RELIGIOUS BELIEF AS ACQUIRED SECOND NATURE

by Hans Van Eyghen

Abstract. Multiple authors in cognitive science of religion (CSR) argue that there is something about the human mind that disposes it to form religious beliefs. The dispositions would result from the internal architecture of the mind. In this article, I will argue that this disposition can be explained by various forms of (cultural) learning and not by the internal architecture of the mind. For my argument, I draw on new developments in predictive processing. I argue that CSR theories argue for the naturalness of religious belief in at least three ways; religious beliefs are adaptive; religious beliefs are the product of cognitive biases; and religious beliefs are the product of content biases. I argue that all three ideas can be integrated in a predictive coding framework where religious belief is learned and hence not caused by the internal architecture of the mind. I argue that the framework makes it doubtful that there are modular cognitive mechanisms for religious beliefs and that the human mind has a fixed proneness for religious belief. I also argue that a predictive coding framework can incorporate a larger role for cultural processes and allows for more flexibility.

Keywords: born believers; cognitive science of religion; naturalness of religious belief; predictive processing

THE NATURALNESS OF RELIGIOUS BELIEF

A prevalent idea in cognitive science of religion (CSR) is that religious belief is natural. The phrase “religious belief is natural” can have multiple meanings...

Hans Van Eyghen is postdoctoral fellow in Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; e-mail: hansvaneyghen@gmail.com.
Some authors argue that religious belief is “cognitively natural” or “maturationally natural.” Justin Barrett and Aku Visala define “cognitively natural” as follows:

[T]here is something about our minds that dispose it to catch religious ideas . . . . [O]ur belief-forming mechanisms would be biased in such a way as to create a tendency or a disposition to acquire, think, and transmit religious ideas instead of some other kinds of ideas. (Barrett and Visala 2018, 69)

In Barrett and Visala’s view, beliefs that are cognitively natural are different from beliefs that are cross-cultural because cognitively natural beliefs are formed by virtue of the way the human mind is structured (Barrett and Visala 2018). In other words, cognitively natural beliefs are the way they are to a large extent because of the internal architecture of the human mind. Barrett and Visala add that not all religious beliefs are natural in this sense. Especially, elaborate culturally specific religious beliefs, like belief in the Trinitarian God, are not. Other, vaguer beliefs, like the belief that there are supernatural agents, would be cognitively natural (Barrett and Visala 2018).

Robert McCauley makes a similar claim and argues that many religious beliefs are “maturationally natural.” He writes: “Maturationally natural cognition concerns humans having (similar) immediate, intuitive views that pop into mind in domains where they may have had little or no experience or instruction” (McCauley 2011, 5, emphasis added) Maturational natural cognition is different than what McCauley calls “practiced natural cognition.” Practiced natural cognition is achieved by extensive experience in dealing with a domain. Clear examples are judgments experts can make in a snap, like an engineer who intuitively knows what materials to use or a chess player who instantly knows the best next move (McCauley 2011). Practiced naturalness requires a great deal of (cultural) learning while maturational naturalness does not.

McCauley gives three arguments in favor of religion’s maturational naturalness. First, religious belief goes way back in human history. Archeological evidence strongly suggests that prehistoric man had some kind of religious beliefs. Second, religious belief is both geographically and historically ubiquitous. Religious belief arose in different cultures and is nonetheless quite similar. Third, species related to man have similar traits. According to McCauley, chimpanzees display behavior that resembles human ritual behavior (McCauley 2011). All three appear to fit better with maturational naturalness than practiced naturalness. Various forms of (cultural) learning were likely less important for prehistoric man than they were for modern man. Similarities in religious belief across cultures are better explained by dispositions of the human mind than by specific cultural learning. Cultural learning is also far less important for apes than for modern humans.
The key in explaining the deep history and ubiquity of religious belief is in looking at the role of cognitive mechanisms or cognitive biases that give rise to religious belief. McCauley and others argue that cognitive biases that reliably develop with cognitive maturation produce elementary forms of religious belief.

In this article, I critically discuss three arguments for the conclusion that religious belief is cognitively natural or maturationally natural. Each argument draws on a different set of theories, all of which are usually ranked under CSR. One can argue that (some) religious beliefs are maturationally natural in three ways. One could draw on adaptationist theories to argue that religious belief served an adaptive function. Maturational natural beliefs are better suited to serve an adaptive function because they are more reliably produced than other (practiced natural) beliefs. One could also argue that religious beliefs are maturationally natural because they are produced by modular cognitive mechanisms. Since modular cognitive mechanisms appear to be part of the architecture of the human mind, their outputs are (mainly) due to the architecture of the mind and not due to cultural learning. Finally, one could argue that religious beliefs are maturationally natural because the mind has fixed content biases (like a bias to form belief in minimally counterintuitive ideas).

I note problems with all three arguments and argue that they can be accounted for by cultural learning. To do so, I rely on the predictive processing (PP) framework. On this framework, religious beliefs are mainly learned and thus not maturationally natural. I argue that PP can be wedded to adaptationist theories, can integrate a bias for religious beliefs, and can include content biases. Integration shows that religious beliefs or biases for religious belief need not be the product of the architecture of the human mind. Instead, I argue that a bias for religious belief and certain content biases could take root at an early age.

As a secondary goal, I aim to show how other (older) CSR approaches can be wedded to PP. By doing so, I aim to show that PP has the potential to unify. PP also leaves more room for (divergent) cultural influences and does more justice to the (limited) flexibility of human minds. It also sheds new lights on various arguments for epistemic implications of CSR.

