

ARE SUPERNATURAL CONCEPTS A COGNITIVE EPIPHENOMENON?

By

JOSEPH J. SOMMER

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Pernille Hemmer and Julien Musolino

And approved by

New Brunswick, New Jersey

January, 2021

ABSTRACT OF THE THESIS

Are Supernatural Concepts A Cognitive Epiphenomenon?

by JOSEPH J. SOMMER

Thesis Directors:

Pernille Hemmer & Julien Musolino

I offer a systematic evaluation of the literature on the memorability of supernatural concepts (e.g., gods, ghosts, souls), itself part of a growing body of work in the newly-emerging cognitive science of religion (Barrett, 2000). Specifically, I focus on Boyer's (1994, 2000, 2001) pioneering Minimally Counterintuitive (MCI) hypothesis according to which supernatural concepts tap a cognitively privileged memory-enhancing mechanism linked to violations of default intuitive inferences. My assessment reveals that the literature on the MCI hypothesis is mired in empirical contradictions and methodological shortcomings which makes it difficult to assess the validity of competing theoretical models, including the MCI hypothesis itself. In light of this fractured picture, I propose a novel and independently motivated account of the memorability of supernatural concepts. This new account is flexible enough to make sense of the heterogenous set of empirical findings in the literature and precise enough to make clear empirical predictions that differ from those of other accounts, including the MCI hypothesis. I conclude with a set of theoretical and methodological prescriptions designed to guide future research on the memorability of supernatural concepts.

Contents

Abstract	ii
1. Introduction	2
2. The MCI thesis then and now	7
3. Empirical challenges to the MCI account	15
3.1. Failed replications? Some reasons to be skeptical	15
3.2. Two key assumptions in the MCI literature.....	19
3.3. Is the MCI effect epiphenomenal?.....	20
3.3.1. Inferential potential.....	22
3.3.2. Coherability	28
3.4. Reassessing the empirical challenge	31
4. Theoretical challenges to the MCI account	33
4.1. The MCI effect as a contextual epiphenomenon	34
4.2. The inherent vagueness of counterintuitiveness.....	39
4.3. The MCI effect operates over beliefs sets rather than beliefs.....	42
4.4. The problem of belief.....	44
4.5 The MCI account and the conceptual combination problem	45
5. Putting the pieces of the MCI puzzle together	48
References	544

1. Introduction

Supernatural concepts such as gods, spirits, and souls have been a ubiquitous feature of human cultures since (at least) the beginnings of recorded history. As Baker and Goetz (2011) put it “Most people, at most times, in most places, at most ages, have believed that people have some kind of soul”. This statement can be easily extended to a number of related entities suggesting that supernatural concepts may represent a human universal (Norenzayan, 2010). From an evolutionary perspective, supernatural concepts pose an interesting puzzle in the sense that gods and souls, unlike trees and lions, cannot be apprehended in any obvious way from our regular modes of perception (Bloom, 2012). To be sure, one can hear a lion’s roar and run to hide behind a tree, but one cannot exactly see an invisible god or hear a soul’s lament. Why then do people the world over entertain such concepts, and why have these ideas been so successfully transmitted across generations?

Given the intimate relationship between human culture and religion, one might be inclined to regard supernatural concepts as the cultural manifestation of an evolutionary adaptation for religion. Indeed, proposals along these lines have been made in recent theorizing on this topic (e.g., Bering, 2006, 2012). However, it is also well-known that not all features of biological organisms are adaptations (Gould & Lewontin, 1979). Thus, an influential alternative to the adaptationist view of religious thought treats supernatural concepts as a byproduct of ordinary cognition. On this view, most famously articulated by Pascal Boyer (1994, 1996, 2000, 2002), supernatural concepts are ordinary concepts that have been minimally altered to give rise their otherworldly qualities. For example, a ghost is a person with human-like mental abilities, albeit one that violates our intuitive expectations in the domains of physics (i.e., does not have a material body) and biology (i.e., is immortal).

The central idea underlying Boyer's pioneering thesis is that a small number of intuitive violations such as those described above enhances the salience of the relevant concepts and hence their memorability. Everything else being equal, such a mnemonic advantage for supernatural concepts makes them better candidates for cultural transmission. If the number of violations remains small, say no more than one or two (Barrett, 2008a), the core concept will remain relatively easy to represent and reason about. In this sense, supernatural concepts, viewed as minimally counterintuitive (MCI) concepts, may be a cognitive optimum for memory, representing the ideal tradeoff between salience and simplicity.

In the years following Boyer's formulation of the MCI hypothesis, its main predictions have received empirical support in Western adults (Barrett & Nyhof, 2001; Johnson, Kelly, & Bishop, 2010), but also cross-culturally (Boyer & Ramble, 2001), and even in young children (Banerjee, Haque, & Spelke, 2013). Specifically, what these studies reveal, across a wide range of participants, is that MCI concepts are indeed more memorable than their intuitive counterparts, and that the mnemonic advantage decreases, and even returns to baseline levels, as the number or complexity of violations increases. In more recent developments, however, a number of challenges have emerged that call into question the robustness of the empirical results obtained in tests of the MCI hypothesis (Gregory & Barrett, 2009; Norenzayan & Atran, 2004; Porubanova-Norquist, Shaw, & Xygalatas, 2014; Purzycki & Willard, 2015) as well as the validity of the theoretical mechanisms underlying the enhanced memorability of MCI items (Franks, 2003; Norenzayan & Atran, 2004; Willard, Henrich, & Norenzayan, 2016). Finally, rival accounts have been proposed to explain the empirical findings ostensibly supporting the MCI effect (Upal, Gonce, Tweney & Slone, 2007; Upal, 2010). These developments, in turn,

raise a number of questions and puzzles for anyone interested in the nature and cultural prevalence of supernatural concepts.

In this article, we present a detailed analysis of the core theoretical and empirical issues surrounding the MCI thesis. In doing so, we distinguish three important aspects of the MCI account. The first is Boyer's hypothesis regarding the nature of supernatural concepts, namely the contention that they are modified versions of ordinary concepts. We call this the *ontological* question. The second is the shape of the effect predicted by the MCI account. Here, the predicted effect has a parabolic shape with MCI concepts (with one or two violations) being the most memorable while intuitive and maximally counterintuitive concepts (with 3 or more violations) register at lower, and roughly equivalent levels of, memorability. We call this the *empirical* question. The third aspect of the MCI account is the nature of the theoretical mechanisms that are believed to be responsible for the enhanced memorability of MCI concepts and the decrease in memorability of maximally counterintuitive concepts. We call this the *theoretical* question.

The upshot of our deliberations is that all three of these questions turn out to be far more complex than initially anticipated. Consequently, more than two decades after the publication of Boyer's original ideas, and in spite of a growing body of experimental results, it may be that the central claims of the MCI account regarding the memorability of supernatural concepts have yet to receive an adequate empirical test. This is in large part because, with the benefit of hindsight, it is now possible to see that a host of variables, many originally unrecognized, are likely to affect memorability and would thus need to be carefully controlled for in order to confirm that the kinds of intuitive violations originally described by Boyer do indeed have the effect predicted by the MCI account. Put another way, we still do not know whether the MCI effect is fundamental, as proposed by Boyer, or whether it is epiphenomenal and emerges from the

operation of other known cognitive mechanisms that do not rely on the special status of the intuitive violations introduced by Boyer (see Barrett, 2008a and Purzycki & Willard, 2015 for similar considerations).

In order to give epiphenomenalism some teeth, we propose a new account of the MCI effect that makes no reference to the special status of the kinds of intuitive violations described by Boyer. This new account of the memorability of supernatural concepts has a number of desirable features. First, it makes clear, testable empirical predictions that are different from those of any of the accounts currently available in the literature, including Boyer's original MCI thesis. Second, epiphenomenalism paves the way to a resolution, or at least a better understanding, of the apparent contradictions inherent in the extant literature. Specifically, under our new account, the shape of the empirical effect predicted by the original MCI hypothesis is expected to obtain under certain conditions but also to vary in a theoretically predictable manner that we spell out. Third, our new account preserves the explanatory value of Boyer's thesis and potentially extends it to new phenomena. Finally, several theoretical issues that challenge the traditional version of the MCI thesis can be more easily accommodated by the epiphenomenal explanation. To be clear, we do not claim that epiphenomenalism is necessarily the correct account of the memorability of supernatural concepts but rather that it is a coherent, parsimonious, and empirically testable hypothesis that needs to be experimentally evaluated, and ruled out, before strong conclusions can be reached about the validity of Boyer's original MCI thesis.

In short, we propose that there is nothing special about violating intuitive ontological theories that cannot be explained by characteristics of the resulting concepts. Instead, the memorability advantage possessed by MCI concepts may result from more general mechanisms

of memory. It has been long observed in the MCI literature that the Von Restorff (1933) effect, which describes a pattern of improved memorability for outlier items in a list, bears a strong resemblance to the MCI effect. Additionally, Boyer (1996, 2002) noted early on that all supernatural concepts are not created equal, suggesting that in addition to salience, concepts required “inferential potential” to be successful. We suggest that variation in these factors are sufficient to explain both general improvements in memorability of supernatural concepts as well as experimental results where MCI items are remembered at lower rates than intuitive items. However, as these variables are unable to explain the reduction in memorability for complex supernatural items, we propose a novel variable, “coherability,” which describes a decreased ability to rely on prior knowledge as supernatural concepts grow more complicated. We believe these variables preserve the spirit of Boyer’s MCI thesis while addressing both empirical and theoretical challenges leveled at the traditional account.

To reach the conclusions outlined above, we organize our discussion as follows. In section 2, we begin by reviewing the fundamentals of the MCI account, including the supportive evidence found in the literature as well as the empirical and theoretical challenges that have emerged since the formulation of Boyer’s original account. Section 3 focuses on the empirical challenges to the MCI account. There, we propose a novel analysis of the memorability of supernatural concepts which regards the MCI effect as a cognitive epiphenomenon. Next, in section 4, we turn to the theoretical challenges that have been leveled against the MCI account and show that none of them pose a direct threat to Boyer’s original thesis which, when interpreted in light of the account we developed in section 3, remains a live option. Finally, section 5 provides an assessment of the state-of-the-art regarding the memorability of

supernatural concepts and offers a number of theoretical and methodological prescriptions to guide future research on this topic.

2. The MCI thesis then and now

In his ground-breaking work on the origins of religious concepts, Boyer (1994, 1998, 2002) introduced several key ideas that have played a pivotal role in the emergence of the newly-established cognitive science of religion (Barrett, 2000). Taken together, Boyer's ideas form what is known as the minimally counterintuitive (MCI) thesis (Barrett, 2008a), an account that rests on two central assumptions. The first is that supernatural concepts, in spite of their otherworldly qualities, are in fact drawn from the store of run-of-the-mill concepts that furnish the human mind. Underlying these concepts are a set of basic ontological categories such as ANIMAL, PLANT, NATURAL OBJECT, and ARTIFACT. A small set of intuitive (or folk) theories allow us to reason about these ontological categories and support a rich set of inferences that are near-universal, early developing, and relevant to specific domains with evolutionary functions (Gelman & Wellman, 1991; Gopnik & Schulz, 2004; Pinker, 1997, 2003; Shtulman, 2017; Spelke, et al. 1992; Spelke & Kinzler, 2007). For example, intuitive physics gives rise to the expectation that solid objects cannot pass through each other and that unsupported objects fall to the ground (Baillargeon, 1994, 1998; Carey & Spelke, 1994; Leslie, 1982). Likewise, intuitive biology prohibits an animal from giving birth to a young from a different species than its own and maintains species specific essences in the face of changes performed on individual animals (Atran, 1998; Berlin, Breedlove, & Raven, 1973; Gelman & Hirschfeld, 1998).

