Filosofia Unisinos
Unisinos Journal of Philosophy
19(3):302-310, sep/dec 2018
Unisinos – doi: 10.4013/fsu.2018.193.12

DOSSIER

Predictive coding and religious belief

Codificação preditiva e crença religiosa

Hans Van Eyghen[1]

ABSTRACT
In this paper I investigate the epistemic implications of a recent theory of religious cognition that draws on predictive coding. The theory argues that certain experiences are heavily shaped by a subject's prior (religious) beliefs and thereby makes religious believers prone to detect invisible agents. The theory is an update of older theories of religious cognition but departs from them in crucial ways. I will assess the epistemic implications by reformulating existing arguments based on other (older) theories of religious cognition.

Keywords: cognitive science of religion, predictive coding, epistemology of religious belief, reliabilist epistemology, the safety condition for knowledge.

RESUMO
Neste artigo investigo as implicações epistêmicas de uma teoria recente da cognição religiosa que se baseia na codificação preditiva. A teoria argumenta que certas experiências são fortemente moldadas pelas crenças prévias (religiosas) de um sujeito e, desse modo, torna os crentes religiosos propensos a detectar agentes invisíveis. A teoria é uma atualização das teorias mais antigas da cognição religiosa, mas se afasta delas de maneiras cruciais. Avaliarei as implicações epistêmicas reformulando os argumentos existentes baseados em outras teorias (antigas) da cognição religiosa.

Palavras-chave: ciência cognitiva da religião, codificação preditiva, epistemologia da crença religiosa, Epistemologia confiabilista, a condição de segurança para o conhecimento.

Introduction

In this paper I investigate the epistemic implications of a recent theory of religious cognition that draws on predictive coding. The theory argues that certain experiences are heavily shaped by a subject's prior (religious) beliefs and thereby makes religious believers prone to detect invisible agents. The theory is an update of older theories of religious cognition but departs from them in crucial ways. I will assess the epistemic implications by reformulating existing arguments based on other (older) theories of religious cognition.

This paper is structured as follows: in the second section, I lay out the predictive coding framework on which the new theory draws; in the third section, I discuss the new theory; in the

[1] Vrije Universiteit Amsterdam
De Boelelaan 1105, 1081HV
Amsterdam. E-mail:
hansvaneyghen@gmail.com

fourth section, I discuss three arguments that draw on theories of religious cognition. They are the 'unreliability argument', 'the unsafety argument' and the 'naturalness argument'.

I argue that the new theory does not support a conclusion for or against a positive epistemic status of religious beliefs. I also argue that the new theory leads to conclusions that are different from those of older theories of religious cognition. I end with some concluding remarks.

Predictive coding

Predictive Coding (PC) is a theory that is gaining traction in neuroscience and cognitive science. The main idea is that the brain is a Bayesian prediction machine which constantly runs and updates mental models of the environment. By operating in this way, the human mind can maximize accuracy and efficiency. The operations of the 'prediction machine' heavily shape perception. In this section, I will give a brief outline of the theory.[2]

Most discussion on PC focuses on perceptual experiences. PC rejects the idea that perceptual experience is determined by sensory input. Instead, perceptual experiences are heavily shaped by the human brain's internal operations. The brain constantly predicts what the subject perceives and thereby shapes perception. The core of PC is adequately captured by Hawkins and Blakeslee when they write: "Your predictions not only precede sensations, they determine sensation" (Hawkins and Blakeslee, 2004, p. 158). During perceptual experiences, the brain runs a model of the world and makes predictions about the causes of sensory input. For example, when a subject walks through a forest, her internal model will predict visual perceptions caused by trees. The model will shape perceptual experiences in such a way that incoming rays of light are processed as trees. The brain can do so because the model it runs has information about the statistical structure of some set of observed input. In other words, the model predicts what objects (or persons) the subject is likely to experience. In this way, the brain is always guessing what will happen next. The expectations derived from the model will influence what perceptual experiences the subject will have. Perception is thus always theory-laden in the sense that our perceptual experiences always depend heavily on a set of priors coming from the internal model.

Defenders of PC often compare the human brain to a Bayesian probability machine. Depending on its internal model, the human brain makes a guess about what the subject is likely to experience. When our subject walks in the forest, she is very likely to encounter trees because the prior probability of finding trees in a forest is high. She is far less likely to encounter tall buildings. Therefore, her internal model will usually make her have experiences of trees rather than of buildings. Experiences of trees in turn strengthen the internal model and the probability of finding new trees rises. In this way, experiences are both shaped by prior probabilities and bootstrap future probabilities.

The human brain is, however, not blind to the external world. While the model (or models) usually tries to match all sensory input to the existing model of the world,[3] the brain also looks for sensory input that does not match its model of the world. When this occurs, the brain makes what is known as 'prediction errors'. Although the probability is rather low, it is possible that there is a building deep in the forest. While visual input of the building will easily be seen as a tree or as something else that is expected in a forest, it can be seen as the building which it is. In cases like these, there is a mismatch between the prediction of the model and the sensory input. These mismatches are called 'prediction errors'. Prediction errors prompt the brain to update its internal model of the world. In our example, the model of the forest is updated to include a building. In this and other situations, the brain is able to use the prediction errors to update its internal model of the world.[4] After the building is perceived as a building, the probability of finding other buildings in the forest increases.