**Arguments for Maturational Naturalness**

In this section, I discuss three arguments for the conclusion that (some) religious beliefs are maturationally natural. Each argument appeals to different kinds of CSR theories.

**Natural Biases for Belief?**

A first argument refers to CSR theories that explain religious belief by pointing to the operations of one or more cognitive biases. If one of these
theories is true, religious belief would be maturationally natural because cognitive biases result from the cognitive architecture of the human mind. Before I discuss the argument, I first discuss two example theories to which it refers.5

One well-known CSR theory argues that religious belief (partly) results from a bias toward agency detection.6 The theory is known as the Hyper-sensitive or Hyperactive Agency Detection Device (HADD). Defenders argue that humans are overly sensitive toward signs of agency. Vague patterns like oddly shaped clouds in the sky are easily seen as animals; sounds of rustling leaves are easily regarded as an indication that someone is approaching; and oddly shaped branches are easily seen as snakes. People would be prone to see natural phenomena as signs of agency because doing so is safe. Detecting too many signs of agency, at worst, causes loss of time and energy while detecting one too little might make you miss an attacking predator. Detected signs of agency are often overridden (or corroborated) by further investigation. In some cases, they could foster the belief that invisible agents are causing the patterns or noises. Belief in invisible agents could easily lead to belief in spirits and belief in spirits could lead to belief in gods.

Another well-known CSR theory is known as “promiscuous teleology.”7 According to the theory, people are prone to see things as purposeful or designed rather than as the result of natural processes. In one experiment, young children were asked why rocks are pointy. Most children preferred the answer “so that animals can scratch themselves” to “erosion made the rock this way” (Kelemen 1999). Deborah Kelemen argues that the bias toward teleological answers is suppressed when adults learn about biology and natural explanations for phenomena like pointy rocks. However, some studies suggest that the bias resurfaces when adults are put under time pressure or forget about the natural explanations (Kelemen and Rosset 2009). A study with Alzheimer’s patients and adults without biological education also showed a preference for teleological explanations (Casler and Kelemen 2008).

Beliefs about teleology or belief that natural phenomena are designed would foster religious beliefs in a designer. In this way, promiscuous teleology could foster the belief in a creator God. Both theories (HADD and promiscuous teleology) argue that religious belief is (at least partly) the result of the operations of cognitive biases. In the next section, I will discuss how the theories support the idea that religious belief is maturationally natural.

Cognitive Mechanisms as Part of the Human Cognitive Architecture

Does the claim that religious belief results from the operations from cognitive biases support the claim that religious belief is maturationally natural? It does if cognitive biases are part of the architecture of the human mind.
and do not result from cultural learning themselves. Below, I discuss why cognitive biases, like HADD and promiscuous teleology, are often regarded as part of the architecture of the human mind. Authors like McCauley suggest that beliefs produced by massively modular cognitive mechanisms are maturationally natural. In this section, I argue that the operations of HADD and promiscuous teleology resemble the operations of massively modular cognitive mechanisms. I also discuss why this would support the idea that religious belief is maturationally natural.

Defenders of both theories I discussed above describe the operations of cognitive mechanisms in a way that fits well with the massive modularity of mind thesis (MMMT). Before I argue how they do so, I discuss the main ideas of MMMT. The main idea of MMMT is adequately captured by Robert McCauley as follows: “They [defenders of MMMT] hold that the human mind is composed of dozens, perhaps hundreds, of specialized mental modules” (McCauley 2011, 52). These modules operate fast and their operations are not easily open to introspection.

As McCauley notes, massive modularity is popular among evolutionary psychologists. Evolutionary psychologists can easily argue that modular cognitive mechanisms evolved to tackle specific evolutionary problems.

The operations of HADD, as described by Stewart Guthrie (1993), share the characteristics of massively modular cognitive mechanisms. They are mandatory and fast. Forming a belief that an (invisible) agent is out there is formed quickly when a subject perceives patterns or vague noises. The operations of HADD are also not easily open to introspection. Barrett (2004) and Guthrie agree that the operations of HADD usually remain unconscious to the subject. HADD would also have been selected because it helped tackle the problem of predator sneak attacks.

Kelemen’s theory also fits well with massive modularity. People would have a disposition to give teleological answers when they need to answer fast. Why teleological answers are preferred is also not easily introspectable. Some experiments were even deliberately set up so that participants would have limited time to answer (Kelemen, Rottman, and Seston 2013).

Now how are massively modular cognitive mechanisms linked to maturational naturalness? They can be linked in three ways. As McCauley notes, massively modular cognitive mechanisms evolved to tackle different evolutionary challenges. If natural selection indeed selected for cognitive biases, it is likely that these are hard-wired in the cognitive architecture of the mind.

A second reason is that HADD and promiscuous teleology appear to operate from a very young age onward. Very young children would not have been exposed to various forms of learning long enough to have learned to detect agents or see teleology in nature. Having maturationally natural cognitive mechanisms can explain why they quickly form beliefs about agency and teleology in any event.
A third reason points to what I call “relapse phenomena.” I discussed how promiscuous teleology “resurfaces” when subjects are under time pressure or forget about alternative mechanistic explanations. Maturational naturalness can better account for such phenomena. If teleological beliefs were formed in another way (practiced naturalness or other) we would not expect such a relapse. Maturationally natural beliefs can be regarded as default states that emerge under normal circumstances when they are not overridden or inhibited. On this view, knowledge of mechanistic explanations overrides natural proneness for teleological beliefs. When this knowledge does not override (because of time constraints or loss of knowledge), maturationally natural beliefs reemerge.