The second assumption underlying the MCI thesis is that supernatural concepts contain characteristics that violate one or more of the expectations associated with intuitive theories. For example, on this view, a ghost, an immortal supernatural entity that can walk through walls, belongs to the ontological category PERSON but violates the solidity constraint associated with intuitive physics (persons, being solid objects, cannot pass through walls) and the mortality constraint associated with intuitive biology (biological organisms have an expiration date). Other properties are preserved such that aside from the conspicuous violations, one can reason about ghosts in essentially the same way that one would about persons in general.¹

The key idea now is to consider what happens when the two core assumptions described above interact. Because the violations of intuitive expectations are surprising, they are salient, which increases the likelihood that they will be remembered. At the same time, the preserved inferential structure associated with the base ontological categories to which the supernatural concepts belong allow these concepts to be represented and reasoned about without much effort. However, Boyer theorized, violating too many intuitive ontological theories would block this inferential structure and inhibit representation of complex concepts. Building on these observations, Boyer proposed that successful supernatural concepts will be “minimally counterintuitive,” possessing only a small number of violations of intuitive expectations because this represents the optimal tradeoff between salience and complexity. On this view, an increase

¹ Note that in the example above, the emergence of a supernatural concept arises through a *breach* of expectations associated with the domains of intuitive physics and biology, because their supernatural characteristics aren’t associated with any natural concept. On Boyer’s account, another way to transform an ordinary concept into a supernatural one is through a *transfer* of expectations from one ontological category to another (Boyer & Ramble, 2001). Consider, for example, a statue that can listen to people’s prayers and cries when it is sad. Here, the base concept, statue, is drawn from the ontological category ARTIFACT. Since artifacts are man-made objects, they are not expected to engage intuitive psychology or biology. However, in this example, properties that typically hold of persons, like the ability to listen and cry, are transferred to a statue, thereby transforming an otherwise ordinary artifact into a supernatural entity.

in the number of violations will lead to decreasing marginal returns of salience for the increased complexity. Thus, a person who can fly might be a good candidate for a supernatural concept, but a person who can fly and who lays eggs that hatch into lizards and who only exists from 2:00p.m. – 4:00 p.m. would be too cognitively taxing for people to entertain despite the salience of its characteristics.

Boyer proposes that all else being equal, concepts that enjoy a mnemonic advantage are more likely to be successfully transmitted and eventually adopted as part of a belief system (Boyer & Ramble, 2001). A desirable feature of this cognitive account of religious thought is that it offers a potential explanation for the striking commonalities in religious concepts found around the world while also accounting for variation among cultures (Barrett, 2000; Boyer, 2003). That is, because the advantage of supernatural concepts is structural, these concepts can achieve success despite variability in superficial characteristics. Another virtue of the MCI account is that it makes testable empirical predictions. First, compared to items with only intuitive characteristics, items with supernatural properties, defined as characteristics that violate² intuitive ontological expectations, should be better remembered. Second, as the number of supernatural properties increases, this memory advantage for supernatural items should correspondingly decrease, as shown in Figure 1 below.

² Henceforth, and for convenience's sake, we use the term 'violate' to encompass both breaches and transfers of ontological expectations.

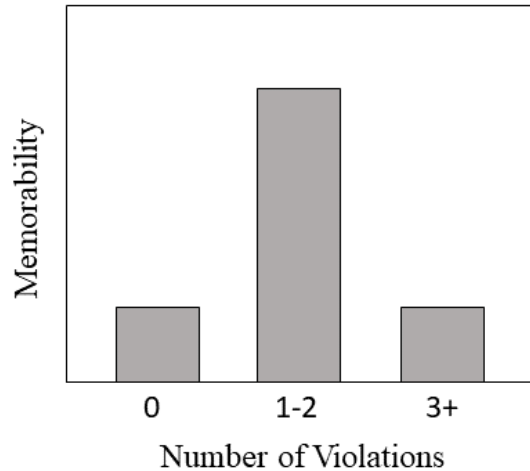


Figure 1: Memory for supernatural items with a given number of violations of intuitive assumptions compared to intuitive items with the same number of mundane characteristics.

Supporting evidence for these predictions has been reported in a number of studies involving participants of different ages and cultural backgrounds (Boyer & Ramble, 2001; Barrett & Nyhof, 2001; Johnson, Kelly, & Bishop, 2010; Banerjee, Haque, & Spelke, 2013). In these studies, participants typically read or listen to stories containing minimally counterintuitive (MCI), intuitive (INT), and maximally counterintuitive items (MXCI) (i.e., with 3 or more intuitive violations), complete a short distractor task, and are then asked to recall as many of the items as they can. Often, after a delay, participants are asked to recall the items a second time. The proportions of items recalled from each category are then compared to one another. These comparisons have supported both predictions of the MCI account simultaneously (Boyer & Ramble, 2001; Banerjee, Haque, & Spelke, 2015; Johnson, Kelly, & Bishop, 2010) as well as separately: MCI items are better remembered than INT items (Barrett & Nyhof, 2001; Slone, et al. 2007) and MXCI items are remembered less well than both INT and MCI items (Gonce, et al. 2006; Slone, et al. 2006).

For example, Barrett & Nyhof (2001) presented participants with unfamiliar Native American folk tales that contained both INT (e.g., a plant) and expectation-violating concepts

(e.g., an animate skeleton). After reading the vignettes and completing a distractor task, participants wrote down the stories as they remembered them. These retellings were coded based on how many concepts of each type participants were able to recall. The results confirmed that counterintuitive concepts were significantly more likely to be recalled than intuitive ones. In a similar vein, Boyer & Ramble (2001) conducted a series of experiments in France, Gabon, and Nepal and found the expected mnemonic advantage for MCI concepts in all three populations in spite of significant differences in culture.

In a more recent study, Banerjee, Haque, and Spelke (2013) offered a developmental perspective on the MCI thesis by asking whether young children would also show the effect reported in adults. These authors presented seven-to-nine-year-old children with a story about two characters who explore a new neighborhood and encounter 12 objects. The objects were drawn from the categories of animals, plants, or objects and had either two intuitive characteristics (INT) (e.g., ‘A lizard that ate insects off the ground and crawled around quickly on all four of its feet’) or one INT characteristic as well as one counterintuitive property (CI) (e.g., ‘A lizard that had a long thin tail and could never die no matter how old it was’). After a distractor task, participants were asked to recall as much of the story as they could, aided with prompting questions such as “What did the kids see or find on their way home?” A week later, participants were brought back to the lab and the same recall procedure was followed.

Banerjee et al. found that items with CI characteristics were better remembered than were items with only INT characteristics both immediately and after a week’s delay. In subsequent conditions, the number of characteristics was adjusted to compare memory for INT items with 2 and 3 characteristics to CI items with 2 or 3 violations of intuitive ontologies. For example, “a lizard that ate insects off of the ground, crawled around quickly on all four of its small feet, and

had a long, thin tail” was compared to “a lizard that always melted in the hot sun, could never die no matter how old it was, and could hear other creatures’ thoughts.” Though CI items continued to enjoy a memory advantage over INT items when they had 2 violations, there was no difference in recall for INT items with 3 characteristics and MXCI items with 3 violations (see figure 1).

While the studies reviewed so far provide support for the main empirical predictions of the MCI account, a broader examination of the extant literature reveals that the full picture is in fact more complicated. Alongside the publication of the studies described above, a number of challenges have emerged, calling into question the empirical validity of earlier findings, the scope of the MCI effect, as well as the nature of the theoretical mechanisms believed to give rise to the effect in the first place. Of particular concern are a number of studies that 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, Shaw, & Xygalatas, 2014), and cases where MXCI items were remembered at a similar rate as MCI items (Harmon-Vukic & Slone, 2009). Further complicating the picture, the MCI effect has been found to reverse, with INT items being remembered better than MCI items, when the material is presented in a bare list as opposed to the richer context of a narrative (Norenzayan & Atran, 2004) (see Table 1 for a list of studies and memorability profiles for INT, MCI, and MXCI concepts).

Table 1: memorability profiles for INT, MCI, and MXCI items in different studies³

Study	Findings
Boyer & Ramble (2001)	$MXCI (BrTr) \leq INT < MCI$ [Original MCI effect]
Barrett & Nyhof (2001)	$INT < MCI$
Norenzayan & Atran (2003)	$MCI < INT$
Norenzayan, et al. (2006)	Immediate Recall: $MCI < INT$ Delayed Recall: $INT < MCI$
Gonce, et al. (2006)	Immediate Recall: $MXCI < INT = MCI$ Delayed Recall: $MXCI < INT < MCI$
Gregory & Barrett (2009)	$INT < MCI$ for participants under age of 25 and after a week delay only. Otherwise, $MCI < INT$
Johnson, Kelly, & Bishop (2010)	Immediate Recall: $MXCI = MCI$ Delayed Recall: $MXCI < MCI$
Banerjee, Haque, & Spelke (2013)	Immediate and Delayed Recall: $MXCI \leq INT < MCI$
Porubanova-Norquist, Shaw, & Xygalatas (2014)	$INT = MCI$
Porubanova, et al. (2014)	$MCI < INT$

Moreover, alternative mechanisms have been proposed to account for the MCI effect. For example, Upal, Gonce, Twenty, & Slone (2007) and Upal (2010) introduced, and experimentally tested, a model which derives the memorability profile of INT, MCI, and MXCI items from properties of the context in which these items occur rather than properties of the items

³ For purposes of the present review and analysis, we conducted a search for articles on Google Scholar with the keywords "Minimally Counterintuitive concepts" and "MCI theory." All papers that included experimental tests of the MCI theory or that contained theoretical contributions directly related to the theory were considered in writing this review.

themselves. Other investigators have underscored the nature of the violations that give rise to MCI items and called for a more nuanced approach anchored in fine-grained semantic analysis with a range of potential empirical and theoretical ramifications (Franks, 2003). In their critical review of the literature on the MCI effect, Purzycki & Willard (2015) point out that the difference between counterintuitive violations (in Boyer's technical sense) and counter-schematic violations (flouting of cultural expectations), which is central to the theory, has not been considered carefully enough, leading to problematic variation among different studies ostensibly designed to test the predictions of Boyer's original account. In yet other developments, Norenzayan and Atran (2004) suggest that the MCI effect isn't restricted to individual concepts, but, crucially, that it extends to "belief sets," which, these authors believe are the more likely units of cultural transmission. Belief sets that contain few MCI items are thought to be MCI belief sets. According to Norenzayan and Atran (2004), if the original account of the MCI effect has any validity, it is quite weak and is swamped by the larger effect at the level of belief sets.

Finally, there is yet another set of challenges that maintains that the MCI effect is insufficient on its own as an explanation of religious belief. Even if MCI items are more memorable and/or transmittable, these advantages may not be enough to lead to belief in these concepts. This problem is supported by popular MCI concepts that are not believed in (the Mickey Mouse problem) or MCI concepts that used to be believed in, but now are not (the Zeus problem) (Gervais & Henrich, 2010) as well as by studies that find MCI items to be reliably rated as less believable than regular concepts (Willard, Henrich, & Norenzayan, 2016).

Overall, our survey of the literature on the MCI thesis reveals that the picture is muddled. On the empirical side, different studies seem to present contradictory results with some reporting

experimental support for the main predictions of the MCI account while others fail to do so and have even reported patterns going in the opposite direction (see Table 1). Assuming that these disparate findings can be reconciled, further complications arise, this time having to do with the nature of the theoretical mechanisms believed to give rise to the MCI effect itself. Of course, if empirical evidence for the MCI effect truly fails to replicate, theoretical debates are irrelevant. Our first task, therefore, before we present our evaluation of the relative merits of the different theoretical accounts on offer, is to address the empirical side of the MCI equation. This is the topic of the next section.

