

Universal cognitive biases as the basis for supernatural beliefs: evidence and critiques

Aiyana K. Willard

Centre for Culture and Evolution,
Brunel University London

Corresponding Author: aiyana.willard@brunel.ac.uk

Hugh Turpin

Centre for the Study of Social Cohesion,
School of Anthropology and Museum Ethnography,
University of Oxford

Adam Baimel

Department of Psychology, Health, and Professional Development,
Oxford Brookes University

Words: 5692 (excluding abstract and references)

Abstract

What explains the ubiquity of religions across time and space, and why do these supernatural belief systems seem to have so much in common? Many cognitive scientists of religion have proposed that cross-cultural patterns in religious belief are, at least in part, the indirect result of reliably developing and otherwise adaptive features of the human mind. These 'cognitive bias' theories propose that religion is a by-product of universal mental architecture. We see similar beliefs recur in unrelated cultural and historical contexts because of biases in how we perceive and interpret the world, and how we remember concepts. This chapter reviews the evidence, merits, and limitations of such theories. In so doing, the chapter addresses the most influential of the cognitive bias theories: the proposed relationships between various religious beliefs and Theory of Mind, anthropomorphism, dualism, teleological reasoning, and minimally counterintuitive concepts. We address both the strengths and shortcomings of these theories in explaining religious belief and suggest where we can go from here.

Keywords: Religion, cognitive biases, anthropomorphism, mind-body dualism, teleology, minimally counterintuitive

Why are cognitive biases important to understanding religion?

Many features of religion, such as ritual practice and the belief in supernatural agents, are widely shared across religions (Tylor, 1871). What might explain the cross-cultural recurrence of these focal elements of religious systems? Early researchers in what became the 'cognitive science of religion' proposed that in order to explain the universality of religion, a new general theory of religion that considered religion's evolutionary origins was necessary (Boyer, 1994; Guthrie, 1980; Lawson & McCauley, 1990). According to this view, the everyday functioning of the human mind makes supernatural/religious reasoning intuitively appealing or a "path of least resistance for our cognitive systems" (p. 1039, Boyer, 2008). This intuitive appeal was not claimed to be adaptive, but rather a by-product of otherwise adaptive functions (Atran & Norenzayan, 2004; Kirkpatrick, 1999). This early work led to an examination of variants of religious beliefs with the goals of establishing their common, or universal, features and identifying their cognitive foundations. Research focused on the ways in which our minds perceive and process information that could give rise to religious-like beliefs. If these beliefs are intuitively supported, then they may have a selective advantage on the cultural marketplace (Sperber, 1996). Intuitively supported beliefs would appear with similar features around the world, and be more likely to survive and persist across generations (Atran & Norenzayan, 2004).

This claim that the foundations of religion are cognitive launched a new era of looking at religion as universal and focusing on its common traits across our species (Sørensen, 2005; C. White, 2021). The key question at the centre of this cognitive and evolutionary revolution was identifying the specific cognitive processes and/or systems that made our species so prone to supernatural reasoning. One such process is mind perception, or Theory of Mind (ToM) – the ability to ascribe mental states (e.g., thoughts, beliefs, desires, emotions, and/or

knowledge) to others. Though this ability clearly evolved for understanding other people and social learning, we do not limit our use of ToM to other humans. We use it to reason about a wide array of animals, objects, weather, and supernatural agents. This overuse, or misapplication, of ToM has been linked to many of the cognitive biases thought to underly religious belief (Willard et al., 2020; Willard & Norenzayan, 2013), such as anthropomorphism (Guthrie, 1993, 1996) – the tendency to see non-human things as human; mind-body dualism (Bloom, 2005, 2007) – the tendency to see minds as separate and not reducible to physical bodies; and teleological reasoning (Kelemen, 1999a, 2004) – the tendency to see things as being intentionally made for a function or purpose.

Another major by-product theory is minimally counterintuitiveness (MCI; Boyer, 2001). MCI theory suggests that religious ideas are more memorable because they violate core ontological categories – described as innate or reliably developing expectations we have for certain categories of things. They are, in essence, exceptions to our most deeply held rules about the world. The claim is that the things we see in religions, like talking animals or ghosts, persist and spread because they violate our intuitions about the world and therefore stand out in our memories. Other by-product theories we will not discuss pertain to behavioural features of religion such as ritual (e.g. Boyer & Liénard, 2007). Here, we focus on the relationship between cognitive biases and religious *beliefs*. These theories have been particularly influential in this domain, and ritual is dealt with elsewhere (Whitehouse, this volume).

These cognitive by-product accounts of religion suggest that religious content is constrained by, and succeeds because of, aspects of our evolved cognitive architecture. That being said, the relationship between the specific content of lived religions and cognitive biases is ambiguous at best. What these biases intend to explain is the tendency for some

broad types of supernatural beliefs to be ubiquitous across religious traditions (e.g., the existence of ghosts or spirits, or creationist myths) rather than the more particularistic features of specific religious beliefs (e.g., the trinity, or that cows are sacred). Even with this caveat, these theories are supported by empirical evidence of varying quality. Many require updating with current developments in our understanding of human cognitive and cultural capacities. We will discuss each major theory in turn, including the evidence that supports them and the challenges they face, before discussing some of the ways forward from here.

