Cognitive mechanisms for the evolution of religious thought

Sabela Fondevila and Manuel Martín-Loeches

Center for Human Evolution and Behavior, UCM-ISCIII, Madrid, Spain. Psychobiology Department, Complutense University, Madrid, Spain

Address for correspondence: Sabela Fondevila Estévez, Cognitive Neuroscience Section, Center UCM-ISCIII for Human Evolution and Behavior, Monforte de Lemos, 5, Pab 14, 28029 Madrid, Spain. sfondevila@isciii.es

The reasons behind the cultural persistence of religious beliefs throughout human history and prehistory still generate unanswered questions requiring scientific explanations. Within the framework of the cognitive science of religion, this article reviews experimental evidence supporting human predisposition for religious thinking and focuses on the hypothesis that a reason why religious beliefs are successful is their minimal counterintuitiveness. According to this hypothesis, religious concepts or stories would be characterized by containing only a small number of world-knowledge violations, which attracts attention while improving memorizability. We conclude this review by summarizing recent findings from our group using brain electrical activity and delving further into these questions. Our research suggests parallels between the natural tendency of the human cognitive system to use metaphors and the minimal counterintuitiveness of religious beliefs.

Keywords: counterintuitions; religion; N400; metaphors; ERPs

The cognitive science of religion

Religious thought is inherent to humans. It is likely to have been present in small groups since prehistoric periods and continues to prevail in large-scale contemporary societies despite cultural changes throughout history. At the beginning of the 20th century, approaches both from social constructivism and the phenomenology of extraordinary experiences constituted the first attempts to undertake a systematic study of religion. The gradual application of the Darwinian naturalistic perspective to the study of cultural expressions has led to the understanding that human behavior is a complex product of natural selection. Religion has not been immune to this process.

More recently, a novel evolutionary viewpoint in the cognitive sciences has focused on the causal explanations of this widespread human phenomenon. The cognitive science of religion combines a large range of methodological and theoretical perspectives, such as anthropology, psychology, philosophy, history, and neuroscience in pursuit of the reasons for the existence of religious thinking. This multidisciplinary approach turns religion into a dependent variable whose causes must be isolated and explained. Most importantly, this discipline views the religious mind as a natural by-product of ordinary cognition. Religious concepts would activate functional and anatomically cognitive systems that have arisen throughout human evolution as adaptations to deal with nonreligious environmental pressures. However incredible they may seem, religious beliefs could be explained by the same processes that humans use for daily life.

To investigate this naturalness of religion, scholars have been devoted to clarifying how humans acquire, represent, and transmit religious beliefs from an epidemiological perspective. At first glance, it is surprising how spirits, virgins giving birth, speaking animals, thinking objects, and similar miracles appear in all mythologies and religions throughout the world, despite their empirical nonexistence. They circumvent our natural expectations but, at the same time, they have easily and successfully spread throughout all cultures.

doi: 10.1111/nyas.12265
Evidence from diverse disciplines has shed light on the mechanisms driving this process, providing insights into the roots of our cognition.

**Human predisposition for religious thinking**

Different findings from developmental psychology have shown early tendencies toward religious thought in children. Kelemen and colleagues carried out a series of experiments demonstrating that 4- to 5-year-old children systematically reason about natural objects, animals, and artifacts in terms of purpose, and this is the case even when children did not receive explicit religious education. This remarkable bias toward teleological reasoning led Kelemen to claim that children are intuitive theists. Furthermore, even adults with scientific backgrounds are prone to providing teleological explanations of natural phenomena in experimental conditions with high cognitive demands.

Another line of research has shown that children do not attribute false beliefs to intangible agents (e.g., God), but they do it to nonsupernatural entities, such as animals or plants. Conversely, the psychologists Justin Barrett and Frank Keil have demonstrated that adults tend to unconsciously anthropomorphize the concept of God despite their theological knowledge about his supernatural nature and capacities. Finally, Bering and his colleagues have explored after-life beliefs in both children and adults, finding that they both naturally attribute emotional and epistemic states as well as desires to an already dead character.

