Evolutionary approaches to epistemic justification

Helen De Cruz
University affiliation: Centre for Logical and Analytic Philosophy, Katholieke Universiteit Leuven, K. Mercierplein 2, 3000 Leuven, Belgium and Somerville College, University of Oxford, Woodstock Road, OX2 6HD Oxford, United Kingdom
e-mail: helen.decruz@hiw.kuleuven .be

Maarten Boudry
University affiliation: Department of Philosophy and Ethics, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium
e-mail: maarten.boudry@ugent.be

Johan De Smedt
University affiliation: Department of Philosophy and Ethics, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium and Uehiro Centre for Practical Ethics, University of Oxford, Littlegate House, 16-17 St Ebbes Str, Oxford, United Kingdom
e-mail: johan.desmedt@ugent.be

Stefaan Blancke
University affiliation: Department of Philosophy and Ethics, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium
e-mail: stefaan.blancke@ugent.be

This is the final draft of a paper that has appeared in *dialectica*, 65(4), 517–535

Abstract
What are the consequences of evolutionary theory for the epistemic standing of our beliefs? Evolutionary considerations can be used to either justify or debunk a variety of beliefs. This paper argues that evolutionary approaches to human cognition must at least allow for approximately reliable cognitive capacities. Approaches that portray human cognition as so deeply biased and deficient that no knowledge is possible are internally incoherent and self-defeating. As evolutionary theory offers the current best hope for a naturalistic epistemology, evolutionary approaches to epistemic justification seem to be committed to the view that our sensory systems and belief-formation processes are at least approximately accurate. However, for that reason they are vulnerable to the charge of circularity, and their success seems to be limited to commonsense beliefs. This paper offers an extension of evolutionary arguments by considering the use of external media in human cognitive processes: we suggest that the way humans supplement their evolved cognitive capacities with external tools may provide an effective way to increase the reliability of their beliefs and to counter evolved cognitive biases.

1. The evolved mind and epistemic justification

Daniel Dennett (1995) has famously compared evolutionary theory to a universal acid—a corrosive substance that eats its way through anything it touches, transforming every field it is applied to. Darwin's idea of natural selection has the power to affect ideas far outside its original domain, including economics, culture, language and epistemology. Since evolutionary theory presents our current best hope to explain design and adaptation from a naturalistic point of view, it is perhaps not surprising that a growing number of philosophers

(e.g., Boulter, 2007; Fales, 1996; Quine, 1969; Stewart-Williams, 2005) incorporate evolutionary arguments in their naturalistic theories of mental content.

The aim of this paper is to examine the implications of evolutionary epistemology for the epistemic justification of beliefs. The term 'evolutionary epistemology' will be used in a broad sense, namely to denote the position that biological evolutionary forces, in particular natural selection, are important in shaping cognition. As we shall see in section 2, evolutionary considerations are being used to either justify or debunk a wide variety of beliefs, including commonsense beliefs, religious ideas, moral judgments and scientific hypotheses. However, as general strategies both justification and debunking are problematic. The former might be subject to circular reasoning, e.g., using induction to justify induction. The latter is potentially self-undermining—if our cognitive faculties are deeply unreliable, why should we buy into evolutionary theory, which is after all a product of those same cognitive faculties? Section 3 reviews and extends strategies to counter the circularity charge leveled against evolutionary arguments. In section 4, we propose that incorporating the extended mind thesis in evolutionary arguments can provide a means to justify beliefs, especially those outside the scope of common sense. The example of temperature will illustrate how using external media allows humans to reach beyond their evolved cognitive biases.

2. Cartesian God or Cartesian demon: the double-edged sword of evolution

Does the fact that the human brain is a product of organic evolution give us a reason to believe that mental states accurately reflect the states of the world, or does it lead to a far-reaching form of skepticism? Evolutionary approaches to the mind have given rise to two mutually incompatible positions. The first position, supported by *evolutionary arguments* (EAs), contends that natural selection will tend to pick out and propagate those types of beliefs and judgments that correspond with the state of the world. The second stance relies on *evolutionary debunking arguments* (EDAs). An EDA is constructed by negating at least one of the crucial EA premises, in particular about the relative importance of natural selection, and about its truth-tracking ability. Although EAs and EDAs share several premises, they reach contradictory conclusions. To see how they differ, let us examine the general structure of both arguments[1].

[1] This formulation of EAs and EDAs is general, in order to capture the similarities and differences in the premises of both arguments. For an alternative formulation of EA, see Boulter (2007); for another rendition of EDA, see Kahane (2011).