Processing of prediction errors can occur at multiple levels. Diuk *et al.* give the following example:

> *[I]magine a gambler who arrives at a city with multiple casinos, holding a set of coupons that allow him to enter any one of the casinos and play a number of different games. The gambler enters one casino and plays blackjack, roulette, and a slot machine. Each time he plays a game, he might observe a difference between what he expected to win and the actual outcome—a "game-level prediction error" that can be used to adjust his future expectations about this game. However, upon playing the last coupon for a casino, he not only learns about the last game itself, but also has enough information to update his knowledge about the casino as a whole: was this a good casino to spend his coupons on? It is at this point that two coincident reward prediction errors would arise: a simple game-related prediction error and a higher-level casino-related prediction error linked to learning the value*

[2] My discussion of PC mainly relies on Andy Clark's overview of the theory (Clark, 2013). His theory has a broader scope than perceptual experience. Since the theory I discuss in the third section focuses on perceptual experiences, I limit the discussion to those experiences in this section.

[3] Andy Clark (2013) calls this 'the brain explaining away the sensory signals'.

[4] This idea goes back to work on neural networks in artificial intelligence. See Churchland (1990), for an explanation of how neural networks learn.

> *of the casino as a whole. These prediction errors are not redundant. For example, the slot machine may have been worse than expected but the casino better than expected (Diuk et al., 2013, p. 5797).*

Like the gambler did in the example, the human brain can bring prediction errors at different levels together. In our example, prediction errors can lead to an update of the internal model of one particular forest but also to updates about forest in general. These levels are not experienced as distinct but are brought together by the brain.

An important question is why PC can be expected to be a reliable guide to navigate the world. The internal model can be completely wrong from the start or not be capable of updating its model in the right way. Karl Friston argues that PC works towards an optimal model by regulating the use of energy. Updates of the internal model of the world are governed by what is he calls 'the free energy principle' (Friston, 2010). The principle states that the brain will work towards minimizing the energy used in perceptual experiences. Since prediction errors lead to friction between the model and the world, it uses up more energy than a smooth fit. Therefore, the brain will use up less energy if it updates its model of the world to prevent prediction errors. Although Friston does not state this explicitly, he suggests that the free energy principle also prevents PC from updating the model too much. Pausing at every sensory input to see whether it matches the internal model in detail will lead to even more loss of energy. Therefore, PC seeks a balance between accuracy and speed in the internal model. Over time, the balance will nonetheless tip in favor of accuracy.

Predictive coding in religious cognition

Recently, predictive coding has been applied to religious cognition. In this section I will discuss predictive coding in religious cognition as it was proposed by Marc Andersen (2017). His theory differs significantly from other theories of religious cognition. Whereas other theories usually state that religious belief (in some rudimentary form) comes natural or intuitive, the PC-based theory states that religious belief is learned.

The theory received its most elaborate defense by Marc Andersen. He draws on two older approaches to religious cognition, Hyperactive Agency Detection (HADD) and neuroscientific study of religious experiences. I will briefly discuss both approaches. In the original theory, the operations of the agency detection device were held responsible for why people acquire religious beliefs.[5] Defenders argue that humans are prone to overdetect agents where there are none. People easily jump to the conclusion that some agent is around upon very limited or ambiguous evidence. Sounds of rustling leaves or a branch that vaguely resembles a snake suffice to conclude that some agent (usually a person or animal) is around. Being hyperactive in detecting agents was evolutionary beneficial. Animals and other humans posed (and often still pose) a serious threat to human survival. Predators or enemies can easily sneak up on humans, so it is safe to be on guard. Detecting too many agents only leads to limited loss of time and energy while detecting one agent too little can lead to instant death. Defenders of HADD argue that hyperactivity in the detection of agents can easily produce religious beliefs. Usually, people look for additional evidence to corroborate or discard their hunches that an agent is around. When they cannot find any evidence, the belief that an invisible agent is around can arise. This can in turn easily lead to belief in spirits or gods.

Marc Andersen notes that the original HADD-theory is not well supported by empirical data. According to Andersen, the original theory predicts that religious believers will be more prone to overdetect agency than non-believers. Studies did not unambiguously support this prediction. One study did find that that paranormal and religious believers reported more 'false alarms' compared to non-believers in detecting human-like faces in artifacts or scenery (Riekki *et al.,* 2013). Another study, however, found that paranormal believers were more accurate than non-believers in detecting faces in a face-house detecting task (Van Elk, 2015).[6]

Andersen's critique of the original HADD theory seems overstated. Defenders of the original HADD theory do not claim that religious believers have a more hyperactive agency detection device than non-believers nor that paranormal believers do. They rather argue that everyone has a hyperactive agency detection device that *could* under certain circumstances foster religious beliefs. They argue that religious believers and non-believers have the same experiences of agency, but the experiences do not lead to religious beliefs for non-believers. Justin Barrett argues that non-believers have various strategies (conscious or not) to prevent HADD-experiences from leading to religious beliefs (see Barrett, 2004, p. 112-115). The theory also allows for the possibility that non-believers never (or rarely) find themselves in situations where their agency detection devices are triggered without apparent agents around. In these cases, they would not from religious beliefs as a result.

The second related body of research on which Andersen draws is neuroscientific studies of experiences of an invisible sensed presence. A sensed presence is the sensation that someone or something living is present in the vicinity of the

[5] Variants of the theory were defended by Guthrie (1993), Atran (2002) and Barrett (2004). My discussion of HADD relies mainly on Barrett's work.