**Born Adaptive Belief**

Many CSR theories refer to evolutionary theory. Some CSR theorists argue that religious beliefs do not have adaptive value. Instead, they argue that religious beliefs arose as a by-product of other adaptive traits. Some prominent CSR theories do argue that religious beliefs have adaptive value themselves. For most theories, the adaptive value is that religious beliefs enable better cooperation. Since cooperation is hugely important for human survival, better cooperation leads to better chances of survival. I will refer to these theories as “adaptationist CSR theories.” I will give an example.

A well-known adaptationist CSR theory is known as the “Big Gods” (BG) theory. Defenders of the theory argue that groups who believe in powerful, morally concerned gods who punish or reward people in accordance to their good deeds had an advantage over other groups. Belief in BG would make people cooperate better. It mainly does so because it lowers the prevalence of free riders. A free rider is someone who reaps the benefits of cooperation but does not contribute anything herself. A high prevalence of free riders undermines the trust that is required for smooth cooperation in a group or society. Free riders also use up more resources than they bring in and thus lower the net gains of cooperation for others. People who believe that one or more powerful supernatural beings exist who are concerned with human behavior, and will punish people in accordance to their behavior, would be far less likely to free ride. Although free riding behavior might go unnoticed by other people, it is not hidden from supernatural beings. Furthermore, free riding behavior will have severe consequences if morally concerned supernatural beings can punish free riders with bad luck or hell. Believing that such beings exist would thus make people think twice before free riding and make them cooperate better.

Defenders of the BG theory do not argue that belief in big gods is a biological adaptation. For the largest part of human history, societies did
not need BG to have smooth cooperation. Societies were small enough to enforce cooperation by social monitoring or kin selection. Belief in BG did yield an advantage in the Neolithic age. Because of their belief in BG, some groups could cooperate well enough to live in large-scale societies. Because large-scale societies quickly grew more powerful than small-scale societies, these groups outcompeted groups without belief in BG.

**Maturationally Natural Adaptive Beliefs?**

A second argument for the maturational naturalness of religious belief is that such beliefs are better suited to tackle evolutionary challenges. Beliefs that are reliably formed under most circumstances, without need for explicit cultural instructions, will likely provide more evolutionary benefits than practiced natural beliefs or other beliefs. For the BG theory to work, subjects need to reliably produce belief in moralizing, punishing gods. Cultural learning appears to be too variable to reliably produce such beliefs. Cultural transmission can take many different courses. Belief in moralizing gods could also be replaced with other (non)-religious beliefs in the course of history. Putting the burden of transmitting adaptive beliefs mainly on cultural transmission might be too much to ask. Maturationally natural beliefs are far less subject to cultural changes. For this reason, they could reliably yield an evolutionary benefit.

Some defenders of the BG theory do claim that cultural learning plays an important role for religious beliefs. They argue that evolved cognitive biases canalize and constrain religious belief, but cannot explain why people hold deep commitments to gods. To explain this, cultural transmission must be taken into account. According to defenders, committed belief in BG is largely the result of cultural learning mechanisms. For example, the fact that many more people share a committed belief to the Trinitarian god and not to Zeus nowadays is largely explained by the fact that most people learn to do so from their environment (Gervais et al. 2011).

The BG theory therefore does rely on cultural learning to some extent. Although evolved cognitive biases can be expected to reliably produce religious beliefs, they cannot be expected to produce the right kind of religious beliefs required for large-scale cooperation. At least some of the burden rests on cultural learning.

**Born Content Biases**

A final argument points to content biases that lead to religious beliefs and are present at a very young age. The most widely discussed content bias of this kind is Pascal Boyer’s cognitive optimum (2002). Boyer’s theory argues that some features about the content of religious beliefs make them memorable and salient. Boyer argues that minimally counterintuitive concept have the best chances of being remembered and transmitted.
According to the theory, people have intuitive ontological categories like “plant,” “animal,” or “person.” These categories allow for predictions. For example, categorizing something under “plant” allows the inference that the thing will not be able to move and will grow under the right conditions. Minimally counterintuitive concepts (MCI) violate some of the expectations that come with their intuitive ontological category. For example, a ghost is usually categorized under “person” but violates the expectation that persons cannot move through walls. At the same time the majority of expectations is retained and a ghost is expected to perceive things in the same way as persons do and to interact with others as persons do.

Minimally counterintuitive concepts differ from intuitive (concepts that violate no ontological expectations) and maximally counterintuitive concepts (concepts that violate many ontological expectations). Intuitive concepts, like “John Doe,” are not memorable because they are ordinary. Maximally intuitive concepts, like a man who is 30 m long, has 11.5 arms, and only appears every second Tuesday of the month, require too much cognitive effort to remember and are too different to make predictions. Most supernatural concepts are minimally counterintuitive and thus easily (optimally) transmitted and remembered.\textsuperscript{16}

\textit{Maturational Natural Content Biases}

Benjamin Purzycki and Aiyana Willard argue that minimally counterintuitive concepts should be sharply distinguished from counterschematic concepts. On Boyer’s theory, the intuitive ontological categories are deeply ingrained in the human mind.\textsuperscript{17} When a being or object is classified under one ontological category, the mind can make “deep inferences” by applying expectations that come with the category. These “deep inferences” stand opposed to “shallow inferences.” The latter stem from more accessible and more specific relations between concepts and reflective information. For example, a rose is expected to bob in the wind or a cross is expected to have a longer vertical axis. Concepts that violate shallow inferences are counterschematic rather than counterintuitive (Purzycki and Willard 2016).\textsuperscript{18}

The distinction between counterintuitive and counterschematic maps well to McCauley’s distinction between maturational naturalness and practiced naturalness. Content biases that are maturationally natural would reliably give rise to intuitive ontological categories and would be present from a very young age on. Practiced natural content biases would be more divergent and more culturally specific and would manifest at a later age.