EA	EDA
1. In order to survive and reproduce, animals need to interact with their environment in a successful way (e.g., find food and shelter, avoid predators, find mates).	
2. An animal's beliefs guide its interactions with the environment.	
3. [from 1 & 2] There is a significant correlation between an animal's beliefs and its fitness.	
4. True beliefs are better guides to action than false beliefs.	*4*. There is no intrinsic relationship between the truth of a belief and its ability to guide an animal's actions successfully.*
5. [from 3 & 4] Animals that have true beliefs, in particular those concerning ecologically relevant circumstances, will have a higher fitness than those that entertain false beliefs.	5*. [from 3 and 4*] Animals that have beliefs that enhance their successful interactions with the world, regardless of their truth-value, will have a higher fitness than those that do not have such beliefs.
6. Natural selection is the primary drive of evolution of cognitive systems.	*6*. Natural selection is of limited importance in shaping cognitive capacities.*
7. [from 5 and 6] Natural selection will select for sensory systems and conceptual cognitive capacities that generate true beliefs.	7*. [from 6*] Evolutionary mechanisms different from natural selection are certainly not truth-tracking.
8. [from 7] Our beliefs are justified, in particularly in domains that are ecologically relevant.	8*. [from 5* or 7* or both] Our beliefs are not justified.

Table 1: Generalized forms of EA and EDA.

Note that 4* and 6* are logically independent defeaters of EA, and as we shall see, several EDAs only make use of one of these premises. An EDA on the basis of 4* emphasizes that natural selection is primarily concerned with fitness[2], not with tracking truths. According to EAs, true beliefs are better guides to action than false beliefs, whereas EDAs that make use of premise 4* (e.g., Plantinga, 1993) see no reason why true beliefs would be privileged. By

[2] Several concepts of fitness are used in biology and in philosophy of biology (see e.g., Ramsey, 2006). *Realized fitness* tracks the actual reproductive success (number of offspring) of individuals within a population. *Propensity fitness* conceptualizes fitness as the propensity of an individual organism to produce a number of offspring. There can be a discrepancy between an organism's propensity fitness and its realized fitness. For example, an animal can have beliefs that increase its propensity fitness (e.g., be properly cautious of predators) and yet fail to realize this high propensity fitness because it is struck by lightning before reaching maturity. For the purposes of this paper, we will use 'fitness' to mean propensity fitness.

contrast, an EDA based on 6* argues that natural selection is just one among many evolutionary forces, and there are lots of factors that may interfere with it, giving rise to non-adaptive traits (see subsection 2.2). EDAs that make use of premise 6* (e.g., Stich, 1990) typically grant that natural selection can sometimes be truth-tracking, but are skeptical of its prominence in shaping cognition. They argue that alternative evolutionary explanations, if anything, are more detrimental to truth-value than explanations that invoke natural selection. Thus, an EDA is an argument where one or both of the crucial premises of EAs are negated. In order to assess the plausibility of these positions, we will now look in more detail at the case for EAs and EDAs, examining their scope and potential problems.

2.1 Evolutionary arguments

The wide applicability of EAs is potentially problematic, as it can lead one to replace the Cartesian God with natural selection. Even if we assume that there is some link between the truth of a belief and its adaptive value (premise 4), that beliefs are partly though indirectly under genetic control, and that natural selection is the primary drive of evolution (premise 6), the scope of EAs should be fairly restricted. EAs have been put forward to justify a wide range of beliefs and belief-producing mechanisms, such as the ability to make inductions (e.g., Quine, 1969, 126), and to draw inferences to the best explanation (Goldman, 1990). The most successful EAs offered to date are in the domain of commonsense beliefs. 'Common sense' refers to those types of beliefs that are obvious and self-evident to the subject, such as the existence of other minds, the occurrence of past events, and the reliability of perception. Stephen Boulter (2007) argues that these are the kinds of beliefs that are most likely to have a bearing on an individual's fitness. Echoing Thomas Reid (1764), he insists that commonsense beliefs are crucial for our ability to act adaptively in the world. Their adaptive value provides evidence for their validity. Steve Stewart-Williams (2005) focuses on one particular commonsense belief, our robust metaphysical belief in an observer-independent world. People are typically not swayed by skeptical arguments that cast doubt on this belief, and even if they proclaim to be skeptical, in practice they still intuitively take the existence of the world to be self-evident. He argues that this deep metaphysical conviction has an innate basis, and that the best explanation for this is that the observer-independent world actually exists. In this way, EAs provide an answer to the radical skeptic or solipsist, who denies the existence of the external world:

> The fact that any normal mind automatically assumes an objective and mind-independent external world may count as proof that such a world does exist. We evolved a mind/brain that creates a sense of an objective, mind-independent external world because this tendency generally contributed to the persistence of the genetic material that gave rise to the tendency. In what kind of world would this tendency be biologically advantageous? It would be advantageous in a world that genuinely exists beyond our fleeting sensory impressions. The fact that this tendency evolved indicates that it was useful, and the simplest explanation for its usefulness is that it is accurate (Stewart-Williams, 2005, 794).

Critics have often complained that EAs are inherently circular, because such arguments rely on a theory that is the product of human rationality to justify the rationality of our beliefs. Thus, a naturalized epistemology based on evolutionary theory is an inherently unstable position (Plantinga, 1993). We will examine two possible responses against this criticism in section 3. A second potential problem for EAs is that they posit that successful (fitness-enhancing) action requires accurate beliefs. As we shall see in the next subsection, EDAs can cast doubt on this premise.

2.2 Evolutionary debunking arguments

As argued above, EDAs attempt to undermine EAs by relying on one of two possible strategies (or both): they attempt to disconnect the link between the truth of a belief and its efficacy (4*), or they cast doubt on the shaping role of natural selection in belief formation (6*). We will examine both strategies in turn. Evolutionary theories of mental content propose that the 'proper' function of our cognitive processes is to promote survival and reproduction (e.g., Millikan, 1984). In analogy to the proper function of organs, such as the heart's function of pumping blood, our cognitive apparatus guides us to perform those kinds of actions that enhance fitness. According to EDAs based on 4*, this should lead us to doubt the accuracy of our beliefs. Natural selection may, for instance, produce risk-aversive and thus error-prone cognitive mechanisms, as in the detection of predators (see below). Natural selection is concerned with *fitness*, i.e., a propensity to produce a greater number of surviving offspring, not with *truth* (premise 4*). Alvin Plantinga, for instance, writes:

> If our cognitive faculties have originated as Dawkins thinks [i.e., through natural selection], then their ultimate purpose or function […] will be something like *survival* (of individual, species, gene or genotype); but then it seems initially doubtful that among their functions—ultimate, proximate or otherwise—would be the production of true beliefs (Plantinga, 1993, 218).

Interestingly, Charles Darwin[3], in a letter to William Graham, also expressed doubts about the justification of our beliefs based on premise 4*:

> With me the horrid doubt arises whether the convictions of man's mind, which have been developed from the mind of the lower animals, are of any value or at all trustworthy. Would anyone trust in the convictions of a monkey's mind, if indeed there are any convictions in such a mind? (Darwin 1881).

But even if premise 4 of EA is secured, establishing a connection between fitness and truth-value, EDA proponents may try to undermine premise 6. Stephen Stich (1990, 56) points out several reasons for why the powers of natural selection are limited (premise 6*), so that evolution does not always produce "good approximations to optimally well-designed systems": the fitting mutations may fail to arise at the right time, or pleiotropic effects[4] and drift[5] may result in cognitive mechanisms that are not fitness-enhancing. Even if the right mutations do occur, it is possible that natural selection gets stuck on a local peak in a fitness

---

[3] In his *Descent of man, and selection in relation to sex* (1871), Darwin argued that human mental faculties are a product of natural selection, so that his doubt must have arisen from premise 4* rather than 6*.

[4] Pleiotropy occurs when one gene influences more than one trait. This may be problematic for selection, because the selection for one trait might favor one specific version of the gene, whereas selection for another trait might favor another version of the gene. As a result, neither trait will be optimal.

[5] As in the case of fitness, the concept of drift has many connotations. It includes indiscriminate sampling, e.g., large-scale events like floods that do not discriminate between fitter and less fit individuals, as both types are equally likely to drown. In this example, fitness differences cannot explain why a particular part of the population dies in an inundation, whereas the other survives; it is a matter of being at the wrong spot at the wrong time. Another case of drift is the founder effect, where part of a population gets isolated, leading to the spread of an originally rare and non-fitness related trait in that subpopulation (see Walsh et al., 2002, for an overview).

landscape[6] without being able to reach a better optimum. In addition, trade-offs may occur between utility on the one hand and expenses in time and resources on the other, resulting in cognitive systems that are less than reliable. Additionally, cognitive mechanisms may also have come about through sexual selection, i.e., through the interaction of phenotypes not with the environment, but with the fancy of the opposite sex (Miller, 2000).