3. Empirical challenges to the MCI account

3.1. Failed replications? Some reasons to be skeptical

The apparent failures to replicate the MCI effect reported in the studies described in the previous section may be cause for concern, especially in light of the replication crisis in psychology (Pashler & Wagenmakers, 2012). However, it is worth remembering that the original results were not limited to a single study from one team of researchers on an atypical population but rather, as discussed earlier, the pattern has been reported a number of times by different investigators, cross-culturally, and in children (e.g., Banerjee, Haque, & Spelke, 2013; Boyer & Ramble, 2001). This raises the possibility that apparent failures to replicate may be due to the effect of other factors, perhaps leading to miscalibrated comparisons between different studies, rather than to the illusory nature of the effect itself (see Barrett, 2008a; Upal, 2010; Purzycki & Willard, 2015 for similar considerations).

In a similar vein, Barrett (2008a) also points out that the empirical support for the MCI thesis has been mixed, with some studies reporting results that are compatible with the predictions of Boyer's account while others describe findings that are inconsistent with those predictions. However, Barrett observes that different investigators have operationalized the notion of "counterintuitiveness" in different ways. To illustrate, consider the phrase 'pieces of furniture that float in the air if you drop them', an MCI item from Boyer and Ramble (2001), and 'thirsty door', an MCI item used by Norenzayan et al. (2004). First, notice the striking difference in linguistic structure between the two items. Second, while a piece of furniture that defies gravity involves a violation of intuitive physics - and is thus counterintuitive in the sense intended by Boyer - a 'thirsty door' is at best ambiguous. Indeed, participants could interpret the phrase to involve a transfer of physiology to an artifact (thirst as a physiological property of door) but they may also interpret the expression metaphorically and come to believe that a 'thirsty door' means that the wood in the door is dry. If so, the resulting concept would not count as counterintuitive in Boyer's technical sense. In light of this problem, Barrett concludes that these differences in the construction of MCI items alone could account for the apparently contradictory nature of the results found in the literature.

Another likely source of variation among different studies of the MCI effect is the nature of the coding scheme that is used to tally what participants remember in those experiments. Imagine a supernatural concept of the form Noun + characteristic 1 + characteristic 2 (see Banerjee, Haque, & Spelke (2013) for specific examples). Recall of such a concept could be coded in a number of ways. For example, participants could get credit only insofar as they are able to correctly recall the whole concept, that is, the noun and both characteristics. Alternatively, partial credit could be assigned if participants recall only the noun, or even only

some (or all) of the characteristics. Since coding schemes are often heterogenous, it is difficult to know how meaningful comparisons across different studies really are (see Barrett, 2008a for related concerns).

Yet another reason to not take apparent failures to replicate at face value is that even if Boyer had never proposed his MCI account, there would be good reasons to expect that something like the MCI effect, or at least the preferential recall of MCI items over INT items, should be found. To be sure, the von Restorff (VR) effect (1933), also known as the bizarreness or isolation effect (McDaniel et al., 1995; McDaniel, Dornburg, & Guynn, 2005), describes findings that in a homogenous list, outlier items are more likely to be remembered. For example, if embedded in a list of tools, a fruit will be disproportionately recalled. This phenomenon bears a striking resemblance to the MCI effect and indeed, past work in the MCI literature has often referenced the VR effect and attempted to demonstrate that the VR and MCI effects are not one and the same (Barrett & Nyhoff, 2001; Boyer & Ramble, 2001; Atran & Norenzayan, 2005; Gregory & Greenway, 2017).⁴

In these studies, recall of CI items is tested against unusual items (often designated BIZ, for bizarre) that deviate from normal or cultural expectations but do not violate any intuitive ontological expectations. Like the results of experiments on the memorability of MCI items, the pattern of results from this body of work is heterogenous and can be difficult to interpret. Some studies find an equivalent advantage of MCI and BIZ items over INT items (Barrett & Nyhof, 2001, exp. 3). Others find that BIZ items outperform MCI items (Atran & Norenzayan, 2005;

⁴ In fact, one of the original motivations for distinguishing between *breaches* and *transfers* was to demonstrate that the memorability advantage of supernatural concepts was due to a violated intuitive ontological theory and not merely due to representing an unusual or impossible characteristic. If the memory effect disappeared for *transfers* of familiar characteristics to an inappropriate domain, it would have suggested that the advantage was solely due to the bizarreness of the *breaches* (Boyer & Ramble, 2001).

Porubanova, et al., 2013, 2014). Yet others find a recall advantage for MCI over BIZ items (Barrett & Nyhof, exp. 4; Boyer & Ramble, 2001; Johnson, Kelly, & Bishop, 2010). When items are rated for oddness or familiarity, these ratings sometimes correlate with recall (Boyer & Ramble, 2001) but on other occasions, they do not (Barrett & Nyhof, 2001).

In light of these conflicting results, the question of whether the enhanced memorability of MCI items can be reduced to the VR effect should not be considered closed. It remains unclear whether the salience produced from violating expectations of group membership, as in VR, is fundamentally different from the salience produced by violating the expectations of an intuitive ontology, as in MCI. The difference between BIZ and MCI items may well be a difference of degree, rather than of kind, with MCI concepts producing the same quality of surprisal as BIZ items but achieving greater memorability through increased quantity of surprisal. However, even if there is a difference in kind between MCI and VR items, it would be strange if the VR effect suddenly disappeared when MCI items are concerned - after all, MCI items are indeed bizarre or unusual. Thus, there are good reasons to believe that MCI items are salient and that they differ from INT items in that regard. If so, the VR effect would straightforwardly predict a memory advantage for MCI concepts. For this reason, results where INT items are better remembered than MCI items, such as Norenzayan and Atran (2004), are surprising and do call for an explanation.

Finally, past studies of the MCI effect have ignored several theoretical variables that have been allowed to vary wildly within and between studies. Apart from the few studies mentioned above, unusualness has not been controlled for among MCI items. Additionally, while the notion of inferential potential has been discussed in the literature surrounding MCI concepts since its

inception (Boyer, 1996, 2002), it has almost never been operationally defined or controlled for, allowing yet another source of variability to persist in these studies.

3.2. Two key assumptions in the MCI literature

The “failed replication” interpretation of the contradictory studies found in the literature on the MCI effect relies on two related assumptions that are often, although not always, tacitly made in that literature. The first is that all MCI items should be equally memorable and that all MCI items should be more memorable than any INT item.⁵ The second assumption is that the

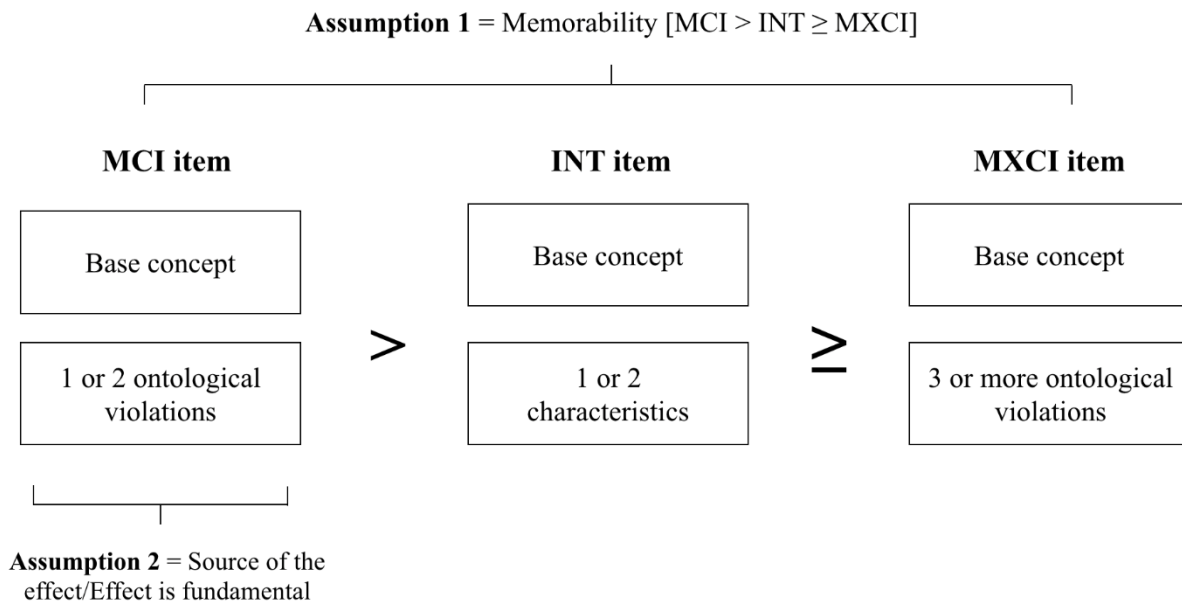


Figure 2: Two key assumptions in the MCI literature.

MCI effect is fundamental rather than epiphenomenal.⁶ That is to say, the violation of one or a

⁵ See Upal (2010) for an explicit rejection of that assumption. The fact that the assumption had to be explicitly rejected as part of a set of predictions made by the context-based account developed by Upal and collaborators attests to the widespread nature of that assumption.

⁶ While Boyer (2001) argues that there is indeed something special about ontological violations, Upal et al. (2007) and Purzycki & Willard (2015) take issue with this conclusion.

small number of ontological assumptions is the only thing that gives rise to enhanced memorability and that it does so above and beyond the contribution of any other variable that would tend to have a similar effect, including VR (Figure 2).

These two assumptions are likely responsible for the heterogeneous stimuli used in previous studies as well as for the failure to control for unusualness and inferential potential. If all MCI items are equal, perhaps because all violations of intuitive theories are considered equivalent, there is no reason to take special care when designing stimuli. At the same time, if violations of intuitive ontologies are special and produce superior memorability, they should swamp all other factors. Unfortunately, these assumptions may have led to undesirable variability across studies, which in turn yielded a distorted view of empirical findings.

For example, under these two assumptions, it would indeed be puzzling to find that MCI items aren't remembered better than INT items, or even worse, that the latter are remembered better than the former. However, these two assumptions are claims that have, to the best of our knowledge, never been systematically assessed. If they do not, in fact, hold, then without controlling for memorability within the class of MCI items itself, comparing different types of MCI items to each other and to different types of INT items across different studies may be the experimental equivalent of comparing apples and oranges.

3.3. Is the MCI effect epiphenomenal?

Following Boyer (2001, 2002) and Boyer & Ramble (2001), much of the literature on the memorability of supernatural concepts has proceeded on the assumption that there is indeed

something special about violations of ontological assumptions. For example, Boyer and Ramble (2001) explain finding that:

Violations of *ontological* expectations - as found in the template for supernatural concepts - are recalled better than what we called “mere oddities”. For instance, “a man who walked through a wall” (ontological violation) was generally better recalled than “a man with six fingers” (violation of expectations, but not of those expectations that define the ontological category PERSON).

This claim can be interpreted in at least two ways. A strong interpretation would be that ontological violations create a memorability effect that is *different in kind* from a typical von Restorff, or bizarreness, effect. The weaker interpretation would be that ontological violations, compared to ‘mere oddities’, simply give rise to a *more pronounced* bizarreness effect. Note as well that it is possible that violations of ontology might produce improved recall for reasons that have nothing to do with salience, per se. Perhaps ontological violations are more likely to produce inferential potential, which is the true driver of improved memorability in these concepts. The important point for our purposes, however, is that the memorability difference between ontological violations and other violations of expectations has not been conclusively settled (see discussion in section 3.1. and Purzycki & Willard, 2015 for a similar conclusion). This in part stems from a commitment to the assumptions that all MCI items are equal, and that the MCI effect is fundamental, which has led to a failure to control for other variables that might influence the memorability of these concepts. Therefore, we do not know if the MCI effect is different in magnitude or in kind from the von Restorff effect.