\textbf{Predictive Processing}

The previous section discussed three arguments in favor of regarding religious beliefs as maturationally natural. Before arguing how religious beliefs can be regarded as a form of practiced naturalness, I will introduce
predictive processing (PP). PP presents a general theory of cognition and how people learn.

*The Predicting Mind*¹⁹

The core claim behind PP is that the human mind is a self-learning, Bayesian prediction machine. When it receives sensory input, the mind is making educated guesses about the cause of that input. It does so by relying on an internal model of the world that provides information about the statistical probability of what can be expected to be around. According to defenders of PP, experiences are constituted by two factors:

1. sensory input, and
2. an internal model of the world that bears information about the likely cause of that input.

Although other models of cognition also claim that input is filtered by top-down processes, PP radicalizes the idea. It argues that all perception is heavily shaped by top-down processes, not just noisy or ambiguous perception. The mind constantly checks its internal model of the world to make statistical estimates of what is likely out there in the world. For example, when out bird watching, a subject who sees a crow will have the visual experience because of two factors:

1. the incoming rays of light, reflected by the bird in her retina, and
2. her internal model that bears the information that the probability of finding crows in the forest is high.

Although both factors are important, the second factor makes the largest contribution to the experience. The subject that is out bird watching is expecting to see birds. She is also in a place where people often see birds and is paying close attention to phenomena that could be birds. All this information makes her mind conclude that a moving black dot in the sky is very probably a bird. This is so because a bird is much more probable to be out there than other black flying objects, given the information the subject has. The subject’s knowledge makes her conclude that the dot is probably a crow.

An important question is why a subject can be expected to have a reasonably precise or reliable internal model of the world that contributes to reasonably accurate experiences. According to defenders of PP, the internal model is reasonably reliable because it is constantly updated when there is a mismatch between the sensory input and the internal model. A mismatch is called a “prediction error.” For example, our bird watcher will have a prediction error when she is not expecting to see a (very rare) bird, but sees one anyway. Her internal model assigns a very low probability to finding
the rare bird and therefore predicts that she will not observe any. When she
does get sensory input of the rare bird, the probability of finding the rare
bird is updated so that it is at least not trivial. By constantly updating the
internal model after prediction errors, the internal model can be expected
to grow ever more reliable and precise.

According to defenders of PP, processing of sensory input happens on
multiple, hierarchical levels. Lower levels deal with events happening at
faster timescales and have greater detail. Higher levels deal with things
happening at slower timescales, which are more abstract in nature. Models
at the higher levels construct plausible hypotheses about the cause of the
sensory input by making predictions. Only the lowest levels receive a
representation of the original sensory input; the next levels merely receive
the prediction error if there is one. If no prediction error was recorded,
higher levels receive no signal and continue as if the hypothesis was correct.

Another important question is what drives the model to be ever more
precise and reliable. According to defenders of PP, an internal model can
be expected to grow more precise and reliable because of the free-energy
principle. The principle states that systems tend to avoid disorder. For this
reason, they tend to minimize the entropy of their sensory states. Entropy
is the measure of uncertainty the internal model has. An uncertain model
will yield a lot of prediction errors and therefore a lot of surprise. Because
models that are better attuned to the environment will yield less surprise, a
system will move toward a more accurate model. The whole process often
remains unconscious.

An important way to reduce entropy and surprise is by taking action. A
subject with an inaccurate model will be prompted to take action to
minimize surprise. Not acting will often result in more surprise, which is
what systems are trying to minimize. For example, a walker who hears a
cry will be prompted to take action and find out what is causing it. Karl
Friston argues that reflexes are often ways in which a subject tries to get
more specific sensory input (Friston 2018).

Although predicting minds will move toward more accuracy, efficiency
requirements imply that a subject will not and should not get everything
right. Like all minds, human minds need to be able to make quick calls
about their environments to survive and to flourish. As a result, the human
mind cannot pause too often to check whether its internal model matches
well with the input it receives from its environment. On many occasions,
friction will go unnoted and the internal model will not be updated.
The predicting mind can thus be expected to have an imperfect model,
which leads to some inaccurate experiences and false beliefs about the
environment. How often these inaccurate experiences occur is unclear.
Marc Andersen (2017) argues that the predicting mind can be expected to
make more errors in low-light environments or when the subjects suffer
from sensory deprivations.
There is another reason why the internal model should not be expected to be perfect. Updating the internal model too stringently runs into the danger of overfitting. The term “overfitting” comes from statistics. A statistical model overfits when it corresponds too closely to a particular subset of the data. A model can overfit in at least two ways. First the model can take too many features into account. Returning to our example, if the bird watcher updates his model to incorporate all features of every observed bird (color, structure of feathers, size of beak, and so on), his model will fail to generalize to new, unobserved birds that lack some of these features. As Rajesh Rao and Dana Ballard (1999) note, the model needs to be efficient. Efficiency requires a certain level of generality.

A second reason why models could overfit is by relying on an unrepresentative set of samples. A bird watcher who has only been exposed to black birds and has never received any information about colored birds might form an internal model wherein all birds are black. The internal model will have little problems correctly identifying crows and ravens as birds. It will have problems correctly identifying many other birds.