Taking into account all these scenarios in which either premise 4 or premise 6 (or both) of EA fails to obtain, it seems that evolution has become more of a latter day Cartesian demon. Plantinga (1993) is perhaps best known for his radical claim that a naturalistic point of view is self-defeating, since evolution through natural selection is not concerned with tracking truths, but with increasing fitness (4*). He argues that only a supernaturalist ontology—where beliefs derive their warrant from the fact that cognitive capacities have been designed by God in such a way that they successfully aim at the truth—can provide an externalist justification for our beliefs. However, most EDAs are less wide in scope than Plantinga's.

Take EDAs against religious beliefs. Empirical evidence from developmental psychology and cognitive science indicates that many elements of religious beliefs arise early in development, and are stably present across cultures, such as an overattribution of agency (we all see faces in the clouds, Guthrie, 1993), the belief that minds persist after death (Bering, 2006), and the intuition that natural objects are designed for particular purposes (Kelemen, 2004). What does this mean for the epistemic justification of religious beliefs? Researchers in the cognitive science of religion have argued that religious beliefs are byproducts of cognitive functions, such as agency detection and theory of mind (the ability to infer mental states). In his attack on religion, Richard Dawkins uses the cognitive science of religion literature to explain away religion as an accidental byproduct of our cognition:

> Religion can be seen as a by-product of the misfiring of several of these modules, for example the modules for forming theories of other minds, for forming coalitions, and for discriminating in favour of in-group members and against strangers (Dawkins, 2006, 179).

But does being a byproduct by itself undermine religious beliefs? Dawkins does not tell, but if that were so, one could argue that science, which is also a byproduct of the evolved structure of human cognition, is likewise undermined. In both cases, the trait in question would be an accidental byproduct of selection, not a direct product of it (6*). One can formulate a byproduct account of religion in a more refined way so that religion, but not other byproducts, is debunked. Paul Bloom (2009), for example, acknowledges that the evolutionary origins of religious beliefs do not provide a straightforward refutation of such beliefs. After all, cognitive scientists also explore why people believe that 5 + 5 = 10, and none of them come to doubt the validity of this outcome as a result. Nevertheless, he thinks that the cognitive science of religion can challenge the rationality of holding on to religious beliefs. Given that religion is an "evolutionary accident", and that a plethora of mutually incompatible religious concepts exists across the world, believers may not be justified in holding the beliefs they do. Thus, it is not the evolutionary origin of religion itself, but the fact that evolved cognitive

[6] Fitness landscapes provide a visualization of the relationship between specific genotypes and reproductive success; peaks represent states where genotypes achieve a high realized fitness. Rugged fitness landscapes have multiple peaks separated by valleys. In that situation, it is difficult for a particular genotype to move away from one peak to reach another, higher peak, because this move (crossing a valley) leads to temporary fitness costs. Organisms with a particular cognitive architecture may be stuck on a local, lower peak, as the costs involved in neural reorganization may prevent them from evolving a more efficient brain.

biases give rise to many incompatible beliefs, that provides a debunking argument against religious beliefs. Similarly, Jesse Bering (2011) argues that, while evolutionary theory does not disprove the existence of God, it nevertheless makes it improbable. In principle, a theist could argue that God instilled religious belief in humans indirectly, through natural selection, but "if scientific parsimony prevails […] such philosophical positioning [i.e., theistic evolution] becomes embarrassingly like grasping at straws" (Bering, 2011, 196). It is not the evolutionary origin of religious belief in itself, but its conjunction with the principle of parsimony that is used in a debunking argument. The case of religion indicates that evolutionary origins of a belief do not always constitute sufficient grounds to dispel it.

