Demystifying Religious Belief

Robert Nola, University of Auckland

The cognitive science of religion (CSR) comprises a large research programme which has grown markedly since its inception in the 1980s. To get a better grip on it, we can split it into a 'CS' bit and an 'R' bit. First, the 'R' bit stands for all the world religions, past and present, with all their differing characteristics such as: religious experiences (those commonly described by mystics (such as ecstatic states), "out-of-body" experiences, "near-death" experiences, mediums and their messages from beyond the grave, etc.); rituals and behaviours (e.g., genuflexion, praying, confessing, tithing, sacrificing (human or animal), fasting, singing songs, etc.); beliefs such as that God (or divinities) exist, that we have a soul which has a pre- and/or after-life, that the Jews are responsible for death of Christ, and so on. The term 'religion' does not denote a natural kind of "thing"; nor does it refer to anything with an "essence" which all the various religions share. This is so for the major world religions such as Christianity, Islam, Hinduism, etc. and their sects. Each religion is best "defined" by an indefinitely large cluster of characteristics drawn from the sort mentioned above; or as a "Wittgensteinian" set of family resemblances based on such characteristics. What stands in need of explanation are not only the various religions (or sects of them) with their distinctive clusters of characteristics, but also each of the characteristics themselves. These are the explananda R for the programme of CSR. This suggests two ways of considering the 'R' bit. First, there is the generic umbrella term 'religion' the scope of which is the various world religions (and their sects) both past and present. Second, each of these various religions can be specified by an (indefinite) cluster of characteristics, or a "family resemblance" set of characteristics. Thus religion is not a "natural" kind, it lacks any essence and the various religions need not share any common characteristic (though some characteristics, like belief in divinities, might be found in many religions).

Second, the 'CS' bit concerns the cognitive sciences which comprises at least the sciences of evolutionary and cognitive psychology. CS is a source of explanatory hypotheses for a science of religion and their application in particular contexts. These hypotheses are the various *explanans*. And what they explain are the *explananda* mentioned above, viz., the two broad kinds of item which comprise R. Third, there is also some kind of explanatory relationship which links the proposed *explanans* provided by CS to the various *explananda* that comprise R. Here we will not devote much time to what this explanatory relationship might be; most models of explanation current in the philosophy of science will serve our purpose. But note that the explanatory success of hypotheses can come in degrees and competing hypotheses may be assessed on the basis their degree of explanatoriness (as in the case of arguments which adopt a form of Inference to the Best Explanation – see section 4). Finally, it should be noted that not only the resources of CS can be called upon to explain some R item. The full programme of the *scientific* explanation of religion is broader than CSR and can call upon explanatory hypotheses which are not obviously cognitive in nature; examples of these sciences are cultural studies, anthropology, political science, and the like. However in this paper the kinds of *explanans* will be restricted to those drawn only from CS.

Those who study religion scientifically are rightly exercised over what the scope of religion R might be, viz., the scope of what is to be explained. Also important are the kinds of hypotheses that they can employ to explain some chosen bit of R. When looked at from the account of the term 'religion' sketched above, there is strictly no such object "religion" to be explained; instead there are the many different world religions with their distinctive clusters of characteristics. Here I will set aside the important issues which these classificatory problems raise and simply adopt the broad classification of aspects of CSR set out in White (2017). White also offers what she calls five key ideas which minimally constitute the field of CSR studies. Important for my purposes is the third which says: 'To explain religion we must first fractionate and reduce it into meaningfully constituent parts'. Here I will adopt the idea of fractionation in which just one aspect of religion R is selected for explanation, and a small set of hypotheses from CS are proposed to do the explaining and are compared for their explanatory success. Section 1 of the paper sets out one pervasive characteristic of nearly all religions as the item to be explained, viz., a person's belief in a divinity or a God; this is part of what is called the "folk" conception of religion.[1] This section also considers a range of hypotheses which might explain this aspect of religion; some of these are drawn from CS but others are not, such as those proposed by Freud. Section 2 of the paper fractionates further and focuses on just one kind of explanatory hypothesis from evolutionary psychology - that provided by the postulation of agency detection devices, such as HADD.

Writers within CSR are somewhat divided over whether or not, if hypotheses like HADD do explain (to some extent and degree) aspects of religion such as belief in divinities, then religious beliefs are thereby explained away or debunked. Thus the editors of a recent collection of papers concerning CSR ask in their introductory chapter: ''Does CSR and allied evolutionary approaches to the study of religion present epistemic challenges to religious belief, or support?' (Trigg and Barrett, 2014, p. 1).[2] The collected papers provide a variety of considerations on this matter. In section 3 we investigate whether or not the supposed existence of HADD as a cause of religious believing provides a debunking of religious belief. There is a *prima facie* case for saying it does. The final section 4 takes a different approach to these issues. It casts the net more widely and asks whether the kind of naturalistic explanations (exemplified by HADD but not only HADD) provide better explanations of belief in divinities than does the folk explanatory schema of traditional folk religion. Controversially, the answer is 'Yes'.[3] The argument here draws upon standard criteria, proposed in the philosophy of science, for comparing rival scientific hypotheses within competing frameworks, or competing "world views".

1. Two Rival Theories for Explaining Religious Belief

There is an evident widespread belief, held by people in most current cultures and throughout much of each culture's history, in the existence of spiritual entities, divinities, gods (within polytheism), a single God (of monotheism), and the like. The core belief that one or more such entities exist is to be distinguished from more religiously specific beliefs such that, for example, God is a Trinity, God sacrificed his only Son to remove the burden of "original sin" from humanity, or God rewards martyrs with a large number of virgins in the next world for their sexual pleasure, and so on. Setting aside these culturally specific beliefs we can ask: 'Why are the core existential beliefs so widespread?' This paper will focus only on this single aspect of religion that stands in need of explanation.

Believers themselves have their own explanations constructed from within their own religious framework as to why they have such core beliefs, viz., that one or more of the entities in which they believe really do exist and that these entities are essentially involved in any explanation of why people believe in them. Here the religious framework is taken at face value to be *sui generis* and is not to be explained, or explained away. Exactly how these explanations are to be constructed is a matter often left to theologians and philosophers of religion to work out; ordinary believers may have no theory of this sort. But central to this view is the counterfactual that if there were no God (or divinity) then there would be no human belief in it. (Note that those who have a rival view might hold that even if there were no God we humans might still have beliefs in God; other casual processes are at work, as will be seen.) What the particular causal processes are between God and human believings that God exists can be left open (though the story told about HADD discussed later might be invoked in such a causal explanation).

Call this explanatory stance "internalist" since explanations of believings invoke the entities of the religious framework in which people also believe. This "internal" stance yields what will be called here the "folk" view of religion of which there are at least the following three aspects.

(i) a set of core beliefs in a divinity or divinities.
(ii) a commitment to the existence of such divinities. (Though (i) can be taken, in the case of believers, to entail (ii), here (ii) is set out explicitly.)
(iii) at least the bare claim that the divinities explain why people have the core beliefs they do (whatever further details the explanation might have).