In light of these considerations, we propose in the following sections an alternative account of the MCI effect that relaxes the commitment to these two assumptions. What remains is a form of epiphenomenalism that relies on independently-motivated cognitive mechanisms, is empirically testable, and provides a natural way to make sense of the apparently contradictory nature of the extant literature on the MCI effect. Specifically, we propose that in addition to the

VR effect, two independently motivated variables, inferential potential and coherability, may contribute toward, or even entirely give rise to the MCI effect.

3.3.1. Inferential potential

Our first proposed variable, the notion of “inferential potential,” has long been theorized to play a role in the success of religious concepts (Boyer, 2002; see also Gregory & Barrett, 2009, and Purzycki & Willard, 2015), but has rarely been measured or controlled for.⁷ Inferential potential (IP) loosely refers to a concept’s ability to “readily generate inferences, explanations, and predictions with little effort” (Barrett, Burdett, & Porter, 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). Consider for example the difference between an invisible potato and a talking potato (Barrett, 2008b) or the difference between a potato that has no spatial location and a potato that eats people (Gregory & Barrett, 2009). In both cases, the latter potato seems more interesting, more likely to generate predictions, and perhaps more memorable than the former.

IP seems to play a role in real world supernatural beliefs as well. Boyer (1999) notes that ghosts and zombies are “symmetrical” CI concepts as one is a body with no mind and the other a mind with no body. However, ghosts and spirits are much more common across cultures than zombies because minds allow inferences to be made about desires and goals. In fact, when zombies are a cultural belief, they are often under the control of an agent, presumably to gain this very property.

⁷ See Barrett (2008b) for an attempt to define inferential potential in terms of the notions of Intentional Agency, Strategic Knowledge, Acting in the World, and Motivating Reinforcing Rituals.

Empirical evidence that IP has an effect on memorability is provided by Gregory and Barrett (2009) who controlled for IP by having participants rate items on a five-point scale from ‘not very thought provoking’ to ‘very thought provoking’ (along with a similar measure of imageability). They then selected MCI items with roughly equivalent levels of IP. Though their study only found a recall advantage for MCI items in a group of younger participants, there was a significant correlation between IP ratings and memorability even within the limited range of items chosen as stimuli (i.e., median IP scores between 2.5 and 3.5 on a 5-point scale).

If IP is one of the variables that gives rise to the memory advantage of MCI items,⁸ we may be in a position to better understand the apparently contradictory nature of the literature discussed earlier. Indeed, variation in level of IP between MCI and INT concepts could lead to a situation where the former are more memorable than the latter, but also the reverse. This situation could arise if a given MCI concept is low in IP and the INT item to which it is compared is relatively higher in IP. If the tacit assumptions of a necessary superiority for ontological violations don’t hold, then in this case, the INT item would be expected to be more memorable than the MCI item, in apparent violation of the original effect (Figure 3).

⁸ While IP has been proposed as a factor that affects the memorability of concepts, little if anything has been said about the mechanisms involved.

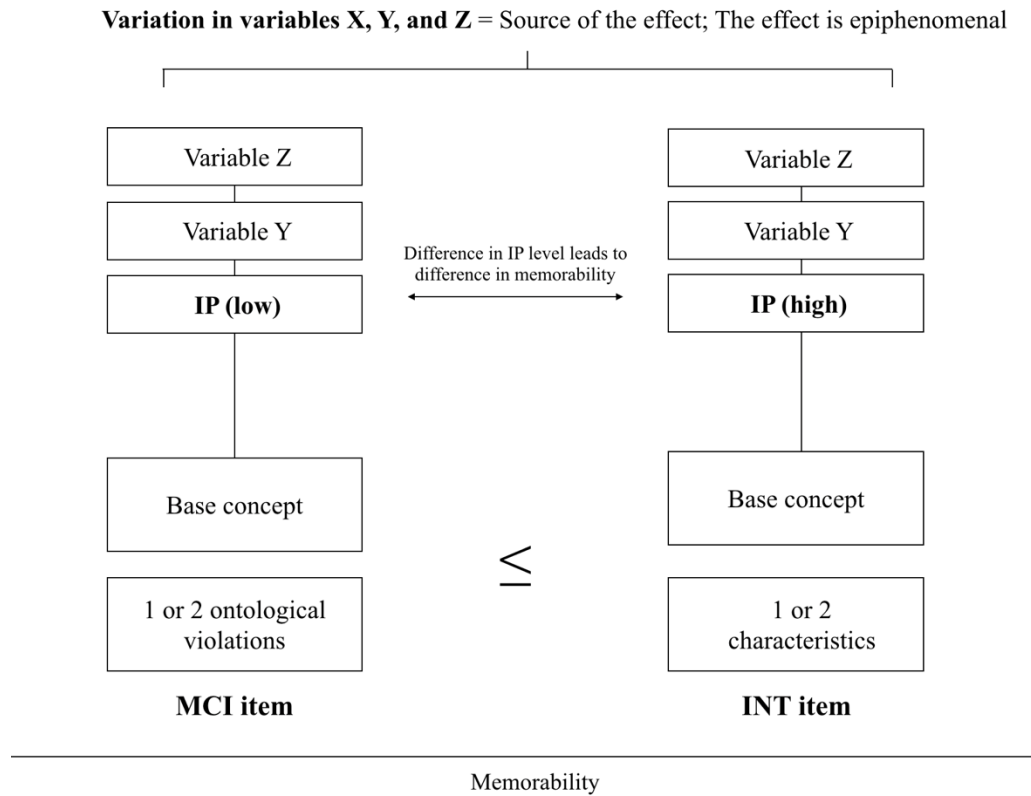


Figure 3: Variation in IP level between an MCI and INT item can lead to a situation where the former is more memorable than the latter.

Characteristics that raise IP, for example the ability to see the future, should make for especially memorable MCI items. Indeed, one could easily generate inferences, explanations, and predictions about a person who can see the future. However, a person who can see the future but immediately forgets about it, while an MCI in a technical sense, would have lower IP, and would therefore also be less memorable (see Boyer, 2002, for relevant discussion). Consequently, if low-IP MCI items are compared to high-IP INT items, it would not be surprising to see the original MCI effect reverse, as has indeed been the case in the literature (Gregory & Barrett, 2009; Norenzayan & Atran, 2004; Porubanova-Norquist, Shaw, & Xygalatas, 2014).

Many examples of low-IP supernatural characteristics in the MCI literature have the kind of structure illustrated in the example above where a single characteristic adds a useful property which is then cancelled out by the addition of a second property such as ‘seeing the future but forgetting it immediately.’ In a similar vein, Boyer (2002) offers examples of poor candidates for religious concepts which include an all-powerful God who ‘exists only on Wednesdays.’ Additionally, low-IP supernatural concepts may be readily created by Boyer and Ramble’s (2001) BrTr⁹ method of creating MXCI stimuli, where instead of adding multiple violations to the same concept, a single complex violation is added, which produces items that are less likely to be recalled despite possessing only one characteristic. For example, ‘a rock that only remembers what does not happen’ takes a breached characteristic of memory and applies it to an object that cannot remember at all in the real world. Other characteristics that yield low IP are simpler in structure but yield few useable inferences, e.g., ‘a cat that never behaves in a goal-oriented manner.’ High-IP characteristics may also be given to objects that cannot utilize them toward any recognizable goal, such as ‘a rock that can read minds’ or ‘a leaf that knows everything that has ever happened,’ which removes much of the characteristics’ utility and produces a low-IP concept.

This variation in IP across MCI items violates the assumption that all MCI items are created equal and suggests that memorability advantage of supernatural concepts may be epiphenomenal. However, the lack of a precise definition makes it difficult to assess the impact

⁹ A “breached transfer.” Recall that, in the MCI literature, a *transfer* takes a characteristic of one item and applies to an item that cannot have that property in the real world, e.g., a tree that thinks. A *breach* occurs when the target property is one that no real creature or object possesses in the natural world, e.g., a person who can see the future. A *breached transfer* takes a breached characteristic and applies it to a concept that does not normally have the original characteristic. For example, Boyer notes that many Catholics believe that certain statues can listen to prayers and that God can hear people anywhere in the world. The statue has been transferred the property of hearing, while God possesses a breached ability to hear from any distance. However, these violations are not combined to make a statue that can listen to prayers uttered anywhere on Earth (Boyer & Ramble, 2001).

of this dimension (if it is indeed a single dimension) on memorability. While further work will be needed to elucidate the components of IP, we suggest a few potential variables that may play a role.

First, Gregory and Barrett's (2009) findings that items rated more highly on a scale comprising thought provokingness and imageability suggests that these dimensions are good candidates for drivers of IP. A candidate mechanism underlying thought provokingness may be increased depth of processing, which is well-established as a means of improving memory (e.g., Craik & Tulving, 1975). Anderson and Reder (1979) argue that the advantage gained by depth of processing is "a result of the number of elaborations subjects produce while studying the material, that these elaborations establish more redundant codings of the to-be-remembered information, and that elaboration is what is critical, especially for long-term retention" (p. 385). Compare this to Barrett, Burdett, & Porter (2009)'s definition of IP as the ability to "readily generate inferences, explanations, and predictions with little effort". On the other hand, imageability is commonly associated with Paivio's (1986) "dual coding" theory, where information that is processed visually as well as verbally has a memorability advantage.

An additional, independently motivated variable, which is likely to be implicated in IP, is the notion of agency. An agent here refers to an entity that one naturally assumes has a mind, like a person or an animal, and whose behavior can therefore be accounted for in terms of beliefs, goals, and desires (Dennett, 1987; Pinker, 1997). We know from the literature on memory that agents are more memorable than non-agents (e.g., inanimate objects) (Nairne, et al. 2013; Nairne, VanArsdall, & Cogdill, 2017; VanArsdall, et al. 2015). Improved memorability for agents has also been demonstrated for MCI items, in particular, with agentic MCI items outperforming supernatural artifacts and plants (Porubanova-Norquist, Shaw, & Xygalatas, 2014;

Porubanova, et al. 2014). Moreover, corpus analyses have found that up to 98% of counterintuitive items in folktales are agents (Barrett, Burdett & Porter, 2009). This memorability advantage for agents may be a byproduct of more general mnemonic advantages for evolutionarily useful content that is relevant to survival (Nairne & Pandeirada, 2016).

Interestingly, the process of creating MCI items, by violating ontological assumptions, may probabilistically increase the likelihood that the resulting concepts, compared to INT items, will have more agency. In the MCI literature, counterintuitive characteristics typically result from violations of a small number of core intuitive systems, namely intuitive physics, biology, and psychology. On this scheme, transfers of psychological properties to a natural object or an artifact will often endow the resulting item with a certain amount of agency that it did not originally possess. Thus, ‘a hammer that can talk’, an MCI concept, will be more agentic than ‘a hammer that has a very sturdy handle’, an INT item. The important point here is that the enhanced memorability of the hammer that can talk compared to the hammer with the sturdy handle may be due in part to the difference in agency between the two items, and not to the special transfer between intuitive ontologies.¹⁰ Likewise, MCI items that differ in agency would also be expected to yield different memorability profiles.