Nonetheless, under specific circumstances, the evolutionary origins of a claim may negatively affect its epistemic status. Thus, EDAs might work if they provide fine-grained reasons for why the evolutionary origins of a particular belief might affect its epistemic justification. We here provide four examples of specific EDAs.

i) Cognitive processes may sometimes *err on the side of caution*. If the costs or payoffs of false positives (detecting a signal in the environment where there is none) and false negatives (failing to detect a signal that is present in the environment) are asymmetric, natural selection will tend to promote beliefs that yield the highest payoffs or incur the least costs (Stephens, 2001). An example of this is agency detection: humans and other animals are prone to detect agency in the environment where none is present (e.g., mistaking wind rustling in the foliage for an approaching animal). This cognitive capacity generates an excess of false positives. The evolutionary rationale for this is that a false positive is less costly than a false negative, as the latter can result in a failure to detect a dangerous predator, a prey, or a potential mate, and the former only results in a small waste of time and energy (Guthrie, 1993).
ii) Animals are bounded in time and space which leads to *trade-offs between accuracy and efficiency*. There is little point in carefully and elaborately choosing the best escape route when faced with a hungry predator.
iii) Some *cognitive illusions may be adaptive* (McKay & Dennett, 2009). The belief of devoted parents that their own children are more beautiful, smarter and kinder than average (the Lake Wobegon effect) is clearly unjustified as not every child can be above average, but it may contribute to the time and energy parents invest in their children, thereby enhancing their inclusive fitness. Wenger and Fowers (2008) found that the majority of a sample of randomly selected biological parents of young children holds unrealistically positive views about their children. The parents in whom this illusion was most pronounced reported the highest degrees of parental satisfaction.
iv) *Intuitive beliefs that have no bearing whatsoever on fitness* are invisible to natural selection, and are thus unreliable. After all, natural selection is the only candidate for a truth-tracking evolutionary mechanism, but it is of limited importance in shaping cognitive capacities. Next to this, we can expect that evolutionary mechanisms such as drift (in the meaning of stochastic processes) are even less truth tracking. Steven Pinker (2005) speculates that this may account for the pervasiveness of cognitive biases and illusions, such as the well-known conjunction fallacy[7].

EDAs of types i–iii rely on premise 4*, as they emphasize the disconnection between truth and fitness. By contrast, EDAs of type iv rely on premise 6*. Such EDAs typically do allow for some connection between the fitness and truth-value of a belief, but they are

[7] In the conjunction fallacy, the probability of two conditions in conjunction (A&B) is regarded as (strictly) greater than that of a single one (A), which is incorrect according to probability theory (Kahneman et al., 1982).

pessimistic about the extent to which natural selection shapes our belief-formation processes. Let us focus on the former type first. EDAs that primarily rely on premise 4* do not give us good reasons to believe in the cogency of scientific reasoning, since they do not guarantee any link between fitness and the truth-value of specific beliefs. This line of reasoning is self-defeating. Evolutionary accounts according to which human cognitive capacities are so deeply biased and defective that knowledge is ruled out are self-undermining. There would be no good reason to assume that scientific theories are justified, or that philosophical reflection and argumentation (such as an EDA) provides us with sound conceptual knowledge. Also, there would be no reason to accept the soundness of psychological studies that indicate cognitive biases and heuristics. The very fact that cognitive psychologists (e.g., Kahneman et al., 1982) are able to recognize the frailties of human reasoning, and that subjects in psychological experiments perform better when their errors have been pointed out to them, suggests that biases and heuristics are not so pervasive as to cloud our reasoning completely, and that faulty reasoning is often corrigible with some mental effort.

In the same vein, EDAs that rely on premise 6* cast doubt on any type of belief that has no bearing on fitness. What kinds of beliefs would we be left with? The relatively high reproductive success of people who are not scientifically literate (like the Amish), or who actively oppose some forms of scientific knowledge (such as fundamentalist Christians) suggests that scientific beliefs do not have much impact on fitness (Kaufmann, 2010). Likewise, there is little reason to expect that our ability for philosophical reflection is subject to natural selection. But if scientific and philosophical beliefs belong to the type of beliefs that have no bearing on fitness, and if EDAs indeed affect any such beliefs, then even fine-grained EDAs that rely on premise 6* are potentially self-defeating. After all, these EDAs themselves are based on scientific theories, notably evolutionary theory, and philosophical reflection. Thus, if we accept that everything that is not directly adaptive or fitness-enhancing will be affected by an EDA, then EDAs of this type will inevitably bloat out to a lot of other beliefs as well, undermining their own coherence and leading one to doubt the cogency of scientific theories and philosophical argumentation. How do we avoid this self-defeat? One plausible solution is that, although higher-order theories as proposed in scientific and philosophical practice may be fitness-neutral, the cognitive skills they are based on need not be. As David Papineau (2000) has suggested, selective pressures may have enhanced human capacities for rational reasoning in the domains of folk psychology and means-end reasoning, for example in our ability to discern causes. An EDA that casts doubt on the adaptive value of *these* basic cognitive capacities (premise 6*) is much less plausible. Indeed, as Evan Fales has argued, the probability that such elaborate neural structures as are needed for these cognitive capacities would have evolved without conferring any adaptive value is *prima facie* quite low, given their high biological costs:

> *Homo sapiens* has, more than any other species, specialized in intelligence as a survival strategy. […] Our heavy investment in big brains and otherwise mediocre bodies makes it all the more unlikely that resources would be wasted on elaborate belief-forming and processing mechanisms that have no practical utility (Fales, 1996, 440).