The internal stance of the "folk" view is not the only explanatory framework available. There are also "external" stances of those who, like anthropologists or advocates of CSR, might wish to study various belief systems without necessarily being committed to them. That is, they "bracket off" any existential commitments of the "folk" view; or more pointedly they consider them to be false. The two stances can be contrasted in the following way. In the case of (i) above, both stances agree that there are beliefs in the existence of divinities to be explained. However externalists adopt a different (ii*) from the internalists' (ii) and downplay, remain neutral about or deny the existential commitment to such divinities indicated in (i). Finally in place of (iii) they claim (iii*): explanations of why there are such core beliefs (as in (i)) are not to be given in terms of divinities but something distinctly non-divine. Overall they eschew the internalist framework.

Clearly naturalists are externalists (but not conversely as some externalists might simply "bracket off" or remain neutral about, the commitments of (ii)). Whatever else the doctrine of naturalism might be, at least it rejects any commitment to the supernatural, such as divinities, spirits, gods, and the like in the explanations it offers. Naturalists will agree with folk theorists that there is a common item to be explained, viz., core *beliefs* in the existence of divinities (call these 'B'); but they will disagree about the existence of such divinities. Consequently they will also disagree about the existential presuppositions of the framework of explanations internalists employ. The "folk" theory and any naturalistic theory will be

rival theories differing not only in their ontological commitments but also in their explanatory hypotheses. However the rival "folk" religious theories and any one of various naturalistic theories share the same *explanandum* in that both attempt to provide explanations of B. What needs to be determined is which offers the better explanation (a matter addressed in section 4).

The above can be made more specific in the following way. Suppose the various items in the core set of beliefs, B, that are to be explained (the *explanandum*) are of the form 'x believes that D exists' (where 'x is a person or believer, and 'D' is placeholder for various divinities that we might suppose exist as expressed in the propositions *that gods exist*, or *that God exists*, or *that spirits exist*, etc.). Importantly, the *explanandum* is an act, or state, of believing, viz., *x believes that God exists*; it is not the contained proposition '*God exists*' which is to be explained. Given this distinction it follows that any explanation of *acts* of believing such as 'x believes that God exists', will not serve as an explanation of what the contained proposition purports to be about, viz., that God exists. It is an important feature of the concept of belief that the *explanandum* 'x believes that God exists' is logically independent of the truth or falsity of the contained proposition 'God exists'. This further underlines the claim that there is no logical link between explanations of acts of believing and explanations of what is believed to be so. Thus there are two distinct "objects" of explanation: an act of believing; and some alleged fact which would be the truth-maker of the content believed. If we now understand 'B' to be about acts of belief (and not the contents of beliefs or their purported truth-makers) then what we need to consider is which of the "folk" theory or rival naturalistic theories best explains acts of believing of the form: x believes that D exists.

The distinction made here is of a piece with Hume's two different enquiries into religion. In the 1750's Hume drafted two books on religion. In 1757 he published one of these as *The Natural History of Religion*. In its 'Introduction' he distinguishes his two enquiries into religion, viz., 'that concerning its foundation in reason, and that concerning its origin in human nature' (Hume 1993, 134). The *Natural History* concerns just the second, naturalistic inquiry into the origins of religious belief in human nature. The first enquiry into the foundation of religion in reason appeared posthumously in 1779 as *Dialogues Concerning Natural Religion*. In this work Hume's concern is whether or not there is any good argument for the truth of the proposition that God exists. As is well known Hume remained ironically sceptical of any attempt to establish its truth. In the light of the distinction made in the previous paragraph, we can say that one of Hume's enquiries is focused upon establishing the truth or falsity of a proposition, viz., *that God exists*. Despite Hume's negative verdict, this enquiry is still actively pursued in the philosophy of religion. Hume's other naturalistic enquiry is importantly different since it concerns the causes of our acts of believing in God to be found in human nature. It does not concern (directly) the truth value of a propositional content, *that D exists*, but rather a hypothesis about the causes of an act of x's believing that D exists. This latter is the domain of modern CSR and its attempts to explain religion.

Hume's naturalistic causal hypothesis can be found scattered throughout *The History* (especially the first three sections) and can be summarised as follows: x's fears anxieties and dreads (due to a hostile world), x's unsatisfied appetites (for life's necessities) and x's terrors (about dying and death) collectively or singly cause x's (act of) belief that God exists. For example, after describing the existential anxieties in which most people live he proposes the casual hypothesis: 'in this disordered scene, with eyes still more disordered and astonished, they see the first obscure traces of divinity' (*ibid*., 140). Hume also supposes that to these divinities we ascribe 'sentiment and intelligence' and '. . . thought and reason and passion, and sometimes even the limbs and figures of men, in order to bring them nearer to a resemblance with ourselves' (*ibid*. 142).

Hume's hypothesis about the causes of religious belief is a version of *anthropomorphism*; alternatively some talk of *projectionism* though Hume does not use either term. But the 'projection' metaphor is suggestive. Certain states of mind (such as anxieties about our human condition) are the cause not only of core acts of belief in divine agents but also our ascribing certain characteristics to them. Further, these beliefs become a "projection" onto the world and thus are taken to be true of the world, though on the Humean casual hypothesis there are no grounds for this supposition. Humean projectionism is non-realist in the sense that the "objects" which our beliefs project onto the world are not strictly there (in much the same way, as in the projection of films, the "objects" projected onto a screen are not always to be strictly taken as part of the real world (though they are still to be counted as "images" on a screen)).

Hume's account suggests the following general schema for a number of naturalistic causal hypotheses. Where 'Φ' stands for some psychological state of a person x and 'B' stands for x's act of believing that some divinity exists, then the "Humean schema" is:

x's Φ-states cause x's B-states.

The task now is to find some appropriate Φ-state as a cause of B-states. In the case of Hume himself, the Φ-states can be filled in by some disjunction of our anxieties dreads, hopes, fears, etc, about our human condition.[4] In the case of Freud the Φ-states are unconscious wishes for a providential (father) figure which remain unfulfilled. Though I would leave it to scholars, Feuerbach is also said to be a projectionist

with respect to God; the Feuerbachean Φ-states are aspects of our mental dispositions such as love, will and reason, which are then projected onto the world in the form of a divinity which also has these dispositions. An important Humean influence can be found in the anthropologist Stewart Guthrie, who proposes '… that religion may best be understood as systematic anthropomorphism: the attribution of human characteristics to non-human things or events' (Guthrie 1993, p. 3). Here the Φ-states are our various acts of anthropomorphism which produce B-states concerning God, or various gods. Though Guthrie's overall thesis is an important source of hypotheses within CS for explaining religious belief, it will not be considered here.

Finally evolutionary cognitive psychologists would fill in the Φ-states by postulating cognitive mechanisms the operation of which generate B-states about seen and unseen divine agents. Developments of this theory will be the focus of the rest of this paper. It will consider not only which of the "folk" or some naturalistic theory provides the better explanation of B (in section 4), but also (in section 3) how reliable the cognitive mechanisms are in generating true beliefs (the answer to which is 'not very').