[6] In the study, visitors of a psychic fair were asked to calassify pictures with different levels of visual noise as representing a face or a house. Van Elk and his team concluded that paranormal believers were more accurate by comparing the number of falsely classified pictures (Van Elk, 2015).

subject. Andersen notes that people often have the sensation when no identifiable stimulus of a corresponding being or living thing is around.[7] Neuroscientist Michael Persinger argues that experiences of a sensed presence are caused by the effect of weak magnetic fields on the temporal lobe.[8] Persinger argues that the human brain contains a sense of the self in each of its hemispheres. Usually, the left hemisphere is dominant and supplies an individual with its everyday experience of self. When the right hemisphere is stimulated by electromagnetic force, the right hemisphere's sense of the self comes to intrude on the usual experience of self. As a result, it produces the experience of another self in the vicinity of the subject. To prove his claims, Persinger had a helmet (the God helmet) constructed that triggers the right temporal lobe with magnetic force. He claims that he successfully produced sensations of a sensed presence in subjects with this 'God helmet' (Booth *et al.*, 2005).

Persinger's theory has also been severely criticized. First, his framework for the self is not supported in conventional theories of cognitive neuroscience. Second, the magnetic stimulation from Persinger's God helmet would be too weak to elicit brain activity. Third, replication of Persinger's experiments with the God helmet failed. A Swedish team used Persinger's God helmet and whole setup to replicate his findings. In their experiment, they had a control group where the God helmet was switched off. The Swedish team found that the control group also reported sensations of a sensed presence. They concluded that the sensations were caused by suggestion rather than by magnetic stimulation (Granqvist *et al.*, 2005).

Although Andersen regards both approaches to religious cognition as flawed, he suggests that some insights can be incorporated in a new theory in line with predictive coding. I will call this theory 'PC HADD'.[9] He argues that people will be more prone to overdetect agents or feel a sensed presence when their internal model of the world expects more agents or persons around. He thereby accepts that humans can be hyperactive in their agency detection and that suggestion can lead to feelings of a sensed presence. When a subject believes that invisible agents are around, the belief will act as a top-down filter on how her sensory input is experienced. She will thus be more prone to identify noises or patterns as caused by an invisible agent. Making a subject aware that someone or something could be around can also act as a top-down filter on what that subject experiences.

Andersen concludes: "The human brain is not predetermined to select agent models to account for ambiguous stimuli. Instead, the models chosen to account for ambiguous sensory evidence rely heavily on context-sensitive subjective estimates of prior probability" (Andersen, 2017, p. 10) He argues that agency detection and experiences of a sensed presence are to a large extent the result of the context in which the subject finds herself. The context comes with certain expectations about what the subject could experience. The context in which a subject perceives and experiences is to some extent determined by her cognitive make-up,[10] but is to a large extent due to social and cultural factors. Social and cultural factors can foster what Andersen calls 'priors about supernatural agents.' One possible source of priors about supernatural agents are religious teachings and texts. Another source is guided processing of ambiguous stimuli. An example is how some Pentecostal Christians are taught to interpret certain bodily sensations as signs from God. Some Pentecostal Christians learn to identify a tingling feeling in their stomachs as caused by the Holy Spirit.[11] The predictive coding framework suggests that once a bodily sensation is identified with a cause (like God), prior beliefs will grow stronger. Andersen also notes that in many religious traditions believers are encouraged to engage in several forms of sensory deprivation. For example, Muslims and Christians are required to fast and some Hindus are encouraged to perform extreme rituals.[12] Andersen argues that these practices force the brain to rely more on priors than under normal circumstances.

Andersen also argues that the nature of many religious concepts makes it hard to filter them out of the internal model of the world. Many gods and spirits are believed to be invisible or even imperceptible. Concepts of gods and spirits are also flexible and evasive. Because of these features, religious concepts are impervious to the Bayesian tendencies to revise error.

Michiel van Elk and André Aleman apply the predictive coding approach to a number of religious phenomena (2017). They argue that religious hallucinations or visions are likely related to imprecise coding of predictive signals. According to Van Elk and Aleman mystical experiences, where the subject feels a loss of ego or identity can be explained by changes in the process of multisensory integration. Multi-sensory integration is the process whereby information from multiple modalities (sense, touch, vision…) is brought together to

[7] Andersen writes: "Sensed presence can manifest itself in a variety of ways, and is often experienced despite the absence of any identifiable stimuli corresponding to the experience" (2017, p. 6).

[8] Persinger set out his theory in a number of articles (Persinger, 1983, 2001; Persinger *et al.*, 2000). My discussion in this paragraph follows Andersen's discussion of Persinger's work (Andersen, 2017, p. 6-7).

[9] The term 'PC HADD' is somewhat confusing as Andersen considers his theory to be a criticism of the original HADD theory. I use the term because Andersen's theory resembles the original HADD theory to a large extent.

[10] Andersen refers to Boyer (2002), who argues that subjects easily remember and transmit concepts that are minimally counterintuitive, and to Whitehouse (2004), who argues that participation in religious rituals has a profound effect on the transmission of religious beliefs.

[11] Andersen got the example from Tanya Luhrmann's work (2012). Luhrmann did fieldwork in a Pentecostal church and noted that people were taught to identify thoughts as infused by God. She argues at length that the identification process requires time and instruction by experts.

[12] Xygalatas discussed the effects of extreme Hindu rituals in Mauritius on people's moral behavior (Xygalatas *et al.*, 2013)

form a coherent image of the body and the environment. Due to some changes, integration could go wrong and result in mystical experiences. They also argue that predictive coding can explain why people pray or sense a supernatural presence. Here expectations that govern social interactions are applied to invisible beings.