So far, we have described how IP is implicated in the memory difference among MCI items as well as between MCI and INT items. We explained how variation in IP could give rise to the original MCI effect, presumably in combination with the VR effect. We argued that such variation could also give rise to the opposite effect, with INT items being as memorable, or even

¹⁰ This problem is likely compounded by a reluctance to violate a single intuitive theory more than once per concept (See Barrett, 2008a). Often the only ontological categories being violated are intuitive physics, biology, and psychology. As the number of violations increases, if violations are generated from intuitive theories without replacement, this leads to a 100% chance of a psychology violation in items with 3 characteristics. On the other hand, if violations of psychology usually produce agency and these violations increase with additional characteristics, we might predict a rise in memorability for MXCI concepts on the grounds that there are more agents among them, but this is the opposite of what MCI theory predicts.

more memorable, than MCI items when the former are endowed with higher levels of IP. Recall though that Boyer's original account also predicts a decrease in memorability for MXCI items (see figure 1). Here, however, it is difficult to see how the VR effect or IP could account for the special properties of MXCI items. To be sure, the VR effect would seem to predict even greater memorability for MXCI items as their distinctiveness is made more pronounced by additional counterintuitive characteristics. Likewise, IP does not appear to be very promising as it is unlikely to systematically decrease when new characteristics are added.¹¹

3.3.2. Coherability

Our second proposed variable is a new concept that we call *coherability*. On Boyer's account, the decreased memorability of MXCI items, compared to MCI concepts, is due to the increased complexity associated with repeated ontological violations. Specifically, Boyer and Ramble (2001) propose that in MXCI items, inferences associated with the base concept are blocked:

In this view, more bizarre material may result in decreased recall because it makes it less likely to activate the relevant schema. However, this operational understanding of "discrepancy" may denote very different phenomena, depending on the kind of "schema" activated. In the present case, what makes Breach+Transfer items different from other types is not that the relevant domain-level expectations were not activated, but that inferences on the basis of these expectations *were blocked* [our emphasis] from Boyer and Ramble, 2001, p. 549.

¹¹ It is worth noting, though, that when researchers give examples of MCI and MXCI items, they often conflate IP and number of violations. This leads to MCI examples such as 'a man who can hear people's thoughts' (1 violation and high IP) but MXCI examples such as 'a rock that can see the future but forgets it and is invisible and intangible and only exists on Tuesdays.' (many violations and also all extremely low IP). This might be indicative of a wider bias in designing MXCI items, but it could just be a tendency to highlight a particularly poor MXCI item when introducing MCI theory.

These inferences, which are crucial for representing the concept, come from the intuitive expectations about the base concepts that are not violated.¹² There is, however, another factor that systematically varies when MXCI items are created but that remains roughly constant for MCI and INT items. When choosing characteristics to add to INT items, one is necessarily constrained by the set of features that people expect such items to have in the natural world. For example, when selecting characteristics for an INT lizard, as in the Banerjee et al. (2013) study, features like greenness, eating insects, being cold-blooded, and so on, easily come to mind. In other words, these properties are compatible with people's prior expectations regarding the base item; in this case, a lizard. To the extent that a set of properties for a given base item is easily associated with members of the relevant category, they can be said to *cohere* with the item and with each other. Thus, greenness, the property of eating insects and being cold-blooded, are all connected to what it means to be a lizard.

Notice now that when an MCI item is created, some amount of *coherability* is lost, as for example in the case of a lizard that is green and can never die. Here, the property of immortality does not cohere well with the lizard or with the property of being green because it is not part of people's prior expectations about lizards and these two characteristics are not often associated with each other. However, as long as the number of ontological violations remains small, as in the case of MCI items, we get the optimal tradeoff between salience and complexity discussed earlier. On this account, the addition of more violations of intuitive ontologies to create MXCI items will naturally lead to even lower levels of coherability. Consider, for example, one of the MXCI items from Banerjee et al.'s (2013) study, namely a lizard that "always melted in the hot sun, could never die no matter how old it was, and could hear other creatures' thoughts". Here,

¹² "*Inferences are governed by background default expectations.* Those intuitive expectations that are *not* violated are the main source of inferences about supernatural situations." From Boyer, 2000, p. 200.

none of the properties associated with the base noun conforms with people's prior expectations about lizards. Moreover, the properties themselves are drawn from unrelated semantic fields and therefore cannot be used to prime each other's recall. In this case then, the set of properties have nothing to do with lizards and they have nothing to do with one another. They do not cohere well at all and memorability is correspondingly affected.¹³

This account of the loss of memorability associated with MXCI items is similar to Boyer's original proposal in how it addresses the empirical question of memorability. In both cases, as ontological violations pile up, the resulting complexity leads to loss of memorability. However, there are important differences between the two accounts in terms of the theoretical question of the underlying mechanisms. The first difference has to do with the source of the complexity associated with multiple ontological violations. On Boyer's account, too many ontological violations negatively affect the inferential structure of the base concept which, in turn, leads to decreased memorability. On the account developed here, the increased complexity associated with MXCI items stems from the increasing lack of coherability among the counterintuitive characteristics and between those characteristics and the base concept. Because each of the characteristics cannot be predicted based on prior expectations, the counterintuitive characteristics need to be separately encoded and recalled which increases the burden placed on the memory system and leads to decreased memorability.

¹³ Coherability may be similar to Barrett's (2008a) "simplicity principle." This principle suggests that when multiple properties from one ontological category are transferred to a second, such as when a rock is given the properties of human thought and biological systems, it may be more easily conceptualized as a human in the form of a rock. Likewise, when transferring multiple properties from the same ontological category, as in a tree that can speak, reason, and feel emotions, people may default to assuming the entire ontological category of human psychology has been transferred, rather than representing each additional violation as a novel characteristic. The fact that Barrett sees these supernatural concepts with highly coherable characteristics as qualitatively different from MXCI items demonstrates the baseline poverty of coherability that MXCI items are expected to possess.

The second difference stems from the differential impact of accumulated violations. According to Boyer, violating an intuitive theory blocks inferences from that intuitive ontology. If too many inferential routes are blocked, it becomes too difficult to reason about the underlying concept. However, the coherability account makes no such prediction and would permit concepts with multiple violations to be represented easily, as long as they have relatively high coherability. The coherability view also allows memorability to vary among MXCI concepts. If violations block inferences entirely, then all MXCI concepts with the same number of violations should be equally difficult to represent, at least if the same intuitive theories are blocked in each of the concepts. Coherability, on the other hand, can vary between concepts with the same number of violations, and this variability is expected to yield differences in memorability.

An important virtue of the coherability-based account is that it makes testable empirical predictions. The first step would be to measure coherability. Participants could, for example, be asked to rate the extent to which characteristics fit with each other or a given base concept. This could be done for INT, MCI, and MXCI items alike. These coherability norms could then be manipulated experimentally with the expectation that, everything else being equal, decreased levels of coherability would correspondingly lead to decreased memorability.

3.4. Reassessing the empirical challenge

We started this section with an empirical puzzle. Why is it that the literature on the MCI effect contains such a disparate, and apparently contradictory, set of findings? In response to this muddled picture, there are at least two alternatives. The first would be to take the empirical contradictions at face value and conclude that the MCI effect is illusory, perhaps another

instance of a psychological finding that failed to replicate. This conclusion is unlikely, however. The effect has already been replicated by a number of investigators in a range of cultural contexts and in different populations. Moreover, there are very good theoretical and independently motivated reasons to expect that such an effect should obtain. If so, what needs to be explained is why different studies ostensibly designed to replicate the MCI effect yield such a vexed pattern. This question, in turn, has led us to reexamine core assumptions at the heart of the MCI literature. The first assumption is that all MCI items are created equal. We have shown that this is unlikely. There are differences in IP between MCI items which have been found to be correlated with memorability (Gregory & Barrett, 2009). A similar situation holds for differences in unusualness, though the relationship with memory is fuzzier, perhaps because IP was not controlled for in these studies (Barrett & Nyhoff, 2001; Porubanov, et al., 2013, 2014). Differences among MCI items in turn challenge the assumption that the MCI effect is fundamental rather than epiphenomenal. Put another way, these considerations call into question the idea that violations of ontological assumptions yield a qualitatively different salience effect, or at least a more pronounced one, compared to other forms of counterintuitiveness discussed, for example, in the literature on the von Restorff effect.

These considerations have led us to propose a new account of the memorability profile of INT, MCI, and MXCI items which regards the MCI effect as epiphenomenal. Indeed, this possibility, initially dismissed by Boyer (2002), has not in fact been considered as carefully as it should have (Barrett, 2008a; see Purzycki & Willard, 2015). Specifically, we have proposed that the enhanced memorability of MCI concepts, compared to INT items, could derive from the effects of independently-motivated memory boosting variables such as the von Restorff effect and inferential potential. Moreover, we introduced the notion of coherability, a general property

of the relationship between a base concept and the set of characteristics used to modify it, and proposed that it could explain the decreased memorability of MXCI items. Since these variables have remained unrecognized, or at least not controlled for, in the MCI literature, it is not difficult to imagine why different studies would yield different, even apparently contradictory, results.

4. Theoretical challenges to the MCI account

We now turn to theoretical challenges that have emerged since the publication of Boyer's MCI thesis and consider the merits of the three main contenders available in the literature. The first is a memorability model proposed by Upal et al. (2007) which derives the MCI effect from properties of the context in which the items occur. This account negates the role of the theoretical mechanism proposed by Boyer and regards the MCI effect as epiphenomenal. Next, we consider a challenge by Purzycki & Willard (2015) who argue that the very notion of counterintuiveness has been ill-defined in the literature and therefore that no conclusive evidence supporting the validity of Boyer's MCI account can be found. We then consider a proposal by Norenzayan & Atran (2004) according to which the MCI effect operates at the level of belief-sets rather than individual beliefs, as originally proposed by Boyer. Yet another challenge notes the insufficiency of the MCI thesis to explain belief considering extant concepts like Mickey Mouse and Zeus which have the MCI structure, but are not believed in. We show that while each of these challenges has merit, they do not irreparably undermine the MCI thesis, which, in the end, remains a live option. Finally, we turn to a critique of the MCI account by Franks (2003) according to which Boyer's original account underestimates the complexity of the conceptual combinations that give rise to supernatural concepts. While Franks' ideas are no doubt

interesting, at worst they point to the conclusion that the MCI account is incomplete. This conclusion, in turn, has little bearing on the memorability issue at the heart of the MCI account.

4.1. The MCI effect as a contextual epiphenomenon

In an interesting departure from Boyer's original account, Upal et al. (2007) propose to derive the enhanced memorability associated with MCI concepts by shifting the explanatory focus from properties of the items themselves (e.g., a lizard that never dies no matter how old it is) to properties of the context in which these items are embedded. This new, context-based approach relies on two key notions that the authors call *predictability* and *postdictability*. Given the occurrence of a particular item or concept in a narrative, predictability refers to "the ease with which the occurrence of the concept can be predicted prior to the concept having been read" (Upal et al., 2007, p. 419). On the other hand, postdictability refers to the "ease with which a concept's inclusion in the text can be justified after the textual unit containing the concept has been read" (p. 419).

To illustrate these two notions, consider the following example adapted from Upal et al.

Odd News of the Day

Bowling Green (KY): July 18, 2004. *"I had just woken up and went to the kitchen to prepare some coffee to drink" said the Kentucky farmer Edwin Smith. "That's when I saw **the cow flying above the trees**. The twister had lifted the 500 pound creature well over 50 feet above the ground and was rotating it around like a doll. That was a scary experience", said Smith.*

In this short narrative, the concept to be recalled is 'the cow flying above the trees.' The italicized text in the passage above is the prior context for that concept and the text that follows the concept is the posterior context. Given the prior context, encountering a concept like 'the cow flying above the trees' is not something that someone would easily predict. In this case, the concept's predictability is therefore low. By contrast, the concept's postdictability is higher.