These five examples of the Humean schema yield hypotheses which are not only naturalistic in character but they are also open to scientific testing. Just how well these hypotheses (or more refined versions of them) fair under evidential test is part of the scientific project of explaining religion. In this paper matters to do with evidence will be set aside; rather the focus is upon what kind of explanation the various hypotheses provide of the core religious beliefs held by people. Of course we would not want to have explanations whose main casual hypotheses fail tests. But we can look for *potential* explanations in the sense of Hempel (1965, 338). *Actual* explanations are those in which what does the explaining is true or there is good evidence for their truth. In contrast a *potential* explanation is one in which the following counterfactual holds: if what did the explaining *were* to be true then a satisfactory explanation *would* be provided. Here the focus will be on potential explanations the hypotheses provide and not the equally important issue of their evidential support.

2. **Agency Detection Devices and their Reliability: From ADD to HADD**

That humans attribute agency, correctly or wrongly, to a wide variety of items in the world is a claim of long-standing, pre-dating even Hume's supposition of it in his own naturalistic approach to religion. Darwin also takes the attribution of agency to be an important source for the core belief in the existence of divinities:

> The tendency in savages to imagine that natural objects and agencies are animated by spiritual or living essences, is perhaps illustrated by a little fact which I once noticed: my dog, a full-grown and very sensible animal, was lying on the lawn during a hot and still day; but at a little distance a slight breeze occasionally moved an open parasol, which would have been wholly disregarded by the dog, had any one stood near it. As it was, every time that the parasol slightly moved, the dog growled fiercely and barked. He must, I think, have reasoned to himself in a rapid and unconscious manner, that movement without any apparent cause indicated the presence of some strange living agent, and that no stranger had a right to be on his territory. The belief in spiritual agencies would easily pass into the belief in the existence of one or more gods. (Darwin 1882, 95)

Darwin's talk of his dog's attribution of agency to the moving parasol in terms of 'reasoning in a rapid and unconscious manner' is of a piece with Kahneman's distinction between "fast" and "slow" thinking of what he calls 'Systems 1' and 'Systems 2' located in our minds (Kahneman 2011, chapter 1). Darwin's dog is a "fast" thinker exploiting "quick and dirty" cognitive mechanisms characteristic of those found in System 1. Unlike us, Darwin's dog does not indulge in the "slow" thinking of System 2 which would be characteristic of someone interested in Hume's first project of attempting to found religious belief in reason. Evolutionary cognitive scientists have developed, in an interesting way, Darwin's suggestion that a System 1 mechanism is used in the detection of agency and this leads to core beliefs in the existence of divinities.[5]

Their hypothesis is that creatures which predate, and/or who are predated, evolved one or more cognitive devices for the fast detection of predators and prey. Suppose a creature avoids predators upon detecting them when they are present either by initiating evasive behaviour directly or through the mediation of a belief something like "predatory agent present" which is then quickly followed by evasive behaviour. Such a creature will survive for another day. But if in the presence of a predator the creature fails to detect it and initiates no evasive behaviour, then it makes a serious Type II error (a false negative) which can be fatal. Alternatively suppose that there is no predator present and the creature still initiates evasive behaviour (or forms a belief which leads to that); then the creature will have wasted time and energy in unnecessary evasion. This is a Type I error (a false positive) which can be costly but not as

costly as the Type II error. Finally suppose there is no predator present and the creature does not initiate any evasive behaviour; there is no cost in this situation and life goes on much as normal.

Error Management Theory[6] (EMT) casts an interesting light on the direction that evolution will take in order for creatures to avoid the hugely fatal cost of Type II errors. Suppose there is an inequality in which the costs associated with Type I errors are not as great as the costs of generally fatal Type II errors. This allows for the evolution by natural selection of creatures which possess predator detection devices which allow for such an inequality. Evolutionary processes can even "fine tune" the detection devices so that they vastly minimize Type II errors while allowing for less costly Type I errors. For example, fire detection alarms could be so tuned that they hardly ever fail to detect a fire in one's house; but this might be at the cost of the alarm going off, harmlessly but annoyingly, when there is no fire present. Suppose one wished to purchase a fire alarm. One might like a perfectly reliable alarm which sounds if and only if there is a fire; but there may be no such alarm available. However there may be a "next best" alarm which has an extremely low probability of not sounding when there is a fire but has a higher probability of sounding when there is no fire. Though not ideal this would still be a good alarm to purchase if it is the only kind available.

If we now treat predators and prey as agents then there are a number of ways in which an evolved agent detection device (ADD) can be unreliable. The *first* source of unreliability is due to EMT; an evolved ADD can be "tuned away" from fatal Type II errors and in the direction of Type I errors which are much less costly but which can generate a larger number of false positives. Thus an ADD may serve a predated tree-dwelling creature quite well when it detects snakes in a tree[7] (with high probability) and so initiates evasive behaviour; but it does not serve it so well when it detects some twigs on a tree which are snake-like and also initiates evasive behaviour. Six further sources of unreliability are suggested in what follows.

Taking the seven sources of unreliability together, there is a case for considering ADDs to be hypersensitive; they are readily triggered to indicate the presence of agents when none are present. In the light of this many theoreticians speak of "hyperactive" or "hypersensitive" agency detection devices. Hence the now common acronym, HADD (coined in Justin Barrett (2000)). The journey from the postulation of ADDs to the postulation of HADDs is already part of the story concerning the first kind of unreliability of our detection devices due to EMT.

The cognitive psychologist H. Clark Barrett proposes a theory of what ADDs might be like for the ancestors of ourselves which evolved ADDs long before they became human. He set out five salient features of what are rather complicated devices involved in detection (Barrett (2005, pp. 203-4):

(1) a perceptual triggering system designed to reliably detect agents (and to discriminate agents from non-agents),
(2) perceptual mechanisms for discriminating between different kinds of agent (e.g., humans versus lions),
(3) perceptual mechanisms for discriminating between possible intentions and behaviours of particular agents (e.g., attacking versus fleeing),
(4) an inferential apparatus that includes amongst other things, to "mind read" and to adopt the intentional stance with regard to agents to make inferences about their behavior,
(5) 'a variety of systems that modulate the agency system proper … The output of the system is in the form of inferences, judgements, and, ultimately, behavioural decisions.

Note that in (4) the term 'intentional stance' is used by Dennett (2006, pp. 108-114) to refer to what others call 'a theory of mind' (ToM). On Barrett's schema, predator-prey detection devices will have a ToM built into them.