Van Elk and Aleman suggest that PC HADD has a very broad explanatory scope. The original HADD theory and other CSR-theories have been criticized for having a too restricted scope.[13] Jong *et al.* (2015) also complained that HADD is massively underdetermined by empirical data. Both problems were used to argue that CSR-theories do not have any epistemic implications. In the next section, I will look at the epistemic implication of PC HADD. I will assume that Andersen, Van Elk and Aleman are right and that PC HADD does not suffer from similar problems from which older theories suffer.

Epistemic implications of predictive coding

From early on, CSR attracted the attention of philosophers. Most of the discussion focused on the question whether CSR-theories implied a negative verdict on the epistemic status of religious belief.[14] The original HADD theory has been used to argue for and against a positive epistemic status of religious belief. Some of the arguments can be restated with the PC HADD theory. We noted above that the PC-HADD theory differs significantly from other theories of religious cognition. This section will also highlight how the new theory leads to conclusions that are different from those of other theories.

The arguments I discuss below argue for or against a positive epistemic status of *religious belief*. I use the term 'religious belief' to refer to belief in the existence of God or other supernatural being. Though the term covers more beliefs, it is used in this more restricted meaning in most of the existing arguments to which I refer.

Unreliability argument

A common line of argument in the debate over the implications of CSR-theories is that religious beliefs are unreliably formed and therefore unjustified.[15] A generic unreliability argument using CSR-theories goes as follows:

(1) CSR-theories show that the mechanisms responsible for religious belief are unreliable.
(2) Beliefs produced by unreliable mechanisms are unjustified.
(3) Therefore, religious beliefs are unjustified.

I will not contest premise 2, but some clarifications are in order. My use of the term 'unreliable' is in line with how the term is used in contemporary discussions over reliabilism.[16] Defenders of reliabilism argue that a belief should be produced by a belief-forming process that is not error-prone to be regarded as justified or to constitute knowledge.[17] Clear examples of error-prone processes are wishful thinking or clairvoyance. While these processes can occasionally produce true beliefs, they usually produce false beliefs. Therefore, the beliefs these processes produce are not justified even when they are true. I will not specify what the ratio between true and false beliefs must be before a process can be regarded as reliable. For my purposes, it suffices that the ratio of true beliefs must be well over 50 %.

I will now assess whether premise (1) holds if PC HADD is taken into account.[18] Andersen himself suggests that it does. He unambiguously states that detections of invisible agents are 'false positives.'[19] He writes: "My fundamental argument is that most false positives in agency detection can be seen as the result of top-down interference in a Bayesian system" (Andersen, 2017, p. 1). Andersen argues that the internal model of the world in the mind of religious believers makes them prone to detect (invisible) agents where there are none. He thereby suggests that detection of agents goes astray because the internal model

[13] See for example Leech and Visala (2011).

[14] The discussion is usually limited to the belief that God exists and does not include beliefs about the nature of God. For an overview of the discussion see McBrayer (2017) and Van Eyghen *et al.* (2018).

[15] Arguments for unreliability were defended by Braddock (2016), Wilkins and Griffiths (2013), Goodnick (2016) and Nola (2013). Wilkins and Griffiths, Goodnick and Nola do not conclude that religious beliefs are unjustified but respectively *undermined, unwarranted* and *debunked*. I lack the space to discuss the differences between these epistemic deficiencies and limit the discussion to justification. The argument I discuss in this paragraph can be reformulated so as to apply to other epistemic deficiencies.

[16] Reliabilist epistemologies state that a belief can be justified or amount to knowledge if it is produced by a reliable belief-forming mechanism (see: Goldman and Beddor, 2015).

[17] In this section, I focus on justification and not on knowledge. If the argument is sound, it will imply that religious beliefs (likely) cannot amount to knowledge since justification is usually seen as a necessary condition for knowledge.

[18] It should be noted that Andersen does not claim that religious belief in its totality, let alone all religious beliefs, are formed by the operations of the mechanism discussed in his theory. He does claim that the PC HADD mechanism is an important contributor to religious belief. If the arguments are true, Andersen's theory will have a major impact on the epistemic status of religious belief. How great the impact is depends on how large of a contributor PC HADD will turn out tob e. For reasons of brevity, I phrased the argument as if religious belief in its totality is formed by PC HADD.

[19] Van Elk and Aleman (2017) do not use this terminology in their discussion of PC.

is flawed.[20] Andersen also argues that the flaws are hard to overcome because of the elusive nature of many religious concepts. He argues that his theory, and especially his claim that a proneness to overdetect agency is hard to overcome, can explain the wide occurrence of religious beliefs in the human population. His claim implies that the flaw can be overcome with the result that the human brain no longer detects agents where there are none. In other words, the human brain can unlearn to overdetect agents and learn to only detect actual agents. Andersen suggests that this educated brain has a more accurate internal model of the world.

Andersen's line of reasoning appears to support premise (1). On his theory, the mechanism (partly) responsible for religious beliefs is the detection of invisible agents guided by the expectations of the internal model of the world. He claims that humans often have a mistaken internal model of the world, which makes it prone to detect invisible agents where there are in fact none. The mistaken detections result in religious beliefs. However, arguing for premise (1) in this way is dependent on an ontological commitment to naturalism. Andersen does not argue that an internal model without invisible agents or gods is more accurate but merely assumes it. He thereby assumes that the actual world contains no invisible agents or gods and thus that naturalism[21] is true. If naturalism is true, any mechanism that produces religious beliefs is unreliable. This shows that an argument for unreliability based on PC HADD ultimately depends on the truth of naturalism. I lack the space to discuss the philosophical debate over naturalism, but a quick glance at the literature shows that the debate is far from settled.[22] This shows that a mere assumption of naturalism will not do to establish premise (1). Without a well-supported premise (1) the conclusion that religious belief is unjustified does not follow.