Another reason why internal models should not be perfect is that humans have a limited amount of cognitive energy. Activities that use up a lot of cognitive energy tend to have negative effects on cognitive performance. For example, job interviews where subjects need to respond to a lot of questions or public presentations often leave subjects tired and distracted afterward. These effects are called “depletion effects.” Other causes of depletion effects are emotion suppression and high self-regulation (Baumeister et al. 1998). Having very detailed and complex models of the world likely poses great demands on the human cognitive system and has negative effects on cognitive performance in the long run. Having slightly less detailed and complex models will likely be less consuming and have a better performance in the long run.

Although efficiency prevents the mind from updating its model too often, internal models can also, among other reasons, be exposed to an unrepresentative set of data. I return to this point below.

**Predicting Supernatural Agents**

Recently, Marc Andersen (2017) applied the PP framework to religious cognition. Andersen argues that the PP framework can explain why people have experiences of supernatural agency. We discussed how two factors contribute to sensory experiences according to defenders of PP in the previous section: (1) sensory input, and (2) a subject’s internal model of the world. For experiences of supernatural agency, both factors would be

1. some ambiguous stimulus, and
2. an internal model with beliefs about supernatural agents.
Andersen’s examples of (1) resemble the examples given by defenders of HADD (see Section 2). A subject could have experiences of supernatural agency after seeing patterns of ambiguous objects or after hearing vague noises. The main contributions for the experience, however, come from the internal model. If the subject believes that supernatural beings exist and engage with people, it is far more likely that internal model will make her have experiences of supernatural agency. This is the case because their internal models predict that the probability of finding supernatural agents is at least nontrivial. Religious believers, with a religious internal model, will therefore be more prone to conclude that some ambiguous sensory input is caused by a supernatural agent and thus experience the input as such.

According to Andersen (2017), experiences of supernatural agency can also be brought about by suggestion. Internal models of subjects who are told that some supernatural agent might be around will assign greater probabilities to finding supernatural agents. This could very well lead them to have more experiences of supernatural agency. Andersen draws support for his claim from a number of studies. In one study, a Swedish team tried to replicate a study conducted by Michael Persinger and his team. Persinger claimed that a helmet could induce mystical experiences by electromagnetic radiation (Booth, Koren, and Persinger 2005). The Swedish team found that a placebo helmet had the same effects Persinger reported. Since their helmet was fake, they attribute the effect to suggestion (Granqvist et al. 2005).

A number of authors have applied PP to mystical experiences. According to Chris Hermans (2015), these experiences arguably involve mental processing at a higher level. Michiel van Elk and Eric-Jan Wagenmakers (2017) agree and argue that the PP framework needs to be expanded to account for these higher level experiences. Van Elk and André Aleman (2017) give some suggestions on how PP could account for various mystical experiences.

**Religious Biases as Practiced Natural**

Andersen puts his own PP theory in sharp contrast to older CSR approaches. In this section, I argue that parts of older CSR theories can be fitted into a PP framework. Integration requires us to rethink the idea that religious belief is maturationally natural. I argue that the intuitions and empirical data that drive older CSR theories fit equally well the idea that cognitive biases for religious belief are the result of something akin to overfitting at a young age. This can also account for the adaptive use of religious belief and content biases that give rise to religious belief.
**Born Overfitters**

We noted that, although systems tend to form ever more accurate models, a case can be made that pragmatic problems prevent them from updating internal models too stringently.

If the error processing takes its normal route, an overfit model (e.g., with only black birds) can be updated. The mind can, however, become rigid. There is evidence from neuroscience that brain cells shrink and connections between different areas of the brain disappear when subjects grow older (Aleman 2014). There is also evidence that young children are more eager to learn and form new ideas (Gopnik 2009). What a subject was exposed to at a young age might thus have a far greater impact on the subject’s internal model than exposure at older age. Older minds might have a harder time updating overfit models. Prediction errors might also be less noticeable to overfit minds. A mind that is trained that all birds are black will be less attentive to colored dots in the sky when the subject is out bird watching.

Below I argue that the cross-cultural phenomena to which older CSR theories refer could result from overfitting at a very young age. Some phenomena, which I call “relapse phenomena,” appear to contradict my claim. I argue that these can be explained by time constraints.

**Overfitting on Features**

Andersen argues that older CSR theories rely too heavily on modular theories of the mind. We noted above that according to modular theories of mind, the human mind has a range of distinct tools for distinct functions. Defenders of PP argue for domain-general models of perception and cognition instead. On these models, the same computational principles are used to process information for a large range of different domains (Andersen 2017, 6–7). Some authors have argued that CSR theories can do without relying on the massive modularity thesis. Most CSR theories allow for flexibility in the operations of cognitive mechanisms and even for conscious intervention. Andersen, however, goes even further. He suggests that there is no distinct cognitive mechanism for religious belief like HADD. Instead, (religious) cognition would rely on one or more general mechanisms, which can process a whole range of input and produce a whole range of beliefs. I argue that although PP leaves little room for distinct cognitive mechanisms, it does leave room for learned cognitive biases.

Andersen’s portrayal of PP leaves little room for distinct cognitive mechanisms. David Maij and Michiel van Elk (2019) argue that PP can make room for cognitive mechanisms by allowing for “evolved priors.” They note that relying heavily on cultural transmission is problematic. The idea runs into the “dark room problem.” The problem states that a predictive mind situated in a completely dark room will be unmotivated to move out
of the room since doing so would lead to error overload.\textsuperscript{28} Evolved priors can solve the problem by specifying what a subject without any cultural input will find surprising. A proneness toward agency detection could be an evolved prior according to Maiij and van Elk (2019). A default state with preprogrammed priors about the world could allow subjects to navigate their environments without much cultural input.