Barrett has a lot more to say about each of the items. But this suffices to indicate minimally what a predator-prey detection devices such as ADD are like. Clearly it would be wrong to call ADD a module of any sort; it is more complex than that. Moreover a creature could possess many different kinds of ADDs based on different kinds of perceptual input, or reactions to predator detection. Importantly as Barrett says: 'we carry with us sophisticated perceptual and inferential machinery for dealing with predators and prey, despite the fact that most of us will rarely, if ever, use it for the purposes for which it evolved' (*ibid.*, p. 200). Recent theories within CS adopt an evolutionary "by-product" view and claim that we now hardly use our inherited ADDs for the detection of predators and prey – their original evolutionary use. Rather we use ADDs more generally outside their original evolutionary context for the "detection" of agency. In particular it is hypothesized to be used for the "detection" of the unseen divine agents of religions. Hence a current role for ADDs which still "indicate" the presence of agents, even when none might be present.

Barrett's five features of ADDs involve further kinds of unreliability which take us in the direction of HADDs. (1) requires creatures to perceptually detect the difference between twigs of a tree (non-agents) and snakes in trees (agents). And (2) requires creatures to tell the difference between rabbits

which are not a threat, and snakes which are; or between human conspecifics which might be a threat and lions which definitely are. These perceptual "judgements" about items in our perceptual field have to be made quickly and accurately, perhaps based only on a quick glimpse. Such perceptual judgements are prone to error along the lines indicated by EMT. Even with the high costs of Type II errors, there still remain considerable costs associated with the false positives of Type I errors due to being over-vigilant and adopting evasive strategies when twigs or rabbits are perceptually encountered (either correctly or mistakenly). These considerations lead to a *second* source of unreliability for HADDs due to the workings of perceptual systems.

A *third* source of unreliability arises as follows. Often a predator is not directly observed but remains unobserved and is inferred from various kinds of evidence; so the predated need to make an "inference" to the presence of, say, a tiger from evidence based on, say, the distinctive movement of the grass in which it lurks. In this case the possessor of HADD needs to make the probability judgement that evidence E (e.g., particular movement of the grass) is more likely given the presence of a predatory agent than in the predator's absence (and, say it is the wind causing the movement, or whatever). More formally the likelihood claim is: p(E, tiger) > p(E, no-tiger). Thus a superior HADD needs to be equipped with an "inferential device" which makes a likelihood judgments and then compares them for the purposes of action.

The inference to the unobserved predator is based on comparative likelihood judgments and so is non-deductive; as is well known, such non-deductive inferences from evidence need not always be highly reliable. This is especially the case for HADD which is a "quick and dirty" inference maker of Kahneman's 'System 1'. HADD may well cause a belief that a predatory agent is present but it need not do this with a high degree of reliability for the truth of the claim that an agent is actually present. This turns on the distinction drawn in the previous section between 'x believes that A is present' and 'A is present'. All that is required is that, in the long run, the number of Type II errors about beliefs be zero or close to zero while the number of Type I errors can be larger (but not so large as to be prohibitively costly). For example, upon acquiring evidence, e.g., that there is a track of bent and broken grass nearby, it would be better to infer that a (predatory) agent is present rather than not, and so be alerted even if there is no agent present; but not to be so alerted when an agent is present is to be subject to a very costly Type II error. This gives rise to an illusion of agency; agency can be projected onto the world when no such agent exists.

This leads to a *fourth* way in which HADD can be unreliable. As will be seen shortly, HADD is not only said to yield as output 'agent present' when the agent is not observed (but could be observed), but also when an agent is unobservable (i.e., cannot be observed either because it makes itself invisible but could appear to some[8], or more strongly is in principle unobservable). Examples of unobservable agents are spirits, divinities and gods. In such cases HADD comes equipped with an encapsulated inference maker within System 1 that makes "quick and dirty" likelihood judgements to the unobservable. But such inferences are also well-known not to be reliable; they inherit all the unreliability of inductive inferences that are made quickly to unobservables. In such cases it would not be correct to think of HADD as a "detection" device since 'detect' is a success word; we detect items which exist and do not detect what is not there. It might be better to think of a HADD as a *postulator* of agents for which the acronym HAPD is more appropriate, though less catchy. Thus HADD as a postulator of unobservables with its encapsulated inference-making device introduces a new dimension of unreliability when it comes to generating beliefs about the existence of unobservable agents (in particular divinities). Along with others, Justin Barrett (see Barrett 2000, 2004 and 2009) is a prominent cognitive psychologist who has extended the operation of HADD, in the direction Darwin suggested, viz., to provide an account of how we come to believe in divinities where these are on the whole unobservable but we infer to them on the basis of something we can observe.

There are three further ways in which HADD can be unreliable. Clark Barrett in his items (3) and (4) attributes an ability to the predated to "mind read" what a predator is up to. That is, HADDs are able to adopt a Dennettian intentional stance; or as we will say, they adopt a "Theory of Mind" (ToM) concerning the predator. As (4) makes clear, such "mind reading" will involve inference to the behaviour of the predator in the immediate future, e.g., it is going to attack. Such "mind reading" based on inferences is clearly not perfectly reliable; so the postulation of a ToM for HADD introduces a *fifth* kind of unreliability for HADD and what it detects (future behaviour).

A *sixth* kind of unreliability for HADD emerges when we consider how HADD operates outside the context in which it originally evolved and is later "co-opted" for purposes other than predator detection. Suppose that a creature which possesses HADD undergoes, through evolution, further cognitive development of the following sort. It develops a more expanded theory of mind, ToM, which enables it to attribute beliefs and desires to other creatures around it (and not only its conspecifics). It also develops a language L, which is sufficiently conceptually rich; in particular it is at least able to

express in language the beliefs and desires it attributes to other creatures (perhaps based on its own desires and beliefs). It also develops the capacity, using L, to tell stories S about the world in which it lives.

Some of these stories might be about what are sometimes called "minimally counterintuitive" agents, MCIs (see Justin Barrett 2004). These are agents in stories which are rather like us, but deviate from us in quite specific and memorable ways. Thus we deviate from our folk physics when we tell stories about agents which can walk through walls or walk on water. We deviate from our folk physiology when we tell stories of agents which do not suffer pain or which do not die. We also deviate from our folk psychology when we tell stories about agents which know what we think and desire because they can see directly into our minds and souls. Finally we deviate from ourselves when we tell stories about agents which are all-powerful, all-knowing and all-benevolent (or totally malevolent as in the case of stories about devils). Stories about such MCIs are the stuff of our religions. The stories will differ from culture to culture but at the heart of each story will be the core belief in the existence of some kind of unobservable agent – a god.

The suggestions of the last paragraph now place HADD in a new context in which other cognitive mechanisms have evolved. Thus there are cognitive mechanisms for language L, mechanisms for theory of mind ToM, and mechanisms for the capacity to tell stories S which might be about MCIs, S(MCI). HADD continues it old work of agency postulation but in conjunction with {L, ToM, S, S(MCI)}. This can be represented as: HADD{L, ToM, S, S(MCI)}. Using the Humean Schema of the previous section concerning the cause of religious beliefs we can set out the following schematic casual hypothesis H which arises out of the project of evolutionary cognitive psychology (where 'D' stands for some divinity):

H: x's HADD{L, ToM, S, S(MCI)} causes x's belief that D exists.