Despite these suggestive findings, studies are still scarce and display some methodological problems. The lack of cross-cultural evidence precludes reliable generalizations of human behavior. Furthermore, the use of explicit tasks does not always provide a critical understanding of the underlying reasoning mechanisms employed by people when dealing with supernatural entities. In spite of these limitations, based on these results, some authors claim that religion is innate and they defend the existence of a cognitive device specifically evolved for religious beliefs. In a similar vein, others underscore the intuitiveness of religion, and argue that young children are equipped with quasi-religious default assumptions. As previously mentioned, however, these conclusions would require further empirical support. Indeed, current evidence supports the idea that explicit reflections do not always govern human conceptions of the supernatural. Instead, a wide range of unconscious inferences naturally and spontaneously applied to any agent are actually being applied to supernatural agents too, beginning early in childhood. Accordingly, it appears more appropriate to consider that there is a natural predisposition in human beings toward beliefs requiring the development of intuitive theories about the ordinary functioning of the world. As the cognitive scientist Ilkka Pyysiäinen has claimed, religion is no more innate than football or politics; what may be innate are the cognitive capabilities underlying these activities.

**The minimal counterintuitiveness of religious beliefs**

The notion of intuitive ontology encompasses a broad catalog of information domains into which world knowledge is organized. When an external input fits into a certain domain, the inference machinery is automatically triggered. According to cognitive scholars, each inference system or domain is based on its function, develops according to specific programs, and involves precise neural circuits. As a result, inference systems do not unequivocally support ontological categories such as “manmade objects” or “living things.” Instead, they are devoted to detecting particular evolved functions such as tool use, biological motion, or intentions. These sets of inferences or core knowledge systems help humans to represent objects, actions, numbers, space, and social partners. They could be framed into more general intuitions about the physical, biological, and psychological functioning of the world, guiding human experience starting in the early stages of individual development. Hence, we intuitively understand that a ball bounces against a wall, that mammals give birth to mammals, or that my friend sympathizes with my happiness. These facts require neither special attention to be comprehended nor too much effort to be processed. They occur around us continuously and unconsciously.

The inference machinery works in a slightly different fashion when religious beliefs are concerned. Actually, religious beliefs typically break our expectations about the natural world (e.g., a talking stone). Cognitive scientists have investigated the distinctive features of religious concepts in this regard, in order to better explain how they were easily
spread as cultures evolved.\textsuperscript{18,23} The anthropologist Pascal Boyer proposed that religious concepts are better characterized as being minimally counterintuitive (MCI), and that this is the reason for their cultural success\textsuperscript{18} (the MCI hypothesis). They are counterintuitive because they violate our core intuitive expectations (in the psychological, biological, or physical domains) related to a particular ontological category (persons, animals, plants, natural objects, or artifacts). These violations occur when an expectation is broken thoroughly (e.g., an invisible person) or when one property from one category is transferred to another category (e.g., a stone sculpture that can hear our prayers). Furthermore, they are MCI because there are few expectations that are violated (i.e., one or two), keeping most other intuitions intact. This allows for inference generalization and, consequently, effortless processing. For instance, the sculpture remains a solid stone object displaying all the features of such an object (it has weight, it can be eroded or broken, etc.) but with the peculiarity that it can hear.

Consequently, religious beliefs are both attention grabbing and easy to process; that is, they are cognitively optimized (cognitive optimums) to the human mind, and this characteristic explains human attraction to religious concepts and, as a result, their spread and success.\textsuperscript{24} This in turn, as developed below, would convey recall advantages over intuitive and maximally counterintuitive concepts (i.e., concepts comprising many violations of core knowledge expectations). Indeed, when facts surprisingly break our expectations in a disproportionate fashion, attention and memory appear distorted. For example, if you see a flying and speaking rock that changes its color when people look at it and gives birth to a particular kind of tropical fish, you would surely direct your attention to it. Yet at the same time, the processing of such a quantity of inconceivable characteristics would require a deep cognitive effort, which in turn would probably cause high memory degradation over time. On the other hand, ordinary events of daily life (e.g., the sun rising everyday) are not surprising enough to be appealing.