Some PP accounts do allow for priors that are not the result of cultural transmission. In response to the dark room problem, Friston, Christopher Thornton, and Andy Clark (2012) replied that human subjects and minds could not survive in a dark room indefinitely. Since human subjects need things like food and heat to survive, a dark room that is closed off from the world will lead to surprise when these needs are not met. They thereby argue that the bodily form, biomechanics, and initial neural architecture of a human subject shape its initial model of the world. These make the human subject “expect” basic requisites for life like food and heat. Friston, Thornton, and Clark’s response can be read as arguing for innate priors like an expectation of food and heat. These could very well have evolved. They are, however, a far stretch from an evolved appetite for agency detection. PP can allow for a hyperactive appetite for agency detection in another way. Hyperactive agency detection can be the result of early-age overfitting on agency.

Instead of being evolved, bias hyperactive agency detection and promiscuous teleology could be learned. Young children quickly learn that animate beings are different from inanimate things. John Opfer and Susan Gelman and John Opfer (2002) conclude from a survey of developmental evidence that children know the animate–inanimate distinction by the age of 10 months. They also note that for children the most important features to distinguish animate and inanimate things are featural cues—in particular whether the thing has a face or not—and dynamic cues—whether the thing can engage in self-generated and self-sustained motion.\textsuperscript{29} Focus on both features could lead to overfit. Children could easily form priors that all things that have faces or engage in self-generated and self-sustained motion are animate. By consequence, they could become prone to classify things with face-like patterns and things that make sudden movements as agents.

Andersen notes that a universal tendency toward agency detection is not supported by the empirical evidence. In response, Guthrie (2017) argues that there is at least evidence that agency is at least privileged in human cognition.\textsuperscript{30} The PP account I outlined above predicts that many subjects will have a proneness toward hyperactive agency detection. It does allow that the proneness can disappear when subjects successfully update their models to include other features of agency.

Overfitting can also explain why people are prone to see teleology or design. Generally, complexity is a good indicator of design and teleology.
Most complex things young children encounter are designed by human agency. The young mind can therefore easily learn priors that all complex things are designed. When subjects learn how complexity can arise gradually by nonagentic forces (e.g., by natural selection or erosion) the prior needs to be revised and subjects become less prone to see teleology.

Kelemen’s work does challenge my PP account. Some of her studies suggest that adults slip back into promiscuous teleology under time pressure (Kelemen and Rosset 2009). I call these phenomena “relapse phenomena.” On a PP account, we should expect adults to update their priors on teleology. Once the priors are updated, promiscuous teleology should disappear permanently. In response to the problem, I note that although adults make more mistakes under time pressures, their responses are still more accurate than those of young children. An adult’s models of teleology therefore appear to be more accurate than those of children. More accurate models can still make mistakes. Making more mistakes under time pressure could be explained by just that: adult predicting minds lack the time to accurately process input and therefore make mistakes. More empirical data on relapse phenomena are, however, needed to see how often they occur and whether they are best explained by maturational natural biases or predictive minds.

Overfitting on Adaptive Beliefs

On Andersen’s account of PP, instruction, learning, and testimony are the main sources of (religious) prior beliefs that shape a subject’s internal model of the world. He does allow for some innate, evolved priors (see above), but they are very basic and not religious. It does not yet explain why cultural transmission can be expected to reliably produce belief in moralizing, punishing gods. A tendency to overfit (on belief in big gods) can solve this problem.

If the human predictive mind is indeed prone to overfit, it can explain why cultural transmission can meet its adaptive responsibilities. PP also allows a natural account for how subjects can come to have a committed belief in big gods. Children are confronted with authoritative figures from a young age. Often authoritative figures exert authority for moral reasons. For example, parents punish children for transgressing moral norms. In this way, children learn to follow moral norms by obeying authoritative figures. In doing so, they form priors about moralizing, authoritative figures. Cultural transmission can tap into these priors with compelling narratives about supernatural authoritative figures. For example, Old Testament stories about God punishing the Israelites for their disobedience can resonate with people because his actions resemble those of human authoritative figures. Compelling narratives that resonate with priors about authoritative figures can make subjects form priors about BG. These priors can foster cooperation.
This account can explain why cultural transmission can be expected to reliably produce belief in moralizing, punishing gods. When the moralizing, punishing nature of big gods is emphasized, it can make the predictive mind make a stronger connection between prosocial behavior and moralizing gods. When subjects are often reminded of the moralizing nature of gods and how prosocial behavior can deter punishment, they will become more committed to belief in big gods. The predictive mind will thus fit more strongly on prosocial priors.

This account also allows for more cultural variation. Ara Norenzayan (2013) argues that belief in big gods is on the decline because the modern welfare state has largely taken over the role of monitoring human behavior. An account where belief in moralizing gods is transmitted and reinforced by compelling narratives can incorporate this. In modern societies, there will be less need to remind people of moralizing gods because prosocial behavior is successfully enforced by the welfare state. It can also explain why some societies have or had smoother cooperation than others.

Minimally Counterintuitive Concepts as Outliers

At first glance, the flexibility of the predictive mind seems hard to reconcile with the rigidity of the human mind in preferring minimally counterintuitive concepts. Andersen’s account of PP suggests that the mind should show more flexibility in its preferences. If cultural transmission favors intuitive concepts, the mind should be expected to remember those best. If it favors maximally counterintuitive concepts, they should be remembered more easily. The mind’s preference for minimally counterintuitive concepts might be the result of overfitting as well.