Five points can be made about this schema. (i) When fleshed out more fully this hypothesis can be subject to test.[9] Since it is expressed in purely naturalistic terms it is open to scientific investigation. (ii) HADD has been co-opted to perform its old function of agency detection, but in the new context of a mind with evolved, additional cognitive capacities in which it causes beliefs about divinities. In the story told so far, HADD has not been directly selected by evolutionary processes to perform the function of divinity detection (or postulation). Rather it has been co-opted to do this; or as some might say its new function is not an evolutionary adaptation but is a by-product of adaptations that have come about through evolutionary processes (such as predator detection). (iii) On this account, the belief that D exists does not casually track any existent D. We can explain belief in the divine without positing the existence of the divine to do so. And this is so because the belief that D exists is the casual product of cognitive mechanisms which are purely naturalistic and have nothing to do with divinities. In the next section we will discuss whether or not there is good reason to think that such beliefs are reliable.

(iv) We can look at H from a Humean perspective and say that characteristically HADD "projects" agency onto the world. Sometimes HADD might be right and there is an agent that it has correctly postulated. But the postulation of unobservable divinities is not the result of HADD successfully tracking such an agent in its new context. Rather it is the operation of a combination of evolved mechanisms which causally produce a particular kind of belief, viz., that some D exists.

(v) Already we have considered six ways in which HADD can be unreliable (where the degree of reliability (on a scale of 0 to 1) can be understood as the frequency with which HADD produces true beliefs over the number of times HADD operates, or the propensity with which it delivers true beliefs). There is now a *seventh* way in which it is unreliable. HADD initially evolved in a context in which we can say that it was sufficiently reliable in that it hardly failed to detect predators when they were present (i.e., hardly any Type II errors or false negatives); that is, we did not die out but managed to survive to reproduce ourselves. But HADD is not so reliable because it is liable to signal 'predator present' when there are none (i.e., a not negligible number of Type I errors). Now in the new context of {L, ToM, S, S(MCI)} HADD operates outside the older context in which it initially evolved. And if it operates outside the environment in which evolution initially designed it to avoid Type I and II errors then we could expect that its reliability decreases in the new context for which it was not designed by evolution but merely co-opted to play its role in new context.

Given the seven ways in which HADD can be unreliable then it is apposite to speak of the illusion of agency that our cognitive mechanism can produce. We have a propensity to postulate agency where there is none. This also underlies our propensity to attribute teleology or purpose to much that happens in the world. This can, of course, be corrected when we learn more about the way the world works and drop the tendency to claim that all that happens in it is to be understood in terms of an anthropomorphic model in which activities of unseen agents with their beliefs and desires are the explainers. This is consonant with Kahneman's distinction between the fast and slow thinking of Systems 1 and 2 of our minds. HADD, in whatever context it is employed, is a typical of the "fast" thinking

provided by cognitive mechanisms of System 1. However it remains open to exploit the resources of System 2 which operates quite differently and can supplant System 1: 'System 2 has some ability to change the way System 1 works, by programming the normally automatic function of memory and attention' (Kahneman 2011, 23). And we can add to this list the "quick" operation of HADD as an agency detector. Its belief outputs are open to revision though some may be more resistant to revision than others.

This section began with Darwin's suggestion that we, and other animals, can postulate unobserved (and/or unobservable) agents which allegedly bring about observable states of affairs. Modern cognitive psychology develops this suggestion in one important way and shows how beliefs in such agents can easily pass into beliefs in the existence of gods or God through the activity of an evolved cognitive device such as HADD. And the stories we tell about such "minimally counterintuitive" agents (MCIs) go well beyond the minimal when we make them all-powerful, all-good and all-knowing. When we see many examples of design in the world we are prone to attribute it to the activities of an unseen agent (or a personal God) which, according to the theory of mind accompanying HADD, will have beliefs and intentions rather like us. Both the teleology we attribute to the world and a purposive God allegedly directing it are due to the operation of HADD. But HADD produces a considerable number of false positive indications of agency and of God; so it is not reliable in what it indicates as existing.

3. Grounds for Debunking Religious Belief

It is now possible to construct a "debunking argument" for belief in divinities.[10] This begins with the Humean "casual schema" in the case of HADD (see the end of section 2). This yields an initial explanatory hypothesis:

(1) Explanatory Hypothesis H about how the states of a belief-forming mechanism (BFM), viz., x's Φ-states, causally explain B, viz., x's belief that D exists,

The explanatory hypothesis is schematic and could be filled by a number of hypotheses from the scientific project of explaining religious belief. The particular hypothesis considered here comes from a specific project within cognitive psychology for explaining religious belief. It has the following form (where 'D' is a placeholder for some divinity or divinities):

H: x's HADD{L, ToM, S, S(MCI)} causes x's belief that D exists.

The belief-forming mechanism, BFM, in this case is HADD; but it operates within the context of other cognitive mechanisms of a highly evolved mental creature. Let us grant that we have a hypothesis such as H and regard it as one of Hempel's *potential* explanations (see the end of section 1). H does not, yet, provide an *actual* explanation as we need evidence for its truth, or evidence which provides a high degree of confirmation.[11] But H is part of the scientific project of discovering and testing theories about the causes of religious belief. By itself H is not enough to produce a debunking argument concerning religious belief; more needs to be added.

We will say that a BFM is reliable for the truth if and only if the ratio of true beliefs that the BFM produces over the number of times the BFM operates is 1, or close to 1. Alternatively we can view the BFM as a cognitive propensity and say that the propensity of the BFM to yield truths is 1 or close to 1. It is in this sense that a BFM is reliable (or as some would say, it "tracks the truth"); otherwise it is unreliable (or as some would say, it is "off-track").

HADD is one of many BFMs we humans possess. So what is its truth ratio? HADD (in the new context in which it operates in person x) is the cause of *x's belief that D exists*. But this provides no grounds for supposing that the propositional content, *that D exists*, is true. The previous section provided seven grounds for claiming that, even though HADD causally produces beliefs, it is not reliable for the truth of those beliefs. The main ground for this is that HADD is tuned by evolution to produce Type I errors (i.e., false positives) but much less probably produce more costly Type II errors (false negatives). Thus we may say:

(2) HADD is not reliable (it is prone to Type I errors, or false positives).

Given (1), person x's HADD does cause x's belief that D exists; but given (2) HADD is unreliable about the existence of D (dues to the proneness of false positives).

We can now offer grounds for a debunking claim:

(3) x's belief that D exists is debunked.

When important long-standing and revered beliefs about religion are shown to be caused in unusual ways (e.g., by a device like HADD) which are unreliable, then the beliefs become suspect. Such a casual story about the genealogy of beliefs in the divine unmasks the real source of the beliefs which has nothing to do with any divinity. The cause is not to be found in something divine but in something mundane and naturalistic. This in turn deflates their more commonly supposed portentous religious origin. When the

beliefs are explained in this alternative way they are demystified; their connections to alleged divinities are severed and this lays the grounds for their debunking.