Indeed, memorability is a crucial psychological factor for successful cultural diffusion. If a concept is easily remembered, it is more likely to be transmitted. In the last decade, a fruitful line of research has been devoted to empirically testing the recall advantages of MCI concepts. In a series of experiments, Barrett and Nyhof\textsuperscript{25} presented Native American folk tales to university students who had to remember and retell them to others. MCI concepts were systematically recalled better than intuitive and bizarre ones (i.e., extremely odd but consistent with natural laws, such as a heavy butterfly) embedded in those narratives even after a 3-month delay. Similarly, Boyer and Ramble\textsuperscript{26} found recall advantages for MCI concepts in people from different cultures: Tibetan Buddhist monks, East Africans, and Western university students. Interestingly, counterintuitions experienced very few transformations upon retellings in these studies, mimicking natural contexts of cultural transmission.

From a slightly different approach, Lisdorf\textsuperscript{27} analyzed Roman prodigies, which are microstories famous for having been faithfully preserved across centuries of retellings. The author concluded that these exceptional narratives are great sources of MCI concepts. Similarly, other authors have also examined large samples of legends and fables and have consistently come up with the same results.\textsuperscript{28} Given the growing academic interest in MCI religious concepts, the psychologist Justin Barrett has proposed a system to conventionalize their coding.\textsuperscript{29} Recently, some have tested Barrett’s proposal by creating concepts with different degrees of counterintuition in order to verify their memorability. Once more, those technically described as MCI according to these criteria performed in a subsequent recall task with the greatest success.\textsuperscript{30} In addition, concepts describing intentional agents were more likely to be found in different folktales\textsuperscript{28} and to be evaluated as religious.\textsuperscript{31} Overall, these studies support the claim that the MCI nature of certain concepts used in religious frames is likely to constitute the cognitive optimum facilitating their cultural diffusion.

Nevertheless, several authors hold a somewhat different viewpoint and stress that the optimality for cultural transmission due to minimal counterintuition does not occur at the level of the concepts assembled \textit{per se}, but rather at the level of whole narratives. Social psychologists Norenzayan \textit{et al.}\textsuperscript{32} have demonstrated this hypothesis by analyzing the Grimm brothers’ fairy tales. They found that the most culturally successful stories exhibited very few counterintuitive (two or three) and many intuitive statements. In essence, these MCI narratives would have survived with greater cultural resonance when compared to those displaying a higher number of...
counterintuitive ideas. In this same vein, several studies have emphasized the role of context in this regard, showing higher recall rates for MCI concepts only when embedded in coherent stories. Furthermore, these concepts should contribute to the global cohesion and integration of the story to attain an improvement of their recall and transmission. In sum, the narrative- or story-level domain seems a crucial variable, complementary or alternative to the conceptual level, in explaining the wide success of minimal counterintuitions, and this, in turn, is a key feature of religious beliefs.

As reviewed so far, the cognitive optimality hypothesis, embodied in minimal counterintuitiveness both at the level of concepts and/or of narratives, has been operationalized and investigated in a vast amount of experimental evidence. Its relevance, therefore, seems well established within the frame of the cognitive science of religion, even if further cross-cultural and ontogenetic studies are still pending in order to attain a status of universality. However, evidence to date comes exclusively from behavioral studies; thus, little is known about how the brain integrates and comprehends religious beliefs. This lack of attention from the cognitive neuroscience field is actually part of a more general lack of experimental work using neuroimaging techniques for the study of cultural phenomena. However, this situation is currently undergoing changes, and the scientific study of religion is incorporating new methods and technologies in order to explore the neurocognitive mechanisms underlying religious thinking.