If evolutionary approaches to the human mind are to be coherent, they should allow at the very least for cognitive capacities that are capable of generating truth-tracking theories, such as evolutionary theory. If we accept internal coherence as an important epistemic virtue, it seems that EAs are more promising than EDAs in the formulation of a naturalistic theory of mental content, since the former are not self-undermining. Therefore, the circularity charge, an important challenge faced by EAs, should be addressed. In the following section, two types of EA are presented that attempt to avoid this circularity problem.

3. Responses to the circularity charge

3.1 Dodging the bullet

One possible way out of the circularity problem is simply to grant the reliability of our inductive capacities as given. This is an externalist position: we need not *know* that beliefs are justified, it suffices that our cognitive processes *are* reliable to make them justified. W.V.O. Quine is perhaps the best-known proponent of this naturalistic position:

> I am not appealing to Darwinian biology to justify induction. This would be circular, since biological knowledge depends on induction. Rather, I am granting the efficacy of induction, and then observing that Darwinian biology, if true, helps explain why induction is as efficacious as it is (Quine, 1975, 70).

In this strategy, the epistemologist simply refuses to address the second-order question of whether our cognitive faculties are indeed reliable. This position is committed to scientific naturalism, which argues that epistemological questions should be approached through empirical science rather than through *a priori* philosophy. A weakness of this strategy is the problem of cheap knowledge, which is common to externalist positions in epistemology. Recall Plantinga's (1993) argument: if theism is true, we can expect that God has designed the human mind in such a way that our cognitive capacities successfully represent true states of affairs. Obviously, naturalists would object to this line of reasoning, but there seems to be little to distinguish between the naturalistic explanation (by appeal to evolutionary theory) and the theistic explanation (by appeal to Christian revelation). The naturalist can appeal to the primacy of natural science. Since evolutionary theory is a scientific theory, it is—according to the naturalist—more trustworthy than skeptical philosophical arguments. Thus, naturalized epistemologists can take evolutionary theory to be correct, and use this theory to explain why our inductions are reliable. Similarly, the theist can appeal to the primacy of God as first cause. However, this problem of cheap knowledge might be circumvented if we consider naturalism not to be an *a priori* philosophical position, but a historical, well-established, though provisional, result of scientific inquiry. In contrast to supernatural accounts, naturalistic explanations have been consistently successful. With the efficacy of induction taken for granted, and bearing in mind the success of the naturalistic program in science, the naturalist can then argue that evolution by natural selection indeed explains induction better than a theistic first cause (Boudry et al., 2010).

3.2 Biting the bullet

Instead of simply refusing to address the problem of justifying our most basic principles of reasoning, some authors have bitten the bullet and proposed ways out of the circularity charge. F. John Clendinnen (1989) has argued that this part of the project of evolutionary epistemology will be inevitably circular, but it need not be viciously so. According to him, interdependence by itself is not sufficient to establish vicious circularity. Consider the triplet of propositions "If P then Q", "If Q then R", "If R then P". Although this form of interdependence should make us wary, the accusation of vicious circularity holds only when the circle is completely closed, i.e., when we have *no other reasons at all* to accept any of the propositions involved. In a deductive model of justification, it is often incorrectly assumed that each step is either completely justified or not justified at all. However, it is possible to have an interdependence of justifications (P, Q, R) in which we start out by an initially very weak and provisional version of P, in which our confidence is gradually raised as evidence accumulates (through Q and R). In particular, Clendinnen (1989) believes that there are two reasons, independent from natural facts, for accepting the principle of induction as minimally rational: the criterion of non-arbitrariness and the principle of simplicity.

Clendinnen (1989) does not believe that we can justify the whole of the scientific method *a priori*. Rather, he argues that, by starting from a minimally rational principle of induction, we may accumulate evidence (including evolutionary theory) that itself lends support to the efficacy of induction: "[i]nduction itself, once accepted as a minimal principle, may be used to interpret the available evidence for or against the thesis that the world is the kind of place in which induction is likely to succeed" (Clendinnen 1989, 468). Thus, by bootstrapping her way out of the circularity, the evolutionary epistemologist is able to justify the inferences on which her own belief in evolutionary theory hinges.