Granted hypothesis H about the role of HADD, we will have a good *potential* explanation of why we have the beliefs in divinities that we do. Note that this does not show that H explains away these acts of believing. HADD is part of our cognitive equipment and it will keep on doing its job of generating beliefs about divinities as long as it is present. If hypothesis H of cognitive psychology is correct, HADD is part of what evolution has bequeathed to us and we will need to invoke HADD to explain what evolution produces, viz., core religious beliefs B. However even though HADD is part of the cognitive mechanisms which play a role in the "fast" thinking in Kahneman's System 1, its outputs can be overridden by other cognitive mechanisms which we have in System 2 and which have also evolved. So while there is a job of explaining why we have the religious beliefs we do, and H may well be a good theory to do this, there is nothing inevitable about us having the religious beliefs we do; they can be overridden by other cognitive mechanisms.

The considerations on behalf of debunking just given might only be provisional and not fully conclusive. It turns on the claim that core religious beliefs arise from a belief-forming mechanism HADD (working in its new context) which is often unreliable with respect to the beliefs it produces. Other belief-forming mechanisms can also give rise to core religious beliefs. However these might have the same defects as HADD if they are also unreliable. What the radical debunker of religious beliefs would have to do is show that these alternative belief-forming mechanisms are also unreliable. This is a much larger but not impossible task for a debunker.

What can be said of believers who get their beliefs in the way suggested by H but they are not aware of this? Given the above account of the workings of HADD and its unreliability, it is not required that any believer know, or be aware, that their core religious beliefs are so formed; it is enough that the beliefs actually be so formed, whether a believer knows this or not. Being reliable or unreliable is an objective matter concerning belief forming processes and not something of which a believer has to be aware.

But what if they are made aware of the unreliability of their belief-forming mechanism by showing them something like the above debunking considerations? That is, person X comes to learn, by studying some cognitive psychology, the following about their unreliable HADD:

(4) X comes to learn (2), viz., that their HADD is not reliable

It would be open to a believer to dogmatically reject what they have learned and stay with their beliefs. In contrast many, if they learn that their act of belief is unreliably formed by a mechanism such as HADD, could agree that there is a defeater of their act of belief. Accepting the second order belief that HADD is unreliable would come to be a defeater of their first order belief that D exists. They might then become sceptics about their belief that D exists. Whether they have recourse to other grounds on which to base their religious belief remains an open question. They might appeal to Hume's other project of founding religious belief not in human nature but in reason. However encountering Hume's arguments (or an updated version of them) against founding religious beliefs in reason might well leave them bereft of any other further grounds for their religious beliefs, and so they become at least sceptics.

4 **Which is the Better Explainer - HADD or Divinities?**

We began by supposing that our various acts of believing that some D exists (B) stand in need of explanation. The "folk" religious theory F purports to explain B by invoking the very entities believed in, viz., the various divinities D, as a central part of any explanatory hypothesis (however that explanation is to be spelled out). In contrast naturalistic hypotheses rule out any appeal to divinities in their explanations of B and instead admit only naturalistic items of which HADD is one example. Let us use 'E_H' to designate a broadly understood naturalistic theory of evolution E applied to biology and psychology, which contains hypothesis H about HADD. We can now ask: which of F or E_H explain B better?[12] This leads to the following schematic versions of Inference to the Best Explanation (IBE) (sometimes also called 'abduction'):

(1) To be explained: B.
(2) Proposed Explainers: F and E_H.
(3) Supposition: E_H is a better explainer of B than F.
(4) Conclusion (a): it is true that E_H.
Conclusion (b): it is reasonable to accept that E_H.
Conclusion (c): Given only the pair-wise comparison of (3), E_H is to be accepted (preferred or favoured) over F.

Premise (1) is a given. But (2) does not propose an exhaustive and exclusive set of rival hypotheses. F and a broad naturalism N are exclusive of one another; but N can contain rival hypotheses pertaining to

evolution other than E_H (these rivals are not considered further here). And perhaps there are further hypotheses not yet envisaged which might do better than either of the suggestions in (2). Shortly we will consider a number of grounds on the basis of which (3) can be accepted. The claim that one hypothesis is a better explainer than another of some given data is commonly expressed as a comparative likelihood, viz., $p(B, E_H) > p(B, F)$.

Granted these three premises what conclusion should we draw? (4a) seems too strong in claiming that we can arrive at the true hypothesis. For that reason some prefer the weaker conclusion of (4b) such as that suggested by Pierce (1955, 151) when he proposed his schema for abduction by placing an epistemic operator in front of the conclusion such as 'there is reason to suspect that', and the like. In order not to be detained by these problems let us adopt conclusion (4c) in which we try to judge, of a pair of rival hypotheses, which is the *better* or the more *favoured*. In fact we are led directly to this by a Law of Likelihood which says: data D favours hypothesis H_1 over H_2 if and only if $p(D, H_1) > p(D, H_2)$.[13]

Consider now the following five criteria (at least) on which (3) is shown to be a better explanation.

(1) *Ontological Parsimony*: other things being equal, prefer that theory which postulates the existence of fewer things or kinds of thing.

On this criterion E_H wins over F. If we suppose that F will include all the naturalistic items that E_H includes but supposes non-naturalistic items that E_H does not, then E_H is to be preferred as the more ontologically parsimonious explainer. Whether or not F actually explains anything in this context is a contested matter (see (2) and (5) below).

(2) *Explanatory Breadth*: other things being equal, prefer the theory which explains a wider variety of facts over a theory which explains fewer.

Consider the domain of facts in biology. Here we need to compare F with not just H but with E_H, the wider theory of evolution which contains H, but a lot more as well. Despite the controversy that surrounds it, the theory of evolution explains more about the facts of biology than F does. Doctrines of "creationism" or "intelligent design" are alternative ways of attempting to make F relevant to the facts of biology; but they make no headway against the theory of evolution when it comes to explaining biological facts. This is a point recognised by Darwin who showed several times over that his own natural selection hypothesis was a better explainer than the creationist hypothesis. In fact on some occasions he argued more strongly that the creationist hypothesis could offer no explanations at all. (Nola, 2013a). So on the score of explanatory breadth the theory of evolution E_H, which contains H, is wider than F; and it is more successful than F which lags badly behind on this score.

(3) *Explanatory and Predictive Novelty*: other things being equal, prefer the theory which leads to new unknown facts or offers an explanation for the first time of old well-known facts, over a theory which does neither.

The God hypothesis of F is hardly equipped to yield such predictive novelty since it is hard to see how such novelty can be obtained out of the God hypothesis alone. If it is to do this it needs much supplementation from our current science. However E, of which H is part, does yield predictive novelty in the sphere of biology (see Coyne 2009, 17-18 for a list of six successful predictions, as well as retrodictions, from the theory of evolution). So on the criterion of making novel predictions E_H does better than F.

(4) *Progressive Explanatoriness*: other things being equal, prefer the theory which is progressive in that it uncovers new facts while a rival theory merely accommodates itself to such new facts which are relevant to it but in which it plays no role in uncovering these new facts.