**Neuroimaging of religious cognition**

During the last decades, event-related brain electrical potentials (ERPs) have been one of the most widely used procedures to investigate cognitive processes. Their high temporal resolution provides real-time measurement of brain activity arguably related to cognitive processes as these occur and develop across time. ERPs typically consist of positive and negative deflections, also known as components. These differ in latency, amplitude, and scalp topography. Among the several ERP components, we will focus on the so-called N400 component; this has been used by the authors of the present work in the study of the neurocognitive mechanisms of minimal counterintuition in religious thinking, as will be summarized below.

The N400 component is a negative peak occurring at around 400 ms after the onset of the critical stimulus (typically, a visually presented word) with centroparietal scalp topography. Most recent functional interpretations for this negative wave claim that the N400 reflects the neural activity underlying semantic retrieval prompted by an eliciting stimulus. The N400 amplitude is primarily modulated by the unexpectedness of a word in the ongoing sentence, finding larger amplitudes for unexpected (typically, semantic violations) relative to expected words. It is also functionally sensitive to the access and organization of long-term semantic memory; in this regard, larger amplitudes are obtained for semantic violations across different categories than for violations within the same category. In addition to its sensitivity to local or sentential semantic incongruencies, and importantly for our purposes, world-knowledge violations also generate larger N400 amplitudes.

With this background, our team has recently approached the study of religious thinking using the N400 component of the ERP as its main tool. One of the premises used in our approach consisted of the above-mentioned sensitivity of the N400 to world-knowledge violations, and, hence, the N400 as a plausible index of the degree of perceived counterintuitiveness of an idea.

In a first basic approach, we wanted to test whether minimal counterintuitiveness might already apply at the level of single statements included in religious narratives, regardless of the number of this kind of statements included in a whole story. For these purposes, we extracted a number of counterintuitive ideas as found in actual religious texts and mythologies throughout the world but unfamiliar to the participants in the experiment. Namely, the religions and mythologies employed were Hinduism, Mesoamerican, Japanese, Egyptian, Greco-Latin, African, Australian, Chinese, Polynesian, and Eskimo. In most of the original sentences, the last word implied a word-knowledge violation, such as “From his mind emerged the moon.” When violation did not occur in the last word, the sentence was rephrased accordingly; in this way, we used the last word as the critical word to elicit the N400. Next to these world-knowledge-violation (in fact, 80% were core knowledge) statements (religious counterintuitions), we constructed correct (i.e., intuitive) sentences by replacing the last word with
Figure 1. N400 modulation by counterintuitiveness. (Left) Brain electrical potentials (ERPs) at a central electrode (Cz) to the processing of counterintuitions extracted from religious texts and mythologies (solid black line), nonreligious counterintuitions (dotted line), and intuitive sentences (gray line). The N400 component of the ERP is modulated by the nature of the counterintuitions, showing that those used in mythologies and religions appear more readily acceptable and accessed from long-term semantic memory than other world-knowledge incongruencies. (Right) Maps of the N400 difference waves: (top) religions minus intuitive and (bottom) nonreligious minus intuitive.

a semantically correct one (e.g., “From his mind emerged the idea”). Finally, a set of semantic incongruencies were also constructed by replacing the last word of each sentence with another one yielding world-knowledge violations (again, 80% core knowledge) that were judged as implausibly appearing in religious narratives, that is, nonreligious counterintuitions (e.g., “From his mind emerged the house”). All the variables known to typically affect the N400 (such as expectedness, frequency, imaginability, etc.) were equivalent across all types of sentences (intuitive, nonreligious counterintuitive, and religious counterintuitive), as detailed in the referenced paper.

Our main finding (Fig. 1) was a significantly reduced N400 amplitude for religious counterintuitive ideas as compared to nonreligious ones. This was interpreted as an index of greater acceptability for these statements relative to nonreligious ideas, even when both were judged as unacceptable by the participants. The differential N400 modulation was clear-cut and, therefore, it was concluded that people find thinking about the moon rather than a house emerging from someone’s head less cognitively demanding. This would, in turn, support the notion that miracles and counterintuitive assertions from mythologies and religious texts are minimally (or at least less) counterintuitive than other counterintuitions by themselves and in isolation. Under this view, the operationalized conception of minimal counterintuitiveness might be amended and expanded: from a reduced number of core-knowledge violations within the frame of a narrative or a concept—as reviewed above—to a single idea that by itself violates core knowledge but at the same time does not seem so extravagant or salient to the human semantic system as other violations, as measured by the N400 component of the ERP.