In their response to the dark problem, Friston, Thornton, and Clark (2012) suggest that humans come equipped with only very basic priors. It is unlikely that they come equipped with full-blown ontological categories like “plant” or “person.” The ontological categories are likely built up inductively. Subjects learn to classify beings as persons by attending to features that define personhood. These features in turn come to constitute intuitions for that category. We noted earlier that fitting a model to a restricted set of samples could make the mind rigid. Applied to plants, most subjects will fit on plants that cannot engage in self-generated movement because the vast majority of plants they encounter indeed do not move out of themselves. Information about a plant that does move out of itself (e.g., a Venus fly trap) yields an error. Like priors about agency detection, the mind can overfit on ontological categories. By encountering mostly immovable plants, the prior “plants cannot move” can become deeply ingrained in the human mind. When the mind gets less flexible by ageing, subjects could find it harder to classify plants that violate the prior as plants.
So far, I have argued that overfitting can give rise to rigid ontological categories. Boyer’s theory (2002), however, claims that concepts that violate a minimal number of expectations of an ontological category will be most salient and best remembered. On PP, minimally counterintuitive concepts will prompt an error. In the normal course, the error would prompt a revision of the internal model. A revised model would no longer have a prior like “plants cannot move.” Moving plants would then stop being minimally counterintuitive. However, since moving plants are outliers among plants, predictive minds could learn not to classify moving plants as just any other ordinary plant. Most statistical models are harder to fit on data with outliers. The predictive mind can learn to fit moving plants as plants, but because moving plants are outliers it could easily learn to classify them as special plants. Matters will be more difficult for maximally counterintuitive concepts. Because these violate a lot of expectations that come with ontological categories, they are likely to be explained away by the predictive mind and not lead to updates of ontological categories.

Rephrasing Boyer’s theory in a PP framework can help explain why not all minimally counterintuitive concepts are religious. The importance of human interaction for human subjects can explain why MCI persons are more salient than MCI plants. MCI persons also fit better in religious narratives. This could help solve the “mickey mouse problem.” Cultural transmission can also explain why people no longer worship ancient Greek gods like Zeus.

Maturational or Practiced Naturalness

Conceptualizing vague religious beliefs as resulting from predictive minds that learn religious biases can explain why (1) we find evidence for religious beliefs in human deep history, (2) why we find religious beliefs cross-culturally with recurrent features, and (3) why related species display similar traits. Prehistoric man had a brain that is not radically different than the brain of contemporary humans. Prehistoric man therefore likely had a similar predictive mind that could easily learn biases for agency detection and teleology and ontological categories. Our human ancestors likely did not have narratives about moralizing gods. This fits well with the claim that belief in big gods became dominant at a later stage in human history.

Since humans have similar minds and are exposed to animate and inanimate things, we can expect them to develop biases for agency detection and teleology cross-culturally. Whether cultures have a belief in big gods will depend on whether they have been exposed to narratives of big gods. This also fits well with what defenders of the big gods theory claim. They claim that belief in moralizing gods grew dominant in the Neolithic age but not that it grew to be universal. We can also expect humans with similar predictive minds to learn similar ontological categories.
Matters are more speculative concerning related species. It is not clear how different the minds of apes are to the minds of humans. Claiming that apes have predictive minds that learn biases or ontological categories is therefore highly speculative. McCauley is, however, also careful to draw strong analogies between human religious behavior and animal ritualistic behavior. For one thing, animal ritualistic behavior appears to lack meaning (McCauley 2011, 150–51).

**Are Overfitters Poor Fitters?**

I argued that religious belief could result from overfitting at a young age rather than being maturationally natural. The claim that humans are natural believers has attracted a lot of attention from philosophers and theologians. Their main focus was whether CSR “debunks” or raises a negative verdict on religious beliefs. A common argument is that belief that is maturationally natural would not be sensitive to truth. John Wilkins and Paul Griffiths (2013), for example, argue that CSR shows that religious belief would have evolved if true or if not true. This would undermine the confidence we should give to religious beliefs. A common response is that God could have directed evolution so that intelligent humans with a maturationally natural belief in God would evolve. God could do so by setting the evolutionary process up in the right way or by intervening in the process where necessary (e.g., Murray 2009).

If religious beliefs result from overfitting rather than from the way the human mind naturally functions, religious beliefs would be more sensitive to truth. Overfit minds do make mistakes. For example, if subjects overfit on classifying agents based on face recognition and self-generated motion, they will (often) wrongfully classify moving objects or face-like objects as agents. How easily subjects make mistakes is not clear. This does, however, not raise a negative epistemic verdict on religious beliefs. To do so, a debunker needs to show that all or most religious beliefs are the result of erroneous detection of agents.

Overfitting on adaptive beliefs also does not obviously damage religious beliefs. I suggested that adaptive beliefs (like beliefs in big gods) are reliably produced in human populations because they fit well with priors about authoritative figures and because compelling narratives are transmitted by means of cultural transmission. To do damage to religious beliefs, a debunker needs to show that the narratives are false or fabricated.