Theory of mind

The link between supernatural agent beliefs and Theory of Mind (ToM) is clear: Gods, ghosts, and spirits are believed to have minds (i.e., capable of thoughts, beliefs, and desires); you need to be able to reason about minds in general to reason about supernatural minds (Gervais, 2013). At the same time, the relationship between measures relevant to ToM and religiosity are inconsistent (Jack, 2016; Lindeman et al., 2015). When found, this relationship is small; perhaps because it is indirect. Although all neurotypical adults have ToM abilities, individuals differ in their capacities for, and tendencies in, deploying their ToM. Those more prone to considering the mental states of others are potentially also more prone to related cognitive biases such as anthropomorphism, mind-body dualism, and teleology (C. J. M. White, Willard, et al., 2021; Willard et al., 2020; Willard & Norenzayan, 2013). These correlated tendencies may lead to higher rates of religious belief, not differences in ToM proneness directly.

To better understand these pathways, we need to consider which aspects of ToM are implicated in the perception of supernatural minds. These minds are not like human minds. They have powers, abilities, or knowledge beyond what we expect from normal human minds. Supernatural minds are often disembodied, meaning most supernatural agents lack

many of the embodied cues we use to determine the mental states of people, such as facial expressions, posture, and gaze. This leaves us nothing with which to understand a god's mind but our imaginations, hallucinations, and what we believe are supernatural actions in the world. Luhrmann (2012) describes how Vineyard evangelical Christians infer God's mental states from fortunate or unfortunate events encountered in daily life. They also learn to perceive some of their own thoughts as divine interjections. Cases such as this suggest culturally acquired and effortful practices for manipulating use of ToM plays more of a role than baseline variation in capacities for mentalizing.

These differences also create problems in how we measure use of ToM in this research. Like most psychological constructs, our tendency to perceive and reason about other people's minds is measured indirectly through what we expect the various types of resulting behaviours and abilities will be. Tests of adult ToM abilities measure things like the tendency to empathise and socialise with others (mentalizing; Baron-Cohen & Wheelwright, 2004), or our accuracy in reading emotion from facial cues (Baron-Cohen et al., 2001). It is not clear these are the same aspects of ToM that allow us to think about supernatural agents.

Other research has shown an inverse relationship between autism, which can include difficulties in ToM and mentalistic reasoning, and religious beliefs (Norenzayan et al., 2012), but this also may not be straight forward. Autistic populations can and do have rich relationships with invisible, imaginary, and supernatural agents, but these agents are idiosyncratic agents of their own making (Visuri, 2012, 2020). These invisible agents may be easier for people with autism to understand because they remove the complexities of navigating emotion recognition and other sensory stimuli involved in deciphering the mental states of humans (Visuri, 2018). This can be reconciled if difficulties with ToM impact the cultural transmission of religious belief, rather than the ability to perceive and reason about

invisible agents. Without the socially learned component of belief, we would expect most supernatural believers to do what these autistic samples do: create their own individual supernatural beings.

Anthropomorphism

Anthropomorphism is the tendency to project human mental states on non-human animals, objects, events, etc. Within the cognitive science of religion, it was first championed by Stuart Guthrie (1993, 1996), although discussions on the role of anthropomorphism within Christianity date to much earlier (Feuerbach, 1841; Hume, 1779). Guthrie's observation was that we tend to see human forms everywhere, and often these sightings are religious in nature (e.g. Huffpost, 2017). We assume that a human-like agent is behind the rustling of leaves or bumps in the night (Barrett, 2004). This tendency should trigger a sense that these things have human-like minds we can interact with. This over perception of agents leads to belief in supernatural agents.

This idea is given evolutionary backing through error management theory (see Haselton & Buss, 2000). Error management holds that when you are uncertain you should err on the side of whatever is more fitness relevant, in this case the detection of agents (Guthrie, 1996). When you hear a rustling in the trees or see a face in the clouds, caution (and evolutionary fitness) dictates you should assume it is an agent because agents are dangerous. If you are wrong, there is very little cost. On the other hand, if you don't detect an agent that is there, the cost could be your life. The theory thus posits that people will be biased to interpret ambivalent cues in the environment as the product of agency. This 'module' was dubbed the Hypersensitive (or Hyperactive) Agency Detection Device (HADD; Barrett, 2000, 2004).

There are several issues with this theory. First, and rather importantly, there is no convincing empirical evidence that we over detect agents in this manner, and not for lack of

trying (Andersen, 2017). Second, there does not seem to be any plausible neurological basis for the existence of a HADD (Lisdorf, 2007). Third, agency detection may not be what we need to explain the existence of supernatural *minds*. A HADD assumes that reasoning about agents is automatic and not cognitively taxing, but it is not clear this is true if we are reasoning about that agent's *mind*, rather than simply perceiving it as alive. Though we can reason about cues to agency like self-directed motion and simple goals effortlessly (Csibra et al., 1999), reasoning about mental states is costly enough to inhibit our ability to do other reasoning tasks while we are doing it (Jack et al., 2013). If you are standing in a crowded room, it is easy to know that other people in that room are alive and have basic goals (get food, socialise, etc). If you tried to think about what was going on in people's minds you would not be able to focus on more than a couple people at a time and doing so would make it hard to track anything else in the room.

Andersen (2017) re-framed the detection problem as one of predictive coding – we over detect agents based on our prior expectation of their presence. This means we are likely to think a bump in the night is a ghost when we are already wary of ghosts. Though likely accurate, it does not solve the problem of how we come to believe in supernatural agents in the first place (Willard, 2017). Though it is effective at explaining the anecdotes about rustling leaves and bumps in the night, it requires pre-existing belief. Even with all of this, Guthrie's observation that people routinely believe in human-like agents that are not human is clearly true. We do anthropomorphise all sorts of things, including randomly moving dots on a screen (Tremoulet & Feldman, 2000), anime pillows (Galbraith, 2019), and of course supernatural agents.

Other theories of anthropomorphism have addressed this phenomenon using a motivational theory of anthropomorphism (Epley et al., 2008). Waytz et. al (2010) looked at