By exploring the N400 component literature in search of possible reasons for this N400 amplitude attenuation to religious counterintuitions, we found that this result might be consistent with electrophysiological findings on metaphor processing. These experiments typically report an N400 amplitude reduction for metaphors relative to other semantic anomalies, showing that metaphor processing is less cognitively taxing. Additionally, behavioral studies have shown that it is difficult for people to ignore metaphors, as it takes longer to judge them as literally false in comparison to semantic violations. According to several authors such as George Lakoff, proponent of the contemporary theory of metaphor (CTM), a metaphorical understanding is the natural approach of our linguistic system, that is, the default mode of interpretation; metaphors would be intrinsically part of our conceptual system, grounded on human experience and cognition, and, therefore, processed effortlessly. In our first study, subjects were instructed to judge whether the statements were likely to occur in the real world, with an indication of not thinking metaphorically when arriving at their decisions. However, this might have been not sufficient to significantly remove metaphorical interpretations given the naturalness and automaticity of these processes, making religious counterintuitions more acceptable (in terms of N400 amplitude), even if considered implausible.

To explore a possible relationship between counterintuitions used in religious stories and metaphorical thinking, we carried out a follow-up experiment in which the same religious and nonreligious counterintuitive sentences used in the previous study were forced to be interpreted either literally or metaphorically. In order to enforce these...
two different processing modes, questions about the literal and a possible metaphorical content of every sentence were set up in a blockwise design. Our findings show that response times to probe questions after religious counterintuitive statements were similar to those for nonreligious ones in the literal mode, but significantly faster in the metaphorical mode. Interestingly, brain activity results paralleled performance. In the literal mode, both counterintuitive ideas evoked similar N400 effects, but in the metaphorical mode the N400 amplitude for religious counterintuitions was significantly reduced, while nonreligious sentences still evoked a strong N400 effect. These findings indicate that counterintuitive statements employed in religious texts are especially susceptible to being understood in a metaphorical mode of thinking; this is not the case for other types of counterintuitions. Accordingly, one critical factor possibly contributing to the minimal counterintuitiveness of certain religious ideas, as reflected in the N400 component of the ERP, might be their amenability to metaphoric interpretation.

The conception of metaphors as natural to the human brain is connected to the aforementioned idea of the naturalness of religion. Indeed, the links between religious counterintuitions and metaphors seem to go beyond this. In this regard, both MCI religious ideas and metaphors would establish conceptual mappings between different semantic domains, as proposed for metaphors by the CTM. According to this theoretical view, conceptual mappings in metaphors occur when a few properties are transferred from the source to the target domain while many others are preserved. Remarkably, this depiction parallels the definition of minimal counterintuitiveness. These conceptual parallels have not been entirely unnoticed in metaphor literature. In fact, the CTM has been successfully applied to the analysis of religious texts, and approaches from cognitive archaeology have also found remarkable links between metaphoric and religious thinking.

Religions and mythologies seem to make use of a special kind of counterintuitive situation, a type that is prone to being interpreted metaphorically, at variance with other types of world-knowledge violations. Returning to the initial question of explaining the cultural success of religious thinking, and taking into account the approach presented here, we speculate that religious beliefs could have benefited from the tendency of the human brain to easily assume and interpret metaphors. This way, world-knowledge violations used in religious narratives may become MCI, hence plausibly benefiting from the memorizability advantages of this property, as seen above at the concept and narrative levels. Although this is an ongoing endeavor and more research remains pending, our results seem to suggest that this feature might have contributed to the evolutionary success and wide spreading of religious ideas.

Conflicts of interest
The authors declare no conflicts of interest.

References