Considering supernatural agents as outliers among the ontological category of persons also does not do any damage. Like Boyer’s original theory, it only explains why supernatural concepts are better remembered and transmitted and more salient. It says little about whether believing that such counterintuitive agents exist is rational or true.
CONCLUSION

I argued against the claim that religious beliefs are maturationally natural. I argued that the phenomena that allegedly support this thesis (adaptive belief, cognitive biases, content biases) can also be explained by a PP model, where religious belief is the result of (cultural) learning from an early age. In this way, the propensity to form religious beliefs is more like what Hermans calls “a pattern of practice” thoroughly shaped by culture and human cognitive abilities (Hermans 2015). The model provides a more economical and more plausible explanation for religious belief. It can also better incorporate an important role for cultural processes and allow for more flexibility.

NOTES

1. For an overview of different meanings, see Barrett and Visala (2018).
2. Other CSR scholars are less firm. For example, Ara Norenzayan merely argues that “[there is] a suite of cognitive faculties [that] reliably develop in children, and regularly reoccur across cultures and historical periods. There are several such faculties, which appear to incline human minds toward religious belief” (Norenzayan 2013, 15).
3. McCauley also argues that religious ritual behavior is maturationally natural (McCauley and Lawson 2002).
4. I noted that Barrett and Visala hold that some religious beliefs are not cognitively natural. Examples would be complex theological beliefs like Trinitarian Christian belief. They therefore only argue that some religious beliefs are cognitively natural. McCauley also argues that, although most religious beliefs are formed easily and quickly, other, theologically complex beliefs are not. In the remainder of this article, I will not repeat this point and use the term “religious beliefs” to refer to those religious beliefs that are formed quickly and easily.
5. Examples of theories of this kind that I do not discuss in this section are Jesse Bering’s “Existential Theory of Mind” (Bering 2002) and Kurt Gray’s “Moral Dyad” (Gray and Wegner 2010).
6. The main defenders are Stewart Guthrie (1993) and Barrett (2004). My discussion is mainly drawn from Barrett.
7. The theory was put forward by Deborah Kelemen (1999).
8. The original modularity of mind thesis (Fodor 1983) states that the human mind has a set of distinct, specialized input systems. Defenders of the massive modularity thesis expand the idea to state that central cognition consists of distinct, specialized input systems as well.
9. Examples are Guthrie’s and Barrett’s theories, which I discussed in Section 2.1. Other CSR theorists who claim that religious belief is a “by-product” are Jesse Bering (2002) and Pascal Boyer (2002).
10. Biologist Stephen Jay Gould compared by-products to spandrels in arches. A spandrel is the space between two arches or between an arch and an enclosure. The spandrel does not have a function in upholding the structure but emerges alongside structures that do. Similarly, an evolutionary by-product does not have an evolutionary “function” itself, but arises alongside adaptive traits that do. Many CSR theories argue that religious beliefs arose in a similar way as by-product of other adaptive traits (Gould 1997).
11. Well-known examples are the “Broad Supernatural Punishment Theory” (Johnson 2015) and the “BG theory” (Norenzayan 2013).
12. Other examples of adaptationist CSR theories are the Broad Supernatural Punishment theory (Johnson 2015) and theories connecting religious belief to sexual selection (Slone and Slyke 2015).
13. See Norenzayan (2013) and Norenzayan et al. (2016) for defenses of the BG theory.
14. A related theory does claim this (see Bering and Johnson 2005).
15. Gervais et al. (2011) do argue that cultural transmission is constrained to transmit beliefs that are “potentially actionable,” “fitness relevant” or “plausible.” These, however, easily allow for religious belief in nonmoralizing gods as well.


17. Purzycki and Willard (2016) closely connect the intuitive ontological categories to modular operations of the mind. They argue that Boyer’s theory fits in a strong modular view of the mind. Modular, encapsulated mechanisms would naturally give rise to ontological categories.

18. Purzycki and Willard (2016) argue that many empirical tests of Boyer’s theory did not take this distinction into account. They do raise severe worries whether the distinction between deep and shallow inferences can be properly operationalized. This would make Boyer’s theory hard to test.


20. See Friston (2010). The concept of free energy was first used in thermodynamics. Here, the change in free energy is the maximum work a thermodynamic system can do in a process at constant temperature. Defenders of PP use a concept of free energy that is more similar to how the term is used in variational Bayesian methods. Here, free energy represents the upper bound on a variational Bayesian model. In PP, free energy is therefore the upper bound on surprise and minimizing that upper bound can reduce surprise.

21. For the relation between action and minimizing surprise, see Friston (2018).


23. The term was first used by Baumeister et al. (1998).

24. Schjoedt et al. (2013) argue that depletion effects can explain why people often fail to process religious events individually and are more susceptible to authority and suggestion.


27. For example, Van Eyghen (2018).


29. A lot of the evidence Gelman and Opfer (2002) survey does argue that the animate–inanimate distinction is innate or modular. Whether the distinction can be regarded as an evolved prior falls beyond the scope of this article. Friston, Thornton, and Clark (2012) would probably argue that it does not.

30. Guthrie argues that agency detection is a privileged default in human cognition. Any default prior that goes beyond basic expectations (e.g., expectations of food and heat) seems problematic on a PP account. Young children who are exposed to other cues of agency (e.g., being self-organizing) could very well develop different priors. However, since engaging in self-generated, self-sustained motion is easier to grasp, classifying something as an agent according to these features might come close to being a default. This could also hold for face recognition because almost all children encounter agents with faces from a young age onward.

31. In statistics, outliers are observation points that are distant from most other observation points.


33. See Barrett (2008).

34. See Gervais and Henrich (2010).

35. In his defense of the BG theory, Ara Norenzayan gives examples of small groups without belief in BG like the Hadza of Tanzania (Norenzayan 2013, 121–22).

36. For an overview, see Van Eyghen, Peels, and Van den Brink (2018).

REFERENCES