The Bayesian epistemologist Tomoji Shogenji (2000) proposes another way out of the circularity problem. He argues that many forms of self-dependent justification are in fact not circular—rather, they are unproblematic forms of Bayesian confirmation. Standard Bayesian confirmation is a procedure that strengthens the subjective belief in a hypothesis H by observation O relative to background belief B if

1. The probability of H & B is strictly positive.
2. The probability of O given H & B is higher than the probability of O given B alone.

In this general form of Bayesian confirmation, we compare Prob(O|H & B) and Prob(O|B). If the former is higher than the latter, we can conclude that O confirms H. Consider as an example empirical evidence for a scientific theory: empirical evidence supports a theory if the evidence is more likely to occur under the assumptions of the theory combined with our background beliefs than it would be under our background beliefs alone. For example, the unusually slow precession of Mercury's orbit around the Sun (O) is in conflict with standard Newtonian mechanics (N) and our background beliefs about the physical universe (B), but this observation is predicted under the theory of general relativity (H). Thus, this observation disconfirms N, Prob(O|N&B) < Prob(O|B), and confirms H as Prob(O|H&B) > Prob(O|B).

Self-dependent justification differs from this general form in that H plays a peculiar role, as it is both the hypothesis that is being tested and a part of the background beliefs. If there is more to background beliefs than H, one can filter H out of B to obtain B*. Thus, in such cases, the hypothesis is contained within the set of background beliefs but it is not identical to the background beliefs. To apply Bayesian confirmation to those cases, we need to replace B with B* & H. The first of the two conditional probabilities then becomes Prob(O|H & B* & H). As Prob(O|H & B* & H) = Prob(O|H & B*), the first conditional can be simplified as Prob(O|H & B*). In this construal, H no longer plays the split role of hypothesis and background belief, but rather, it plays the role of hypothesis twice in predicting the probability of O. As H is no longer part of the background beliefs under this construal, the proper second conditional becomes Prob(O|B*).

Shogenji (2000, 294) applies this model to evolutionary epistemology. Here H stands for "perceptual process P is reliable"; O represents "S believes that P is reliable", and B* stands for "when it is used in empirical investigation, perceptual process P generates a belief in the perceiver, to which she has an introspective access to form a metabelief; her introspection, memory, etc., which do not depend on her perception, are reliable." In this case, O (S's belief in the reliability of her perceptual processes) does confirm H, because there is no reason from B* alone to assume that S's perceptual processes would be reliable; her perceptual processes could generate any kind of belief, and it seems very unlikely that among those beliefs would be her conviction that P is reliable. On the other hand, H confers a high probability on those beliefs. In this way, the reliability of perceptual processes can be tested by simple Bayesian confirmation.

Shogenji's (2000) case for evolutionary epistemology could be strengthened. The case as he presents it relies on a theory-dependent observation O. But current evolutionary