Unsafety argument

Another argument claims that CSR-theories show that religious belief could easily be believed falsely.[23] I will call this the 'unsafety argument.' The theories allegedly show that mechanisms responsible for religious belief would still produce religious belief if nothing supernatural exists. For example, on the original HADD theory, people are prone to form beliefs in invisible agents even when no such agents are around. Therefore, if HADD were to produce true beliefs, its success is in a relevant way due to luck.

The idea that a belief should not easily be believed falsely is the main idea behind the safety condition for knowledge.[24] A belief is safe if there is no very close nearby world and not many nearby worlds in which the belief is false and the subject would still continue to form the belief in the same way. On the original HADD theory, it seems that the hyperactive agency detection device produces unsafe beliefs. The HADD would produce beliefs in invisible agents in many nearby worlds where there are no invisible agents because vague noises of patterns that vaguely remind of agency suffice to form these beliefs. An unsafety argument goes as follows:

(1) CSR-theories show that subjects will form religious beliefs in nearby worlds where nothing supernatural exists.
(2) Religious beliefs formed in a world where nothing supernatural exists are not true.
(3) Therefore, CSR-theories show that subjects will easily form religious beliefs that are not true.
(4) If subjects easily form religious beliefs that are not true, religious beliefs cannot amount to knowledge.
(5) Therefore, CSR-theories show that religious beliefs cannot amount to knowledge.

In what follows, I will not contest premise (4) (i.e. the safety condition for knowledge applied to religious belief).[25] Instead, I focus on premise (1). Before proceeding, it should be noted that discussing possible worlds where nothing supernatural exists runs into a problem. In classical forms of theism, God is regarded as a perfect being. Among other things, being perfect entails existing necessarily.[26] A being that exists necessarily exists in all possible worlds. Therefore, on this version of theism there is at least one supernatural being that exists in all possible worlds. Being necessary can, however, be cashed out in a different way. Richard Swinburne argues that God's existence should not be regarded as logically necessary. He argues that God's being is necessary in a weaker sense, namely that "if He exists at any time He exists at all times" (Swinburne, 1993, p. 274). On Swinburne's account, there are possible worlds where God does not exist and reasoning about possible worlds

[20] The flaw could result from evolutionary pressures as was argued in the original HADD theory. Andersen, however, suggests that the flaws are put in place by religious instruction and socialization.

[21] I take 'naturalism' to be the metaphysical theory that nothing supernatural exists.

[22] Notable recent criticisms of naturalism were raised by Rea (2002) and Plantinga (2002)

[23] A similar argument using other theories of religious cognition is discussed and criticized by Clark and Rabinowitz (2011).

[24] Varieties of the safety condition for knowledge were defended by Sosa (1999), Pritchard (2009) and Greco (1999). Williamson (2002) also discusses safety but does not regard it as a necessary condition for knowledge. Though all three accounts of safety differ in some regards (see: Rabinowitz, 2014), they share a commitment to the idea that a belief must not be easily believed falsely. In this section, I rely on Pritchard's account of safety.

[25] I note that the safety condition for knowledge has been criticized (e.g. Comesaña, 2005).

[26] For a deeper discussion see Plantinga (1974).

without any supernatural beings therein is possible. In this section I assume Swinburne's account.

If Andersen's PC HADD theory is true, it seems as if subjects will form religious beliefs in worlds without anything supernatural. It is plausible that in such worlds humans will have an internal model of the world with supernatural beings because of religious instruction or because having such a model is evolutionary beneficial. It is also plausible that the internal model will not be corrected easily because of the elusive nature of religious concepts.[27] However, it is not clear why there is religious instruction or socialization in worlds where nothing supernatural exists. Religious believers usually trace the origins of religious instruction back to a revelatory event or divine inspiration. In a naturalistic world, there can be no revelatory event or divine inspiration. To argue that there will be religious instruction and socialization without any such event, an account of where the instruction comes from is needed. Some authors have proposed accounts. Inspired by ideas of Sigmund Freud, some argued that religious systems were designed by rulers to keep their subjects in check. Accounts like these are not popular anymore. Alternative accounts are around. For example, Durkheim (1971) argues that religious instruction emerges from totemist practices. In totemist practices, groups gathered around a representation of an animal. The animal represented something the group stood for. Hunter-groups would be inclined to gather around an image of a wolf while fisher-groups would prefer a bear. Durkheim argues that groups in fact pay homage to themselves or to something they find important. These gatherings gave rise to rituals, which in turn gave rise to stories and beliefs.

I cannot properly assess Durkheim's theory here.[28] I do note that the other theories of religious cognition, like the original HADD-theory, did provide a fuller naturalistic account of the genesis of religious belief. It is clear that an unsafety argument needs a plausible naturalistic account of how religious instruction and socialization emerged. Moreover, this account cannot rely on falsely detected invisible agents, for such an account would be circular. Andersen does not provide such an account. Therefore, premise (1) is not obviously established by PC HADD.

PC HADD could support premise (1) with an evolutionary account of how religious beliefs show up in the internal model of the world. On this version of the theory, subjects in nearby worlds without supernatural beings with a similar evolutionary history and a similar cognitive make-up would (erroneously) form religious beliefs. However, without an evolutionary account (which Andersen did not provide) this remains speculative.