Which theory, F or E_H, was employed in the discovery that the human mind contains a cognitive device, HADD, for the detection of agency? It was not F but E_H. F makes no such discovery about the human mind but rather piggy-backs on discoveries made in the rival science E_H; and then it incorporates these discoveries into a seemingly new explanation of what "folk" religious belief can do. This is what Lakatos would call a "degenerating problemshift" (Lakatos 1978, 112). Discoveries about how a supposed God makes himself apparent to us humans, which would be relevant to F, are in fact made in a rival progressive programme, E_H, which does not need the divine postulates of F at all. However F takes on board the fact of (the newly discovered) HADD and the beliefs it causes and attempts to accommodate these into its framework. But, if it were a more successful research programme than its rival E_H, it ought to have anticipated these new facts. So F fails criterion (4) while E_H passes it.Consider other ways in which explanations crafted within F can illegitimately piggy-back on discoveries in cognitive psychology, such as the postulation of HADD. Religious believers regard cognitive psychology as a threat to a belief about the existence of divinities. Cognitive psychology postulates HADD which causes acts of believing in divinities without the truth of the content believed (viz., that there are divinities); F makes no such postulations. So once HADD emerges within cognitive psychology as a proposed explanatory mechanism, attempts might be made to incorporate it into F thereby showing that evolutionary processes leading to

HADD provide one way, admittedly quite circuitous, in which a divinity can make itself evident to humans in the world.

One kind of incorporation is suggested by van Inwagen (2009 pp. 134-8 and footnotes 10 and 11). It is supposed that there is a creator who considers the full range of possible worlds one of which he makes actual: 'God chooses that one and says, 'Let it be' (*ibid*., footnote 10). Moreover the kind of naturalistic account of the world that includes item like HADD is said to be such that: 'Any naturalistic explanation of any phenomenon can be incorporated without logical contradiction into a "larger" more comprehensive supernaturalistic explanation of that phenomenon' (*ibid.*, 134). But what is the kind of incorporation here? It is claimed that a naturalistic cognitive psychology E_H which does its work in postulating HADD, can get incorporated into a supernatural theory S without contradiction. But this can be done too easily. Just form the conjunction (S& E_H), on the supposition that this entails no contradiction. Then the conjunction (formed by the irrelevant conjunction of S to E_H) will explain, in the Hempelian sense, what E_H explains by itself. Moreover the incorporation can occur with a false, or a true, S (with the proviso that the conjunction is consistent). This conception of "incorporation" by mere conjunction is not something that any naturalist should fear.

Justin Barrett, like van Inwagen, claims that the divinity can contemplate all possible worlds and then make one of them actual, viz., the one in which evolution occurs and which culminates in us humans with HADD-like capacity to believe in, and love, God: 'God could have instantiated this world out of all the possible worlds because in this world natural selection brought about the kind of creatures capable of a loving relationship with Him' (Barrett 2009, 97). But a HADD-like capacity to detect God does not necessarily imply that what is detected is also loved. Too much might be expected from a supposed evolved capacity of humans to be in a loving relation with God; there would be little ground to suppose that evolution bequeathed such a capacity to all humans (in much the same way as it has bequeathed to us a HADD which merely has the capacity to detect agency which it might fear rather than love). There are non-believers and anti-theists who have, or want, no such loving relation (either because they simply do not, or because they lack the capacity). So it would be, to some extent, a misjudgement on the part of the deity to make a world that has non-loving non-believers. Moreover, as Hume suggests, it turns on the belief 'that the Deity has human passions, and one of the lowest human passions, the restless appetite for applause [love]' (Hume 1993, 128).

(5) *Absence of Untestable Assumptions*: other things being equal, prefer that theory which has no untestable assumptions.

F fails in comparison with E_H in another way; it turns on an argument from design which fairs unfavourably with E_H in that it invokes untestable auxiliary hypotheses about the intentions of a divinity. There are two parts to the argument.

(i) In line with the above claims about God, suppose that God is all-powerful in that he can contemplate all possible worlds and he can make one of them the actual world. He chooses that world in which laws of evolution operate, and over time it comes to contain humans who have HADD devices which cause belief in divinities. To this we need to add something about what God desires or intends.

(ii) God seems to desire two things: (a) a world with creatures in it like us who can at least have a belief in God and not lack any such belief; (b) the creatures who have such beliefs form a loving relationship with God (and do not remain indifferent to him or hate him). Such abilities and desires then fit a belief/desire model for explaining why God with beliefs and intentions made this world with us in it with our HADD mechanisms and did not make some other world, or did make this world but without evolutionary processes which would guarantee the emergence of a HADD device for "detecting" him.

Let us focus on (ii) (a) as a premise about God's intentions to make a world in which creatures like us evolve and have HADD mechanisms for believing in God. What evidence do we have for God having such intentions? It is not enough to merely cite the fact that there is such a world with creatures in it like us with HADD like devices in our mind (brains). Assuming this in part begs the question about a creator in the dispute between naturalists and creationists. The fact that the world is the way described is not good grounds for assuming that a creator had the intention to make such a world. This becomes an untestable assumption.[14]

The above turns very much on an "intelligent design" argument. One problem is that HADD seems not to be the best way in which an all-powerful divinity could get humans to believe in him. HADD is unreliable and a more reliable detection device might have been in order if we are supposed to detect him. Couldn't God have made a better device than HADD to detect more reliably his presence? And as suggested, there are problems with (ii); there is no independent evidence for what God desires and wants. Sure, this world does exist and the laws of evolution seem to operate in it. And if E_H is right, then we do have HADD which works away causing acts of believing in divinities. But the fact that such a world exists with an evolved HADD is no evidence for the claim that God desired to have such a world or

that he intended to make it. So in the absence of any evidence for God's desires and intentions to make this world, *independent* of the fact that this world does exist and HADD operates in it, is to make an independent assumption about God's intentions and desires which is not open to any possible scientific test.

(6) Fewer *ad hoc* assumptions: other things being equal, of two theories prefer that which has no *ad hoc* assumptions or which makes fewer *ad hoc* assumptions.

It is not always clear what *ad hocness* means. Making untestable auxiliary assumptions as discussed in (5) above is one form of *ad hocness*. Another proposed by Popper (1959 section 20) concerns the decrease of the degree of testability of a given hypotheses using further saving auxiliary hypotheses. However it is not clear yet than in the debate concerning the degree of explanability of F and E_H that this kind of *ad hocness* arises.

This completes the discussion of which of HADD or divinities best explains our religions beliefs (of the sort specified in B). An argument of the form of IBE was proposed and then at least five different criteria were considered to determine which provides the better explanation (but more criteria than this could be considered). The outcome is that E_H wins over F. The argument is general enough to enable one to consider which of a naturalist or supernaturalist framework best explains B, our set of believings in divinities; it need not be confined to the existence and workings of HADD alone. The verdict is that naturalism provides the better explanation.[15]

References

Barrett, H. Clark (2005) ‘Adaptations to Predator and Prey’, David M. Buss (ed.) *The Handbook of Evolutionary Psychology*, Hoboken NJ, John Wiley & Sons: 200-223.