Indeed, mention of a twister in the posterior context allows the reader to easily justify to herself having read about a cow flying above the trees. Upal et al. (2007) propose an ideal agent model of memorability where predictability is inversely related to memorability such that the more predictable a concept is in a given context, the less memorable it should be (because it can be generated from prior knowledge). When predictability is low, but postdictability is also low, the information or event is unlikely to be worth the effort of storing in memory, as it represents a strange and probably rare occurrence. Crucially, though, when predictability is low and postdictability is high, the agent stands to gain a lot of information by committing the event to memory. Because the event has high postdictability but was not predicted, it suggests a mismatch between the agent's knowledge and the world, which should be corrected.

Upal's tradeoff between predictability and postdictability ostensibly yields exactly the memorability profile that one would expect for INT, MCI, and MXCI concepts under the original MCI account. INT concepts are high in both predictability and postdictability, making them entirely unsurprising to the agent and thus uninformative. By contrast, MXCI items are low in both predictability and postdictability. Here the potential information gain resulting from low predictability is cancelled out by the low level of postdictability. Finally, MCI concepts are expected to be more memorable than INT and MXCI items because they are low in predictability, thereby representing an opportunity for updating the agent's model of the world, but unlike MXCI items, they are more postdictable (Figure 4), which means the effort to update the model is likely to be worthwhile.

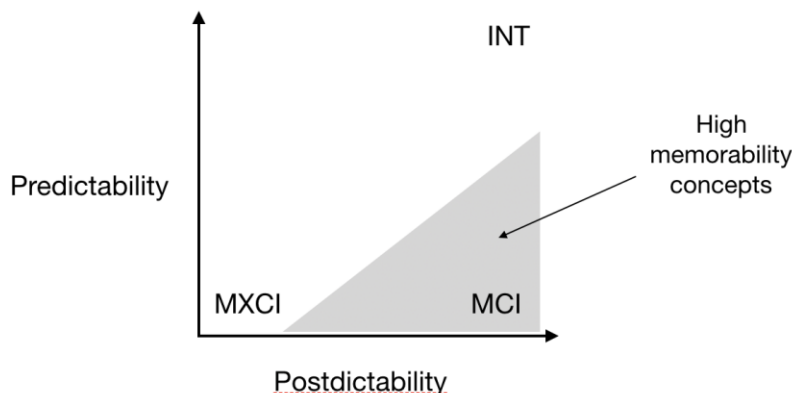


Figure 4:

nature of MCI

the optimal
concepts for

memorability

Thus, Upal et al.'s account predicts that MCI concepts should be more memorable than INT or MXCI ones. Furthermore, their account makes a number of interesting empirical predictions, namely that manipulating the context would entail that:

1. If a concept is made more predictable, it should become less memorable.
2. If a concept is made less predictable, it should become more memorable.
3. If a concept is made more postdictable, it should become more memorable.
4. If a concept is made less postdictable, it should become less memorable.

Upal et al. provide experimental evidence for their memorability hypothesis by testing postdictability values for INT, MCI, and MXCI items. They report that these values differ precisely as expected by their model, with $(MXCI) \text{ postdictability} < (MCI) \text{ postdictability} < (INT) \text{ postdictability}$. Moreover, they show that experimentally induced changes in context impact the memorability of concepts as predicted in (1-4). Upal et al. conclude that “there is nothing inherent or magical about minimally counterintuitive concepts. Rather than a property of the item itself, as most previous researchers have assumed, it is the properties of the item in a particular context that matter” (p.422).

On this view, the MCI effect is epiphenomenal in yet a different sense from the one that we introduced in section 3.3. For Upal et al., regardless of whether an item is INT, MCI, or MXCI, its memorability profile will be determined by its combined level of predictability and postdictability in a given context. As Upal (2010) explains, “There is no sharp boundary between INT, MCI, and MXCI concepts in terms of memorability.” What this means is that in certain contexts, it would not be surprising to find that INT or MXCI are more memorable than MCI concepts. This approach does pose a serious challenge to the theoretical mechanism that Boyer originally proposed in that it has the potential to explain away the MCI effect without attributing any privileged status to ontological violations as a source of enhanced memorability. On this view, Boyer’s MCI effect would be epiphenomenal but also variable in the sense that it would only emerge under some contextual conditions, but not others (both of which can be specified). Another interesting feature of Upal and colleagues’ contextual view is that it may also help explain the apparently contradictory nature of the empirical findings reported in the MCI literature (see section 2 and also Norenzayan & Atran (2004) for related discussion).

Until further empirical tests are carried out, however, Upal, et al.’s conclusion that there is “nothing inherent” about MCI concepts might be too strong. Without denying the effect of context, it remains reasonable to expect, at least in principle, that INT, MCI, and MXCI items give rise to different memorability profiles if the context is held constant. This question becomes all the more relevant if we adopt the view developed in section 3.3. on which the MCI effect arises as variables such as inferential potential and the von Restorff effect interact. These variables are properties of the concepts themselves and not of the context. If so, and if these properties affect memorability, this leaves a role for the concepts themselves to play in combination with the context.

Moreover, Upal's suggestion that MCI items are at a cognitive optimum between INT and MXCI items may be premature. Granted, MCI items are less predictable than INT items and more postdictable than MXCI items with 3+ characteristics. However, those same MXCI items would be less predictable than MCI items and more postdictable than an item with 100 characteristics. In other words, for any number of characteristics added to a noun, the resulting concept is less predictable than it would be if fewer characteristics had been selected. It would similarly be more postdictable than if it possessed additional characteristics. For example, imagine a concept with 5 characteristics. This concept is likely to be less predictable than a concept with only 3 characteristics simply because there is more to predict. Additionally, it is more easily postdictable than a concept with 7 characteristics because there is less to justify. Notice that this 5-characteristic concept exists in a middle ground between (more) predictable items (with fewer characteristics) and less postdictable items with even more characteristics. This is precisely what Upal claims creates the cognitive optimum for MCI items. That is, that they are less predictable than INT items but more postdictable than MXCI items.

Now, one might reasonably claim that a traditional MXCI item, with 3 characteristics, would be better remembered than a 7 or even 100+ characteristic items. However, it would be a mistake to claim that this intermediate status represents a cognitive optimum - we know that these concepts are not the cognitive optimum because they are remembered poorly relative to MCI items. In the same vein, why assume that MCI items are a cognitive optimum just because they sit between the categories of INT and MCI? It is possible that the true optimal tradeoff between predictability and postdictability occurs when concepts have BIZ characteristics, or even INT characteristics with low coherability, such as a lion that was born on Christmas, with

MCI items possessing very poor postdictability scores. That MCI items should be a cognitive optimum is not so easily derived from Upal's theory.

4.2. The inherent vagueness of counterintuitiveness

Purzycki & Willard (2015) offer a critique of the MCI account that calls into question the scope and usefulness of Boyer's approach as well as the validity of its central tenets. They begin by identifying the four assumptions underlying the MCI thesis, namely:

- (1) The human mind contains a set of innate inferential systems that allow us to apprehend the world around us.
- (2) Concepts that violate a small number of those inferences (i.e., MCI concepts) are more memorable than their INT and MXCI counterparts.
- (3) MCI concepts are a central feature of religious traditions.
- (4) The ubiquity of religious concepts is in part due to the fact that MCI items enjoy a memorability advantage.

In their assessment of the literature, Purzycki & Willard (2015) conclude that empirical tests of the MCI thesis have, for the most part, focused on addressing (2) while taking (1) and (3) for granted. To the extent that the relevant evidence supports (2) then, the validity of (4) is thereby established. The main thrust of Purzycki & Willard's (2015) critique stems from the observation that the relationship between (1) and (2) is far less obvious than has been assumed and therefore, that the very concept of an MCI item remains ill-defined. Consequently, empirical tests of the memorability of MCI concepts have relied on markedly different operationalizations of the notion of counterintuitiveness, leading to the kind of muddled empirical picture that we described earlier.

To bring these issues into sharper focus, Purzycki & Willard (2015) distinguish between what they call deep vs. shallow inferences. The former corresponds to Boyer's ontological

violations and they are grounded in implicit knowledge stemming from core cognitive faculties of the kind revealed by developmental psychologists (e.g., intuitive physics, biology, psychology). The latter, by contrast, include inferences between concepts that are, for the most part, explicitly accessible and that need not necessarily be consistent with deep inferences. To illustrate, consider the concept of a dog that can walk through walls. Developmental psychologists have demonstrated experimentally that infants expect physical objects to be solid and to not be able to pass through each other (Baillargeon, 1998; Carey & Spelke, 1994; Leslie, 1982). Thus, a dog that can walk through walls violates a core principle of intuitive physics and therefore a deep inference. On the other hand, the expectation that dogs salivate at the sight of a bone would be regarded as a shallow inference because it isn't part of our intuitive conception of what it means to be an animal or a living organism.

With that distinction in mind, Purzycki & Willard (2015) point out that the literature on the MCI thesis is mired because the notions of deep and shallow inferences have all too often been conflated resulting in the creation of MCI items that cannot be meaningfully compared to one another (see Barrett, 2008a for a similar concern). Moreover, even MCI items that seem to violate deep inferences, as for example “a virgin mother”, “a crying mailbox” (from Banerjee et al., 2013), and “a table that breathes” (from Boyer, 2000), are suspicious because, according to Purzycki & Willard (2015), “we have no evidence that *breathing*, *crying*, or *sex makes babies* are default inferences of ANIMALs, “folk biology”, or “the domain of living things”.” (p.20).

Finally, Purzycki & Willard observe that disembodied minds such as ghosts, spirits, and ancestors have been argued to be deeply *intuitive*, perhaps because we are intuitive dualists, and that this is precisely why these concepts are ubiquitous in religious traditions (Barrett, 2004;

Bering, 2011; Bloom, 2007 among others).¹⁴ An important upshot of these considerations is that the repeated conflation of deep and shallow inferences suggests that the MCI effect has never really been teased apart from other well-known bizarreness effects such as the von Restorff effect; precisely the conclusion that we reached, albeit on different grounds, in section 3.3.

We certainly agree with Purzycki & Willard (2015) that there is confusion in the literature regarding what should count as an MCI concept both in terms of the nature of the violations involved as well as the number of such violations (Barrett, 2008a). We also agree that the distinction between deep and shallow inferences can sometimes be subtle or difficult to make. For this reason, it is possible that Boyer's original claim that violations of intuitive ontological assumptions produce superior memorability has never truly been tested. However, on the account we develop here, there is no need to distinguish between shallow and deep inferences. If indeed MCI concepts' memorability is due to the VR effect, IP, and relatively high coherability, then shallow inferences are all that need to be violated to produce the effect. Nonetheless, future work should attempt to carefully assess memorability differences produced by violations of shallow and deep inferences, while controlling for factors such as IP and unusualness, to determine whether the MCI effect is fundamental or ephiphenomenal.

It is also worth pointing out, regarding assumption (3) above, that it hasn't been merely taken for granted in the literature on the MCI thesis, as Purzycki & Willard's (2015) suggest. Indeed, corpus analyses of religious traditions and folktales have confirmed the centrality of MCI concepts (Gibbon, 2008; Norenzayan, Atran, Faulkner & Schaller, 2006; Barrett, Burdett &

¹⁴ This may be the result of two definitions of "counterintuitive" in this context. The first, more colloquial, definition is simply that some things are easy to grasp. Religious concepts in this sense, may not be counterintuitive at all. The second definition is Boyer's notion of 'countering' an intuitive theory. On this definition, supernatural and religious concepts are axiomatically counterintuitive, regardless of how easily they are represented in the mind or how possibly real they seem.

