in memory. It further permits comparisons between CI and BIZ concepts, which should shed light on the relationship between the VR and MCI effects. Finally, not only can MCI concepts be compared to MXCI concepts (with three violations) while controlling for IP, but the same can be done for "MXBIZ" (maximally bizarre) concepts with three bizarre properties. To our knowledge, no published study has yet tested the critical prediction that MXCI concepts should be less memorable than MXBIZ concepts due to their proposed unique loss of IP or conceptual breakdown. Resolution of these outstanding questions is imperative for achieving an understanding of the nature of the MCI effect as well as the mechanisms underlying the prevalence of supernatural concepts across cultures.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13428-022-01826-3.

References

Atran, S., & Norenzayan, A. (2005). Religion's evolutionary landscape: Counterintuition, commitment, compassion, communion. *Behavioral and Brain Sciences, 27*(6), 713–730.

Baillargeon, R. (1994). How do infants learn about the physical world? *Current Directions in Psychological Science, 3*(5), 133–140.

Baillargeon, R. (1998). Infants' understanding of the physical world. In M. Sabourin, F. Craik, & M. Robert (Eds.), *Advances in Psychological Science* (Vol. 2, pp. 503–509). Psychology Press.

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

Barrett, J. L. (1999). Theological correctness: Cognitive constraint and the study of religion. *Method & Theory in the Study of Religion, 11*(4), 325–339.

Barrett, J. L. (2004). Counterfactuality in counterintuitive religious concepts. *Behavioral and Brain Sciences, 27*(6), 731–732.

Barrett, J. L. (2007). Cognitive science of religion: What is it and why is it? *Religion Compass, 1*(6), 768–786.

Barrett, J. L. (2008a). Coding and quantifying counterintuitiveness in religious concepts: Theoretical and methodological reflections. *Method & Theory in the Study of Religion, 20*(4), 308–338.

Barrett, J. L. (2008b). Why Santa Claus is not a god. *Journal of Cognition and Culture, 8*(1), 149–161.

Barrett, J. L., & Keil, F. C. (1996). Conceptualizing a non-natural entity: anthropomorphism in God concepts. *Cognitive Psychology, 31*, 219–247.

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

Beebe, J. R., & Duffy, L. (2020). The memorability of supernatural concepts: Effects of minimal counterintuitiveness, moral valence, and existential anxiety on recall. *The International Journal for the Psychology of Religion, 30*(4), 322–341.

Bendixen, T., & Purzycki, B. G. (2021). Competing forces account for the stability and evolution of religious beliefs. *The International Journal for the Psychology of Religion, 31*(4), 307–312.

Boyer, P. (1994). Cognitive constraints on cultural representations: Natural ontologies and religious ideas. In: *Mapping the mind: Domain specificity in cognition and culture*, 391–411.

Boyer, P. (1996). What makes anthropomorphism natural: Intuitive ontology and cultural representations. *Journal of the Royal Anthropological Institute*, 83–97.

Boyer, P. (1998). Cognitive tracks of cultural inheritance: How evolved intuitive ontology governs cultural transmission. *American Anthropologist, 100*(4), 876–889.

Boyer, P. (2001). *Religion Explained: The human instincts that fashion gods, spirits and ancestors*. Random House.

Boyer, P., & Ramble, C. (2001). Cognitive templates for religious concepts: Cross cultural evidence for recall of counter-intuitive representations. *Cognitive Science, 25*(4), 535–564.

Burdett, E. R., Porter, T. J., & Barrett, J. L. (2009). Counterintuitiveness in folktales: Finding the cognitive optimum. *Journal of Cognition and Culture, 9*(3), 271–287.

Carey, S., & Spelke, E. (1994). Domain-specific knowledge and conceptual change. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and cognition*. Cambridge University Press.

Carney, J., & Mac Carron, P. (2017). Comic-Book superheroes and prosocial agency: a large-scale quantitative analysis of the effects of cognitive factors on popular representations. *Journal of Cognition and Culture, 17*(3-4), 306–330.

Chee, Q. W., & Goh, W. D. (2018). What explains the von Restorff effect? Contrasting distinctive processing and retrieval cue efficacy. *Journal of Memory and Language, 99*, 49–61.

Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition, 38*(3), 213–244.

Gonce, L. O., Upal, M. A., Slone, D. J., & Tweney, D. R. (2006). Role of context in the recall of counterintuitive concepts. *Journal of Cognition and Culture, 6*(3), 521–547.

Gopnik, A., & Schulz, L. (2004). Mechanisms of theory formation in young children. *Trends in Cognitive Sciences, 8*(8), 371–377.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619–619.

Gregory, J. P., & Barrett, J. L. (2009). Epistemology and counterintuitiveness: Role and relationship in epidemiology of cultural representations. *Journal of Cognition and Culture, 9*(3), 289–314.

Gregory, J. P., & Greenway, T. S. (2017). The mnemonic of intuitive ontology violation is not the distinctiveness effect: Evidence from a broad age spectrum of persons in the UK and China during a free-recall task. *Journal of Cognition and Culture, 17*(3-4), 253–280.

Harmon-Vukic´, M. E., & Slone, D. J. (2009). The effect of integration on recall of counterintuitive stories. *Journal of Cognition and Culture, 9*(1/2), 57–68.

Johnson, C. V., Kelly, S. W., & Bishop, P. (2010). Measuring the mnemonic advantage of counter-intuitive and counter-schematic concepts. *Journal of Cognition and Culture, 10*(1), 109–121.

Kelley, M. R., & Nairne, J. S. (2001). von Restorff revisited: isolation, generation, and memory for order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 54.

Leslie, A. M. (1982). The perception of causality in infants. *Perception, 11*, 173–186.

Lisdorf, A. (2004). The spread of non-natural concepts. *Journal of Cognition and Culture, 4*(1), 151–173.

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*(5-6), 389–395.

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science, 24*(10), 2099–2105.

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science, 26*(1), 22–27.

Norenzayan, A., & Atran, S. (2004). Cognitive and emotional processes in the cultural transmission of natural and nonnatural beliefs. *The Psychological Foundations of Culture*, 149–169.

Paivio, A. (1986). *Mental representations: A Dual-Coding Approach*. Oxford University Press.

Pinker, S. (1997). *How The Mind Works*. WW Norton & Company.

Pinker, S. (2003). *The Blank Slate: The modern denial of human nature*. Penguin.

Porubanova, M., Shaw, D. J., McKay, R., & Xygalatas, D. (2014). Memory for expectation-violating concepts: The effects of agents and cultural familiarity. *PloS One, 9*(4), e90684.

Porubanova-Norquist, M., Shaw, D. J., & Xygalatas, D. (2013). Minimal Counterintuitiveness Revisited: Effects of cultural and ontological violations on concept memorability. *Journal for the Cognitive Science of Religion, 1*(2), 181.

Purzycki, B. G., & Willard, A. K. (2016). MCI theory: A critical discussion. *Religion, Brain & Behavior, 6*(3), 207–248.

Shtulman, A. (2017). *Scienceblind: Why our intuitive theories about the world are so often wrong*. Hachette.

Sommer, J., Musolino, J., & Hemmer, P. (2022). The Memorability of Supernatural Concepts: Some Puzzles and New Theoretical Directions. *Journal of Cognition and Culture, 22*(1–2), 90–135.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science, 10*(1), 89–96.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99*(4), 605.

Stubbersfield, J., & Tehrani, J. (2013). Expect the unexpected? Testing for minimally counterintuitive (MCI) bias in the transmission of contemporary legends: A computational phylogenetic approach. *Social Science Computer Review, 31*(1), 90–102.

Swan, T., & Halberstadt, J. (2021). Anxiety Enhances Recall of Supernatural Agents. *The International Journal for the Psychology of Religion*, 1–17.

VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N., & Cogdill, M. (2015). Adaptive memory: Animacy effects persist in paired-associate learning. *Memory, 23*(5), 657–663.

Von Restorff, H. (1933). Ueber die Wirkung von Bereichsbildungen im Spurenfeld. Analyse von Vorgängen im Spurenfeld. I. Von W. Köhler und H. V. Restorff. *Psychologische Forschung*.