Naturalness argument

While most arguments based on CSR-theories conclude to a negative epistemic status of religious belief, some argue for the opposite conclusion. The arguments (I call them 'naturalness arguments') claim that CSR-theories speak in favor of religious belief because they show that religious belief comes naturally.[29] The term 'natural' should be read as synonymous with intuitive or spontaneous. The argument goes as follows:

(1) CSR-theories show that religious belief is natural.
(2) Beliefs that are natural are justified in the absence of defeaters.
(3) Therefore, religious belief is justified in the absence of defeaters.

The idea that religious belief is natural is widely affirmed by CSR-theorists, though not undisputed.[30] Theorists claim that various cognitive mechanisms and biases give rise to religious beliefs without any need for religious instruction or learning. They do not argue that religious belief is innate, but rather that it emerges spontaneously much like beliefs about good and evil emerge.[31] Clark and Barrett (2010) argue that beliefs that come naturally can be regarded as 'innocent until proven guilty.' They note that humans often rely on beliefs that come naturally, like beliefs about good and bad, and are justified in doing so. Natural beliefs can be revised or defeated, but this requires additional evidence. They argue that there is no reason to deny the same status to religious beliefs. Their line of reasoning is not without criticisms as well,[32] which I will not go into. For now, I will grant premise (2) and focus on premise (1).

It is not clear that premise (1) is true if PC HADD is plugged in. Andersen states that invisible agents and gods are mainly introduced in people's internal model of the world by means of religious instruction. If this is the case, religious belief are not natural. Reinforcement of religious beliefs by corroborating experiences still occurs without instruction, but this process depends on prior religious beliefs.

I argued earlier in the paper that invisible agents or gods could appear in people's internal model of the world because it was evolutionary advantageous. It is not implausible that humans are not a blank slate and that their default internal

[27] I assume that the belief-forming mechanism remains fixed in the nearby worlds. As Rabinowitz argues, this is in line with how safety is usually assessed. He argues that the belief-forming method should remain fixed when assessing the safety of a belief it produces (Rabinowitz, 2014).

[28] I do note that Durkheim's theory has been criticized for relying on outdated anthropological evidence.

[29] 'Naturalness arguments' were defended by Clark and Barrett (2011, 2010) and Braddock (2018).

[30] For criticisms see Shook (2017) and Banerjee and Bloom (2013).

[31] Barrett (2002) argues that children are 'born believers' like some people are 'born singers' or 'born artists'. They were not born as singers or artists, but were born with the necessary (cognitive) equipment to become good singers or artists.

[32] Adherents of evidentialism tend to deny that natural beliefs merit a positive epistemic status. See for example Feldman and Conee (2004).

model of the world includes gods or invisible agents.[33] If this is the case, religious beliefs are still natural.

Again, much depends on how invisible agents and gods ended up in the internal model of people. If they were included after religious instruction or socialization, religious beliefs cannot be regarded as natural. If they result from evolutionary pressures, they can. Future research might settle this question. For now, it is not clear whether religious beliefs are natural on PC HADD. As a result, premise (1) is not established if PC HADD is plugged in. Therefore, the argument is not sound.

Conclusion

The goal of this paper was to assess what epistemic implications new approaches to religious cognition that draw on predictive coding have. After discussing three arguments, the conclusion is that the theory, in its current form, is neutral. Arguments against a positive epistemic status require a strong case for naturalism or a naturalistic account for the origins of religious instruction. Both are not obvious. The new theory also does not obviously show that religious belief is natural and hence justified.

For all arguments, progress can be made by future research on predictive coding. Future research can shed light on how invisible agents and gods end up in the internal model of the world. It can also shed more light on the role of religious instruction and evolutionary pressures. Since the theory is still very young, new developments could very well bolster the case for any of the three arguments. Though the conclusions I drew after discussing each of the three arguments are not far-reaching, they show that philosophers of religion, theologians and anyone interested in the epistemic status of religious belief should keep up with new developments.

My discussion also shows that arguments based on older CSR-theories cannot always easily incorporate new theories. This is most obvious in naturalness arguments. While older CSR-theories appeared to support it, the new theory does not in an obvious way.

Acknowledgments

I thank Marc Andersen, Lieke Asma, Wout Bisschop, Lieven de Cock, Naomi Kloosterboer, Rik Peels, Chris Ranalli, Tamarinde van Haven and Rene van Woudenberg for their valuable comments on earlier drafts of this paper.

References

ANDERSEN, M. 2017. Predictive Coding in Agency Detection. *Religion, Brain & Behavior*, p. 1-20. https://doi.org/10.1080/2153599X.2017.1387170

ATRAN, S. 2002. *In Gods We Trust: The Evolutionary Landscape of Religion (Evolution and Cognition)*. Oxford, Oxford University Press, 400 p.

BANERJEE, K.; BLOOM, P. 2013. Would Tarzan Believe in God? Conditions for the Emergence of Religious Belief. *Trends in Cognitive Sciences*, **17**(1):7-8. https://doi.org/10.1016/j.tics.2012.11.005

BARRETT, J.L. 2004. *Why Would Anyone Believe in God?* Walnut Creek, Altamira Press, 141 p. (Cognitive Science of Religion Series).