Porter, 2009). In the end, we take Purzycki & Willard's (2015) critique of the MCI thesis to be a call for more theoretically and methodologically precise work on the MCI thesis rather than a clarion announcing its demise, as some investigators have concluded (Guthrie, 2015).

4.3. The MCI effect operates over beliefs sets rather than beliefs

Norenzayan and Atran (2004) offer a critique of Boyer's MCI thesis which begins with a puzzle. If MCI concepts enjoy the kind of memorability advantage predicted by Boyer's account, why don't they dominate myths, religions, and other popular tales? As Norenzayan and Atran point out, the Bible and the Grimm Brothers' folktales, for example, contain a succession of ordinary concepts and events peppered with only a few counterintuitive elements.¹⁵ Moreover, Norenzayan and Atran call into question the validity of the conclusions drawn from the few studies which, at the time, were taken to provide experimental support for the privileged memorability status of MCI concepts over intuitive ones (Boyer and Ramble, 2001; Barrett and Nyhof, 2001). Specifically, Norenzayan and Atran argue that contextual effects, rather than item-level properties, may have given rise to the results obtained by Boyer and Ramble (2001). For example, the fact that the stories are about an intergalactic ambassador may have biased participants to preferentially attend to and recall science-fiction like, counterintuitive items. The full force of these considerations was revealed by Upal et al.'s (2007) context-based model of memorability discussed in section 4.1.

¹⁵ For example, Norenzayan and Atran point out that Little Red Riding Hood, one best-known tale from the Grimm brothers, contains exactly two counterintuitive items, namely a talking wolf and the fact that the little girl and the grandmother come out of the wolf's belly alive.

As an alternative, Norenzayan and Atran (2004) propose that the kind of cognitive optimality that lies at the heart of Boyer's MCI thesis may operate at the level of *beliefs sets* rather than the level of beliefs (i.e., individual concepts). In order to test their hypothesis, these authors presented sets of concepts with varying proportions of INT and MCI items to their participants. The items were presented in lists, to avoid potential contextual effects. The stimuli they used transferred an adjective from an appropriate context (e.g., "closing door," "thirsty cat") to an inappropriate one (i.e., "closing cat," "thirsty door"). Lists of items containing only INT concepts were compared to those containing mostly INT and a few CI concepts, equal numbers of both, and lists with few INT concepts and many CI ones. Recall was measured immediately and after a delay. MCI items did display an enhanced fitness over the delay, decaying more slowly than INT items, but, crucially, they were recalled less well than the INT items in each of the conditions except when the CI items were the majority, where item types were remembered at a similar rate. MCI belief sets, containing mostly INT items with a few MCI items, were remembered better and had a lower rate of memory decay compared to INT and MXCI belief sets.

The belief-set approach developed by Norenzayan and Atran (2004) certainly constitutes an interesting research avenue that should be further explored. We also sympathize with the concerns that these authors express about possible contextual effects affecting the results of earlier memorability studies, especially in light of the model developed by Upal et al. (2007). At the same time, we remain unconvinced that contextual effects are the whole story, as discussed earlier. We also believe that Norenzayan and Atran's (2004) puzzle about the prevalence of MCI concepts only has limited force. One reason why MCI concepts do not drown more mundane items in stories, religious traditions, and folktales is that regular people, animals, objects, and

artifacts abound and play a critical role in everyday life (Barrett & Nyhof, 2001). It is therefore not too surprising that they should also dominate the stories we tell each other. MCI items may indeed very well enjoy a memorability advantage, but reality also has a potent effect on our cognitive lives that we shouldn't dismiss too easily.

4.4. The problem of belief

Another challenge to Boyer's thesis stems from the observation that the MCI nature of a concept is not sufficient to yield believability, and therefore that the MCI account is at best an incomplete theory of supernatural beliefs. For example, The Mickey Mouse problem asks us to consider the fact that while Mickey is quite popular, almost no one believes he is real (Atran, 2002). Similarly, The Zeus problem highlights the fact that even if an MCI used to enjoy widespread believability, it may still come to fall into widespread disbelief (Gervais & Henrich, 2010). These problems would seem to imply that not only are mechanisms other than MCI structure required to yield believability, these mechanisms may not be properties of the concepts at all, with more nebulous, cultural factors driving belief in supernatural concepts.

Though the Mickey Mouse and Zeus problems show that possessing the MCI structure is insufficient for a concept to be believed in, Boyer's original account was never intended as a theory of belief formation. The MCI thesis was proposed, first and foremost, as an account of the enhanced *memorability* and *transmission* of supernatural concepts. The usual corollary to discussions of the MCI effect is that everything else being equal, concepts that are better remembered and transmitted will tend to be more believable (e.g, Boyer & Ramble, 2001). However, since the world is a complicated place, and that all else is rarely held constant, a

straightforward correspondence between memorability and believability is neither entailed by the MCI account nor likely to hold in the first place.

Furthermore, notice that with the proper scope of the MCI account in mind, the Mickey and Zeus problems become self-defeating. Despite the fact that people do not believe that Mickey Mouse is real and no longer believe in Zeus, these concepts remain frequently transmitted. Why should people be so interested in bizarre concepts that they do not even think exist? People continue to consume and transmit stories about MCI concepts be they members of ancient pantheons of gods, mythological creatures, or superheroes. The popularity of these concepts, even though no one believes they are real, is precisely what the narrow version of the MCI effect would predict. In other words, despite a lack of belief in these concepts, they are remembered and transmitted quite well. To be sure, it is critical to identify the mechanisms that do lead to belief in supernatural concepts (e.g., Willard, Hernich, & Norenzayan, 2016), but we do not view this gap as a failing of MCI theory.

4.5 The MCI account and the conceptual combination problem

Boyer's contention that supernatural concepts are modified versions of ordinary concepts is what we called the *ontological* question. Two aspects of this questions can be distinguished. The first regards the provenance of supernatural concepts; the idea that they are drawn from the set of ordinary concepts that human beings entertain. Thus, the supernatural concept of a ghost belongs to the natural category PERSON. The second aspect concerns the nature of the mechanisms that tranform natural concepts into supernatural ones. According to Boyer, this happens either via "breach" where a characteristic that doesn't apply to anything in the natural

world is added to a concept (e.g., the ability to see the future, which is not something that any natural entity possesses) or via “transfer” where a property of one category is transferred to a member of a different category (e.g., a statue, drawn from the category ARTIFACT, that can cry, a property associated with the category PERSON).

In his discussion of the MCI account, Franks (2003) observes that supernatural concepts, on Boyer’s scheme, are created via general procedures used for conceptual combination. To quote Franks:

These contradictions, breaches and so on only arise because, in forming religious representations, other, non-religious representations are combined, and the properties of those combined non-religious representations are in some way incompatible. That is, religious representations are a type of concept combination ... (p. 11)

However, Franks points out that given what is known about the mechanics of conceptual combination, the MCI account exploits only a subset of the combinatorial options available. This, in turn, glosses over a number of subtleties which may turn out to be important in understanding the cultural and psychological properties of supernatural concepts. To appreciate the complexity of conceptual combinations, let us briefly consider the main mechanisms available. The first, known as property mapping, takes two concepts, A and B, and creates a new concept, AB, in which the properties of A and B are unified. If, for example, A is the concept red, and B the concept apple, then a red apple, AB, has all the properties of an apple and all the properties of something that is red.

When the features of A and B are inconsistent, however, two strategies are possible. The first is negation where a property of one concept directly negates a property of the other concept. Examples, drawn from Franks’ (2003) discussion, involve *fake gun*, *invisible picture*, and *stone lion*. In such cases, the modifying adjective directly negates the core property, or essence, of the head noun. Hence, a fake gun isn’t a real gun, an invisible picture isn’t a real picture, and so on.

The second strategy to deal with inconsistencies between two concepts is what Franks calls *doubt*. Examples include *apparent friend* or *alleged criminal*. In such cases, the modifying adjective casts doubt on the property expressed by the head noun. Hence, we do not know whether an apparent friend is a real friend or not, nor do we know whether an alleged criminal actually committed a crime. Another broad strategy to deal with conceptual combinations, aside from property mapping, is what is known as relation mapping. This happens when the connection between two concepts, A and B, is established via a specific semantic relation. Examples include *soldier ant* which is an ant that IS a soldier, *daisy chain*, in which the daisies MAKE a chain, *tear gas*, in which gas CAUSES tears, and *tax law* in which the law is ABOUT tax.

Given the processes described above, Frank's criticism of the MCI account is that only conceptual combinations involving property mapping, and more specifically, negation, are taken into account. However, when there is an incompatibility between two concepts to be combined, the tension involved can be resolved via doubt or relation mapping, in addition to negation. To illustrate, Frank asks us to consider the example of bread and wine at a catholic mass and their relationship to the body and blood of Christ. Using property mapping, one can argue that the ideal essence of Christ has been transferred to the bread and wine. Alternatively, using relational mapping, one could imagine that the bread and the wine REPRESENT the body and blood of Christ. Finally, using doubt property mapping, one can cast doubt on the actual relationship between the bread and wine, on the one hand, and the body and blood of Christ on the other. The unresolved tension stemming from such doubt might contribute to the aura of mystery surrounding such religious notions.

In sum, Frank argues that the MCI account, because it relies on only a subset of the options available for conceptual combinations, isn't granular enough and might therefore be unnecessarily restricted in its explanatory power. He goes on to show that a more nuanced approach to conceptual combinations in the creation and interpretation of supernatural concepts may indeed be in a better position to shed light on important cultural questions pertaining to religious cognition. Frank may very well be right that as originally formulated, the MCI account is too restrictive and not nuanced enough. However, such criticism, even if valid, would seem to have little impact on the memorability of supernatural concepts created through "breach" and "transfer" of ontological assumptions. Other methods of conceptual combination notwithstanding, the question of whether supernatural concepts enjoy a memory advantage remains an interesting and open one to be settled through careful experimental investigation.

5. Putting the pieces of the MCI puzzle together

The cognitive turn in the study of religious thought and behavior within the past two decades or so, and in particular Boyer's pioneering MCI thesis, have given rise to exciting new developments and a rapidly growing literature. What has emerged from this body of work is a heterogenous set of conclusions which at times have seemed to confirm the predictions of the MCI thesis (Boyer & Ramble, 2001; Barrett & Nyhof, 2001; Johnson, Kelly, & Bishop, 2010; Banerjee, Haque, & Spelke, 2013) while on other occasions, the same literature has called into question virtually every aspect of Boyer's original account (Gregory & Barrett, 2009; Norenzayan & Atran, 2004; Porubanova-Norquist, Shaw, & Xygalatas, 2014; Harmon-Vukic &

Slone, 2009; Upal, et al., 2007; Upal, 2010; Purzycki and Willard, 2015). To make sense of this fractured picture and map the road ahead, we return to the ontological, empirical, and theoretical questions that we introduced at the beginning of this article. As we have shown, each of these has turned out to be more complicated than originally anticipated.

The ontological question asks about the nature of supernatural concepts and their relationship to other mental entities. Here, Boyer's position is that concepts such as gods, souls, and ancestors are in fact ordinary concepts (e.g., PERSON, ARTIFACT) that have been modified to give rise to their otherworldly qualities. Although the simple model of modification proposed by Boyer has been disputed (e.g., Franks, 2003), it seems reasonable to assume that at least some supernatural concepts are ontologically epiphenomenal.