Barrett, Justin 2000, ‘Exploring the Natural Foundations of Religion’, *Trends in Cognitive Science*, 4(1); 29-34.

Barrett, Justin 2004. *Why Would Anyone Believe in God?* Walnut Creek CA, Altamira Press.

Barrett, Justin 2009, ‘Cognitive Science, Religion and Theology’, in Schloss and Murray (eds.) 2009, pp 76-99.

Barrett, Justin 2012. *Born Believers: The Science of Children’s Religious Belief*, New York: Free Press.

Barrett, Justin and Lanman, J. A. (2008). The science of religious beliefs. *Religion*, 38(2), 109–124.

Coyne, Jerry 2009. *Why Evolution is True*, New York: Viking.

Darwin, C. 1882, *The Descent of Man*, London: John Murray. 2d edition; http://darwin-online.org.uk/contents.html#descent

Dennett, Daniel. 2006. *Breaking the Spell: Religion as a Natural Phenomenon*, New York: Viking.

Haselton, M. and Buss, D. 2000, ‘Error Management Theory: A New Perspective on Biases in Cross-Sex Mind Reading’, *Journal of Personality and Social Psychology* 88 (1), 81-91.

Haselton, M. and Nettle, D. 2006, ‘The Paranoid Optimist: An Integrative Model of Cognitive Biases’, *Personality and Social Psychology Review* 10 (1) 47-66.

Hempel, Carl G. 1965. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, New York: The Free Press.

Hume, David 1993. *Dialogues and Natural History of Religion*, Oxford: Oxford University Press.

Isbell, Lynne A. (2006). Snakes as agents of evolutionary change in primate brains. *Journal of Human Evolution*, 51: 1–35. 10.1016/j.jhevol.2005.12.012.

Isbell, Lynne A. (2009). *The Fruit, the Tree, and the Serpent: Why We See So Well*. New York, Harvard University Press.

Jong, Jonathan and Halberstadt, Jamin (2016) *Death, Anxiety and Religious Belief: An Existential Psychology of Religion*, London, Bloomsbury. Kahane, Guy 2011. ‘Evolutionary Debunking Arguments’, *Nous* 45(1): 103-25.

Kahneman, Daniel 2011. *Thinking Fast and Slow*, New York: Farrar, Straus and Giroux.

Lakatos, Imre 1978. *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1*, Cambridge: Cambridge University Press.

Law, Stephen (2016) ‘The X-claim argument against religious belief’, *Religious Studies*: first published online: 21 December 2016: 1-21.

Leech, David and Visala, Aku (2011) ‘The Cognitive Science of Religion: Implications of Theism?’ *Zygon*, 46 (1): 47-64.

Nola, R. (2013a) ‘Darwin’s Argument in Favour of Natural Selection and Against Special Creationism’, *Science & Education*, 22 (2): 149-171. doi:10.1007/s11191-010-9314-3

Nola, R. (2013b), ‘Do Naturalistic Explanations of Religious belief Debunk Religion?’ In Gregory Dawes and James Mclaurin (eds.) *A New Science of Religion*, London, Routledge: 162-188.

Peirce, Charles, 1955. *Philosophical Writings of Peirce*, Justus Buchler (ed.), New York: Dover.

Popper, Karl (1959) *The Logic of Scientific Discovery*, London, Hutchinson.

Schloss, Jeffrey and Murray, Michael (eds.) 2009. *The Believing Primate*, Oxford: Oxford University Press.

Sober, Elliott (2008) *Evidence and Evolution: The Logic behind the Science*, Cambridge, Cambridge University Press.

Trigg, Roger and Barrett, Justin (eds.) (2014) *The Roots of Religion: Exploring the Cognitive Science of Religion*, Farnham, Ashgate.

Van Inwagen, Peter 2009. ‘Explaining Belief in the Supernatural’, in Schloss, Jeffrey and Murray, Michael (eds.) 2009, pp 128-8.

White, Claire, (2017). What the Cognitive Science of Religion is and is Not. In Hughes, Aaron W. (ed.) *Theory in a Time of Excess: Beyond Reflection and Explanation in Religious Studies Scholarship*, 95-138. Sheffield: Equinox.

[1] An exception here might be Buddhism versions of which are not monotheistic and are even atheistic. However many versions of Buddhism are polytheistic and do postulate divinities of some sort.

[2] Here I do not attempt to investigate the claims the various contributors to this collection make concerning the challenge CSR makes to religious belief. Similarly for a host of other papers such as that of Leech and Visala (2011) who argue that CSR and theism are not incompatible and that CSR is world-view neutral (pp. 60-61). Here the view is that CSR is not world-view neutral and provides grounds for naturalism.

[3] One writer who agrees with the conclusion that CSR and religion are not consistent is Dennett, 2006, particularly Part II.

[4] For an excellent discussion and critique of the Humean claim that our fears of death cause beliefs in god or divinities see Jong and Halberstadt (2016).

[5] For one of many suggestions along these lines, see Barrett and Lanman (2008)

[6] See Haselton and Buss 2000; Haselton and Nettle 2006.

[7] Snakes are mentioned because there is good evidence that the presence of predatory snakes played a big role in primate evolution, from sight and colour discrimination to the development of HADDs. For example see Isbell 2006 and 2009.

[8] The normally invisible Abrahamic God is said to put in an appearance when he displayed his "back parts" but not his face to Moses; see Exodus 33:23 (King James translation, though other translations are less specific about what was displayed).

[9] Much empirical evidence is being found for the activities of cognitive devices such as HADD. For a sampler see Dennett 2006, chapter 4; Barrett 2012.

[10] Debunking arguments are suggested in Kahane (2011, 106). They are set within a reliabilist conception of knowledge. This form of debunking argument was followed in Nola (2013b). However the main point being made against religious belief can get caught up in problems that arise for reliabilist epistemology itself. These can be separated as is done in the presentation of a related "debunking" argument in Law (2016). Here I follow Law's suggestion for presenting a slightly different set of considerations on behalf of debunking.

[11] A number of evidential considerations on behalf of HADD are provided by Justin Barrett, the inventor of the acronym HADD; see Barrett, 2012. So the HADD hypothesis is not without some evidential support. This is a matter which will not be evaluated here.

[12] The contrast between F and E_H is closely related to the contrast between the supernaturalist and naturalist frameworks. Here an attempt is made to adjudicate between the frameworks using the same criteria as one would use in assessing the epistemic worth of rival scientific theories.

[13] This Law is given in Sober 2008, p. 32; it is expressed here using different symbols.

[14] For a good account of what goes wrong here, and how untestable assumptions arise, see Sober 2008, section 2.12, pp. 141-7.

[15] I would like to thank Hans Van Eyghen and an anonymous reviewer for comments on an earlier draft which helped improve this revised version.