BARRETT, J.L. 2012. *Born Believers: The Science of Children's Religious Belief*. New York, Free Press, 320 p. BOOTH, J.N.; KOREN, S.; PERSINGER, M.A. 2005. Increased Feelings of the Sensed Presence and Increased Geomagnetic Activity at the Time of the Experience During Exposures to Transcerebral Weak Complex Magnetic Fields. *International Journal for Neuroscience*, **115**:1053-1079. https://doi.org/10.1080/00207450590901521

BOYER, P. 2002. *Religion Explained: The Human Instincts that Fashion Gods, Spirits and Ancestors*. London, Vintage, 430 p. BRADDOCK, M. 2018. An Evidential Argument for Theism from the Cognitive Science of Religion. *In:* H. VAN EYGHEN; R. PEELS; G. VAN DEN BRINK, *New Develpments in Cognitive Science of Religion: The Rationality of Religious Belief*. Dordrecht, Springer. (New Approaches to the Scientific Study of Religious Belief), p. 171-198. https://doi.org/10.1007/978-3-319-90239-5_10

BRADDOCK, M. 2016. Debunking Arguments and the Cognitive Science of Religion. *Theology and Science*, **14**:268-287. https://doi.org/10.1080/14746700.2016.1191878

CHURCHLAND, P.M. 1990. On the Nature of Theories: A Neurocomputational Perspective. *Minnesota Studies in the Philosophy of Science*, **14**:59-101.

CLARK, A. 2013. Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *The Behavioral and Brain Sciences*, **36**(3):181-204. https://doi.org/10.1017/S0140525X12000477

CLARK, K.J.; BARRETT, J.L. 2010. Reformed Epistemology and the Cognitive Science of Religion. *Faith and Philosophy*, **27**(2):174-189. https://doi.org/10.5840/faithphil201027216

CLARK, K.J.; BARRETT, J.L. 2011. Reidian Religious Epistemology and the Cognitive Science of Religion. *Journal of the American Academy of Religion*, **79**:1-37. https://doi.org/10.1093/jaarel/lfr008

CLARK, K.J.; RABINOWITZ, D. 2011. Knowledge and the Objection to Religious Belief from Cognitive Science. *European Journal for Philosophy of Religion*, 3:67-81. https://doi.org/10.24204/ejpr.v3i1.381

COMESAÑA, J. 2005. Unsafe Knowledge. *Synthese*, **146**(3):395-404. https://doi.org/10.1007/s11229-004-6213-7

DIUK, C.; TSAI, K.; WALLIS, J.; BOTVINICK, M.; NIV, Y. 2013. Hierarchical Learning Induces Two Simultaneous, but Separable, Prediction Errors in Human Basal Ganglia. *Journal of Neuroscience*, **33**(13):5797-5805. https://doi.org/10.1523/JNEUROSCI.5445-12.2013

DURKHEIM, E. 1971. *The Elementary Forms of the Religious Life*. New York, The Free Press.

[33] The idea that people are not blank slates and are pre-equipped with a set of default cognitive inclinations and biases was a core idea of cognitive science of religion. Boyer (2002) is best known for arguing that the religious concepts people hold are constrained by the operations of the human mind.

FELDMAN, R; CONEE, E. 2004. Evidentialism. *In:* E. CONEE; R. FELDMAN (eds.), *Evidentialism: Essays in Epistemology*. Oxford, Clarendon Press, p. 83-108. https://doi.org/10.1093/0199253722.003.0005

FRISTON, K. 2010. The Free-Energy Principle: A Unified Brain Theory? *Nature Reviews Neuroscience*, **11**(2):127-138. https://doi.org/10.1038/nrn2787

GOLDMAN, A.; BEDDOR, B. 2015. Reliabilist Epistemology. *In:* E.N. ZALTA (ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2016. Metaphysics Research Lab, Stanford University. Available at: https://plato.stanford.edu/archives/win2016/entries/reliabilism/. Accessed on: 12/04/2018.

GOODNICK, L. 2016. A De Jure Criticism of Theism. *Open Theology*, **2**(1):23-33. GRANQVIST, P.; FREDRIKSON, M.; UNGE, P.; HAGENFELDT, A.; VALIND, S.; LARHAMMAR, D.; LARSSON, M. 2005. Sensed Presence and Mystical Experiences Are Predicted by Suggestibility, Not by the Application of Transcranial Weak Complex Magnetic Fields. *Neuroscience Letters*, **379**:1-6. https://doi.org/10.1016/j.neulet.2004.10.057

GRECO, J. 1999. How Must Knowledge Be Modally Related to What Is Known? *Philosophical Topics*, **26**(1/2):373-384.

GUTHRIE, S.E. 1993. *Faces In the Clouds: A New Theory of Religion*. New York, Oxford, Oxford University Press.

HAWKINS, J.; BLAKESLEE, S. 2004. *On Intelligence*. New York, Owl Books/Times Books.

JONG, J.; KAVANAGH, C.; VISALA, A. 2015. Born Idolaters: The Limits of the Philosophical Implications of the Cognitive Science of Religion. *Neue Zeitschrift* für Systematische Theologie *und Religionsphilosophie*, **57**:244-266. https://doi.org/10.1515/nzsth-2015-0012

LEECH, D.; VISALA, A. 2011. The Cognitive Science of Religion: A Modified Theist Response. *Religious Studies*, **47**:301-316. https://doi.org/10.1017/S0034412510000302

LUHRMANN, T.M. 2012. *When God Talks Back: Understanding the American Evangelical Relationship with God*. New York, Vintage, 335 p. McBRAYER, J. 2017. Explanations of Religion. *Oxford Bibliographies*. Available at: http://www.oxfordbibliographies.com/view/document/obo-9780195396577/obo-9780195396577-0358.xml. Accessed on: 23/03/2018.