Turning to the empirical question, namely the predicted superior memorability of MCI concepts compared to their INT and MXCI counterparts, experimental results diverge more sharply. Sometimes MCI concepts have indeed been found to be more memorable than INT and MXCI items and sometimes they haven't been. On certain occasions, INT items have even been found to be more memorable than MCI concepts. One can make sense of this apparently paradoxical picture by recognizing that multiple factors, in addition to the ones originally proposed by Boyer, can have an effect on memorability. These include, as discussed earlier, the notion of Inferential Potential, the von Restorff effect, coherability, and perhaps other factors such as the linguistic structure of MCI items, the kind of conceptual combination involved in the creation of the items, coding schemes to assess the number of violations and recall patterns, the nature of the violations, as well as contextual effects. Since studies designed to test the predictions of the MCI thesis have not systematically controlled for the effects of these variables, it should come as no surprise that the resulting empirical picture is muddled.

What all of this means is that researchers shouldn't expect a strict hierarchy where MCI items are always more memorable than INT or MXCI concepts. Given what we know today, empirical departures from this rigid memorability hierarchy do not necessarily entail that Boyer's MCI thesis has been falsified. At the same time, the discovery of a growing set of potentially confounding variables that have not been systematically controlled for in previous studies also means that we still do not know if violations of ontological assumptions lead to the memorability effects predicted by Boyer's account or if the effect is epiphenomenal. In other words, multiple experimental studies notwithstanding, the jury is still out regarding the privileged role of ontological violations for enhanced memorability.

On the theoretical side, we suggest that epiphenomenalism is a strong contender for explaining the true underlying mechanisms driving the MCI effect. Assumptions that held all MCI items to be equally memorable and the mechanisms underlying the effect to be fundamental may have been warranted when the effect was first proposed. However, considering the confusing pattern of memory results, we believe that these assumptions should be relaxed. What is left is a view of the MCI effect that preserves the empirical predictions as well as the explanatory power of Boyer's provocative idea. In other words, on the view proposed here, we should expect both that MCI items will outperform INT and MXCI items, when factors like IP are held constant, as well as that supernatural concepts around the world should have this structure. However, the present account also extends the predictions made by the original account in a number of interesting ways.

First, we should expect the most successful concepts to be highly agentic, either due to having an agent, such as a person, as the base of the concept or through characteristics that increase IP and agency. Second, we expect some successful concepts to be more complicated

than the traditional MCI account would permit, but under predictable conditions of retaining high coherability (and perhaps high IP). Third, concepts that are low in IP, despite having an MCI structure, should be poorly remembered and infrequent among extant supernatural concepts. More speculatively, by removing the superiority of violations of intuitive ontologies, or deep inferences, in favor of BIZ or violations of shallow inferences, the MCI effect might expand to account for cultural concepts that are not necessarily supernatural. To be clear, we are not suggesting that there is nothing special about supernatural characteristics. We believe that these characteristics are excellent at producing salience and IP, particularly in the form of increased agency. However, BIZ concepts also produce some salience, may easily produce IP, and might compensate for their slightly reduced salience with improved coherability.

While supernatural concepts are incredibly common across cultures and are often invented and transmitted today, many successful concepts are not supernatural, but fit nicely into the account we have developed here. For example, many cultural myths involve heroes and heroines who are uniquely (but not supernaturally) skilled in combat, leadership, strength, or intelligence. These are precisely the sorts of violations of shallow inferences, coupled with improved agency, for which the epiphenomenal MCI theory predicts successful memory and transmission. Today, many popular television shows are about “natural” high powered and rich doctors, lawyers, athletes, and politicians. These characters are not MCI in the sense first proposed by Boyer, but could it be that their cultural success is due to the same mechanisms?

As we noted above, we do not claim that the case is closed in favor of epiphenomenalism. In order to disentangle the original MCI thesis and the alternative account that we propose here, the memorability profiles of BIZ and MCI items will need to be measured as factors like unusualness and IP are held constant. If, under such conditions, MCI items are no more

memorable than their BIZ counterparts, there may be reasons to believe that the enhanced memorability of MCI items reported in some studies may indeed be epiphenomenal. If, on the other hand, such controlled comparisons still reveal a memorability advantage of MCI over BIZ items, this would suggest that Boyer may have been correct about ontological violations having some special quality (though, of course, there may be additional variables involved that we have not identified). Additional experimental work will also be needed to test the predictions of our coherability-based account of the decreased memorability of MXCI concepts. All this work is currently under way in our laboratory.

Alongside the theoretical considerations discussed above, future work on the memorability of religious concepts will also benefit from methodological prescriptions that have emerged from the literature on the MCI thesis. Echoing Purzycki and Willard's (2015) conclusions, the following list provides some, although by no means all, of the relevant suggestions:

- (1) The notion of counterintuitiveness will need to be defined more precisely (e.g., Purzycki and Willard's deep vs. shallow inferences) to more readily distinguish between epiphenomenalism and fundamentalism.
- (2) The linguistic structure as well as the kind of semantic combinations that are used to create MCI items should be taken into account (Franks, 2003).
- (3) The contextual predictability/postdictability of INT, MCI, and MXCI items will need to be carefully measured and controlled for (Upal et al. 2007; Upal, 2010)
- (4) More rigorous coding schemes to count the number of violations, along the line of Barrett (2008a), should be used.
- (5) A publicly available database of INT, MCI, and MXCI concepts that have been normed on relevant dimensions could be created and used in future studies.
- (6) Authors should provide a complete list of test materials and instructions to participants used in their studies so as to ensure meaningful comparisons and replication efforts.
- (7) Variables such as inferential potential, agency, coherability, and the degree of unusualness caused by the violations should be measured and controlled for.

The theoretical and methodological considerations outlined in this article will, we hope, lead to a deeper level of understanding regarding the role played by known cognitive

mechanisms in the memorability of supernatural concepts. At the same time, the MCI effect, whether it remains fundamental or is derived from other known mechanisms, will also need to be integrated into a broader account of religious thought and behavior that goes beyond assessing the memorability profile of supernatural concepts (Gervais, et al. 2011; Norenzayan & Atran, 2004; Parren, 2017).

References

- Anderson, J. R., & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. In L. S. Cermak and FIM Craik, Eds., *Levels of Processing in Human Memory*, 385-404.
- Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, 21(4), 547-569.
- Atran S. *In Gods We Trust: The Evolutionary Landscape of Religion*, Oxford University Press; 2002.
- 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.
- 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)*. London: Psychology Press.
- Baron-Cohen, S., Wheelwright, S., Spong, A., Scahill, V., & Lawson, J. (2001). Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome. *Journal of Developmental and Learning Disorders*, 5(1), 47-78.
- Barrett, J. L. (2004). Counterfactuality in counterintuitive religious concepts. *Behavioral and Brain Sciences*, 27(6), 731-732.
- 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., Burdett, E. R., & Porter, T. J. (2009). Counterintuitiveness in folktales: Finding the cognitive optimum. *Journal of Cognition and Culture*, 9(3), 271-287.
- 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.
- Berlin, B., Breedlove, D., & Raven, P. (1973). General principles of classification and nomenclature in folk biology. *American Anthropologist*, 74, 214-242.
- Boyer, P. (1992). Explaining religious ideas: elements of a cognitive approach. *Numen*, 39(1), 27-57.
- 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. (1999). Cognitive aspects of religious ontologies: how brain processes constrain religious concepts. *Scripta Instituti Donneriani Aboensis*, 17(1), 53-72.
- Boyer, P. (2002). *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.
- 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: Cambridge University Press.
- Chi, M. T., & Slotta, J. D. (1993). The ontological coherence of intuitive physics. *Cognition and Instruction*, 10(2-3), 249-260.
- Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268.
- Dawkins, R. (1976). *The Selfish Gene*. UK: Oxford University Press
- Delius, J. D. (1991). The nature of culture. In *The Tinbergen Legacy* (pp. 75-99). Springer, Dordrecht.
- Franks, B. (2003). The nature of unnaturalness in religious representations: negation and concept combination. *Journal of Cognition and Culture*, 3(1), 41-68.
- Gelman, S. A., & Hirschfeld, L. A. (1998). How biological is essentialism? In D. L. Medin & S. Atran (Eds.), *Folkbiology*. Cambridge, MA: MIT Press.
- Gelman, S. A., & Wellman, H. M. (1991). Insides and essences: Early understandings of the non-obvious. *Cognition*, 38(3), 213-244.
- Gervais, W. M., & Henrich, J. (2010). The Zeus problem: Why representational content biases cannot explain faith in gods. *Journal of Cognition and Culture*, 10(3), 383-389.
- Gervais, W. M., Willard, A. K., Norenzayan, A., & Henrich, J. (2011). The Cultural Transmission of Faith: Why innate intuitions are necessary, but insufficient, to explain religious belief. *Religion*, 41(3), 389-410.
- 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.
- Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 205(1161), 581-598.
- 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.
- Guthrie, S. (2015). On Purzycki and Willard's critique. *Religion, Brain & Behavior*, 6(3) 48-50.
- Hatano, G., & Inagaki, K. (1994). Young children's naive theory of biology. *Cognition*, 50(1-3), 171-188.
- 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.
- Leslie, A. M. (1982). The perception of causality in infants. *Perception*, 11, 173-186.

- McDaniel, M. A., Einstein, G. O., DeLosh, E. L., May, C. P., & Brady, P. (1995). The bizarreness effect: It's not surprising, it's complex. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 422.
- McDaniel, M. A., Dornburg, C. C., & Guynn, M. J. (2005). Disentangling encoding versus retrieval explanations of the bizarreness effect: Implications for distinctiveness. *Memory & Cognition*, 33(2), 270-279.
- Norenzayan, A. (2010). Why we believe: Religion as a human universal. In *Human morality and sociality: Evolutionary and comparative perspectives*, Macmillan International Higher Education. 58-71.
- 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.
- Nairne, J. S., & Pandeirada, J. N. (2016). Adaptive memory: The evolutionary significance of survival processing. *Perspectives on Psychological Science*, 11(4), 496-511.
- 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.
- Paivio, A. (1986). *Mental representations: A Dual-Coding Approach*. New York: Oxford University Press.
- Parren, N. (2017). The (possible) Cognitive Naturalness of Witchcraft Beliefs: An Exploration of the Existing Literature. *Journal of Cognition and Culture*, 17(5), 396-418.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7(6), 528-530.
- Pinker, S. (1997). *How The Mind Works*. WW Norton & Company.
- Pinker, S. (2003). *The Blank Slate: The modern denial of human nature*. Penguin.
- 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.
- 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.
- 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 UK.
- Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review*, 99(4), 605.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1), 89-96.
- Sperber, D. (1985). Anthropology and psychology: Towards an epidemiology of representations. *Man*, 73-89.
- Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In *The Adapted Mind: Evolutionary psychology and the generation of culture*, 19.
- Tweney, R., Gonce, L., Upal, M. A., Edwards, K., & Slone, D. J. (2006). The creative structuring of counterintuitive worlds. *Journal of Cognition and Culture*, 6(3-4), 483-498.

- Upal, M. A. (2007). What is More Memorable, Counterintuitive Concepts Interpreted Metaphorically or Literally? In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 29, No. 29).
- Upal, M. A. (2010). An alternative account of the minimal counterintuitiveness effect. *Cognitive Systems Research*, 11(2), 194-203.
- Upal, M. A., Gonce, L. O., Tweney, R. D., & Slone, D. J. (2007). Contextualizing counterintuitiveness: How context affects comprehension and memorability of counterintuitive concepts. *Cognitive Science*, 31(3), 415-439.
- 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*.
- Willard, A. K., Henrich, J., & Norenzayan, A. (2016). Memory and belief in the transmission of counterintuitive content. *Human Nature*, 27(3), 221-243.