NOLA, R. 2013. Do Naturalistic Explanations of Religious Beliefs Debunk Religion? *In:* G.W. DAWES; J. MACLAURIN (eds.), *A New Science of Religion*. New York/London, Routledge, p. 162-188. (Routledge Studies in Religion).

PERSINGER, M.A. 1983. Religious and Mystical Experiences as Artifacts of Temporal Lobe Function: A General Hypothesis. *Perceptual and Motor Skills*, **57**:1255-1262. https://doi.org/10.2466/pms.1983.57.3f.1255

PERSINGER, M.A. 2001. The Neuropsychiatry of Paranormal Experiences. *The Journal of Neuropsychiatry and Clinical Neurosciences*, **13**(4):515-524. https://doi.org/10.1176/jnp.13.4.515

PERSINGER, M.A.; TILLER, S.G.; KOREN, S.A. 2000. Experimental Simulation of a Haunt Experience and Elicitation of Paroxysmal Electroencephalographic Activity by Transcerebral Complex Magnetic Fields: Induction of a Synthetic 'Ghost'? *Perceptual and Motor Skills*, **90**(2):659-674. https://doi.org/10.2466/pms.2000.90.2.659

PLANTINGA, A. 1974. *The Nature of Necessity*. Oxford, Clarendon Press, 255 p. (Clarendon Library of Logic and Philosophy).

PLANTINGA, A. 2002. Introduction: The Evolutionary Argument Against Naturalism. *In:* J.K. BEILBY (ed.), *Naturalism Defeated? Essays on Plantnga's Evolutionary Argument Against Naturalism*, Ithaca, Cornell University Press, p. 1-14.

PRITCHARD, D. 2009. Safety-Based Epistemology: Whither Now? *Journal of Philosophical Research*, **34**:33-45. Available at: https://www.pdcnet.org//pdc/bvdb.nsf/purchase?openform&fp=jpr&id=jpr_2009_0034_0033_0045&onlyautologin=true. Accessed on: November 27, 2018.

RABINOWITZ, D. 2014. The Safety Condition for Knowledge. *In: Internet Encyclopedia of Philosophy*. Available at: http://www.iep.utm.edu/safety-c/. Accessed on: 13/02/2018.

REA, M.C. 2002. *World Without Design: The Ontological Consequences of Naturalism*. Oxford, Clarendon, 246 p. https://doi.org/10.1093/0199247609.001.0001

RIEKKI, T.; LINDEMAN, M.; ALENEFF, M.; HALME, A.; NUORTIMO, A. 2013. Paranormal and Religious Believers Are More Prone to Illusory Face Perception than Skeptics and Non-Believers. *Applied Cognitive Psychology*, **27**:150-155. https://doi.org/10.1002/acp.2874

SHOOK, J. 2017. Are People Born to Be Believers, or Are Gods Born to Be Believed? *Research Gate*, **29**(4). Available at: https://www.researchgate.net/publication/312371406_2017_Shook_-_Are_People_Born_to_be_Believers_or_Are_Gods_Born_to_be_Believed. Accessed on: 01/02/2018.

SOSA, E. 1999. How Must Knowledge Be Modally Related to What Is Known? *Philosophical Topics*, **26**(1/2):373-384. https://doi.org/10.5840/philtopics1999261/229

SWINBURNE, R. 1993. *The Coherence of Theism*. Oxford, Clarendon Press/Oxford University Press, 312 p. https://doi.org/10.1093/0198240708.001.0001

VAN ELK, M. 2015. Perceptual Biases in Relation to Paranormal and Conspiracy Beliefs. *PLOS ONE*, **10**(6):e0130422. https://doi.org/10.1371/journal.pone.0130422

VAN ELK, M.; ALEMAN, A. 2017. Brain Mechanisms in Religion and Spirituality: An Integrative Predictive Processing Framework. *Neuroscience and Biobehavioral Reviews*, **73**(Feb.):359-378. https://doi.org/10.1016/j.neubiorev.2016.12.031

VAN EYGHEN, H.; PEELS, R.; VAN DEN BRINK, G. 2018. Introduction. *In:* H. VAN EYGHEN; R. PEELS; G. VAN DEN BRINK (eds.), *New Developments in the Cognitive Science of Religion: The Rationality of Religious Belief*. Dordrecht, Springer.

WHITEHOUSE, H. 2004. *Modes of Religiosity: A Cognitive Theory of Religious Transmission*. Rowman, Altamira.

WILKINS, J.S.; GRIFFITHS, P.E. 2013. Evolutionary Debunking Arguments in Three Domains: Fact, Value, and Religion. *In:* G. DAWES; J. MACLAURIN (eds.), *A New Science of Religion*. London, Routledge, p. 133-146.

WILLIAMSON, T. 2002. *Knowledge and Its Limits*. Oxford, Oxford University Press, 352 p. https://doi.org/10.1093/019925656X.001.0001

XYGALATAS, D.; MITKIDIS, P.; FISCHER, R.; REDDISH, P.; SKEWES, J.; GEERTZ, A.W.; ROEPSTORFF, A.; BULBULIA, J. 2013. Extreme Rituals Promote Prosociality. *Psychological Science*, **20**:1-4. https://doi.org/10.1177/0956797612472910

Submitted on September 4, 2018
Accepted on September 21, 2018