approaches to the human mind frequently appeal to observations that are not dependent on evolutionary theory itself—of course, they are dependent on other theories, but crucially, they do not depend on the specific evolutionary hypotheses they set out to test. Such observations, which are neutral with respect to the theory that is being tested, are denoted with O* (a term borrowed from Adam, 2004). Over the past decades, evolutionary psychologists[8] have used findings from neuroscience and cognitive, comparative and developmental psychology to test evolutionary hypotheses (see e.g., Dunbar & Barrett, 2007). Such O* types of observation do not assume that evolutionary theory is true, and therefore can confirm evolutionary hypotheses about cognition without circularity. Take, for example, the observation in comparative psychology that humans and monkeys are fast and accurate in their categorization of various kinds of stimuli. Rhesus monkeys can reliably sort pictures into food and nonfood categories, even if the pictures show items they are unfamiliar with (Fabre-Thorpe et al., 1998). Moreover, they can do this in a very brief period (pictures are flashed for a duration of only 80 ms), and they are correct in approximately 90 % of the trials. Next to this, a wealth of experimental data from developmental psychology (e.g., Farroni et al., 2005) shows that human newborns have a visual preference for face-like stimuli. Moreover, young infants can recognize their own mothers from other women with similar clothes and hairstyle within a few hours after birth (Bushnell, 2001). These observations do not rely on the supposition that evolutionary theory is correct, but they do strengthen evolutionary arguments for the reliability of our beliefs. The categorization studies indicate that monkeys are capable of correctly categorizing stimuli. Their proficiency given the hypothesis that natural selection has endowed them with cognitive capacities that are at least truth preserving under some conditions is more likely than their proficiency given everything else we know about the physical world: the complexity of visual scenes, together with the lack of previous experience with the test items leads to the expectation that the monkeys would perform at chance level, $Prob(O^*|H \,\&\, B^*) > Prob(O^*|B^*)$. Pace Darwin's (1881) earlier-mentioned skepticism about monkeys' minds, there are indeed good reasons to trust a monkey's convictions, at least when it comes to discriminating food from nonfood. Likewise, the studies with face-recognition in human newborns provide us with an observation that seems vastly improbable given our background knowledge about their lack of experience (having spent their time in the dark environment of the womb) and their poor visual acuity. By contrast, the observation is more likely if we accept an evolutionary hypothesis that proposes that humans are equipped with an evolved, unlearned capacity to recognize faces. Such an ability would have been adaptive for primates living in social groups in order to recognize each other, especially given that diurnal primates like ourselves have less-developed olfaction compared to most other mammals. In sum, circularity can be avoided by a careful rephrasing of EAs, and by reliance on observations outside the domain of evolutionary theory. Nevertheless, it remains to be seen whether non-circular EAs could be constructed for other cognitive processes, such as inductive inferences. After all, neuroscience and other scientific disciplines do not deliver brute facts, but require interpreting, cognizing minds. If these inferential processes are at stake, it is hard to see how one can construct a non-circular EA.

4. Extended cognition and evolved cognitive biases

As we have seen, EAs are most successful for commonsense beliefs. What justification can we get for other types of beliefs, like scientific knowledge, mathematical results, or

[8] Evolutionary psychologists examine cognitive capacities (especially those of humans) as a product of functional design caused by natural and sexual selection.

philosophical argumentation, which may be outside the grasp of natural selection—in other words, how can we counter premise 6* of EDAs? Additionally, how can we counter premise 4* for cognitive capacities where we may expect that natural selection has shaped cognitive capacities that do not track truth? In subsection 2.2 we suggested that although scientific and philosophical beliefs may be fitness-neutral, this does not imply that the cognitive skills on which they are based would be fitness-neutral. A capacity like induction is useful in a wide variety of settings, not just in the context of academic reflection. Additionally, in section 3 we outlined several strategies that allow for self-dependent justification in EAs, and that avoid the circularity charge.

In what follows, we explore a further strategy to argue that humans can have reliable beliefs despite cognitive biases. We will argue that the way humans naturally supplement their evolved cognitive capacities with external tools may provide an effective way to increase the reliability of their beliefs and to counter evolved cognitive biases. Outsourcing cognitive tasks to the external environment enhances cognition in several ways. First, it improves conceptual stability. Thinking sometimes involves complex manipulations of conceptual structures, as in logical reasoning or carrying out large calculations. The stability of these manipulations (e.g., substituting a constant by another in a logical proof) is greatly enhanced by writing down each step (Clark, 1996; De Cruz & De Smedt, in press). Second, extending the mind may provide a way around evolved cognitive biases. For example, a large experimental literature (see Loftus, 2003, for a review) indicates that people's episodic memory (i.e., biographical memory of personal experienced events) is highly constructive and liable to distortion. For instance, people typically remember their worst train-missing experience when simulating how painful and inconvenient a next train-missing experience will be (Morewedge et al., 2005). This puzzling feature of episodic memory can be explained by the hypothesis that its function is not one of disinterested representation of true events, but one of building simulations that guide our actions in adaptive ways. Overestimating the discomfort of an unpleasant experience may help us to avoid that situation in the future. Adaptive as this may be, it poses severe limitations on the reliability of our long-term episodic memories, which typically get distorted over time. Artifacts, such as books, journals, electronic storage devices, measuring instruments, calendars, or even simple tallies, allow us to store information that is cognitively challenging to memorize, and to protect it from memory distortion. In this way, an EDA that would call the reliability of human memory (a commonsense belief) into question could be countered as follows: human episodic recall may be biased, but humans can mitigate this by using external memory systems.

The use of external media is not limited to contemporary societies, but seems to be a pervasive element of human cognition at least since the late Pleistocene (ca. 120,000-12,000 years ago). From this period onward, archeologists find notched pieces of ochre and bone, shell beads and representational art, demonstrating that humans conveyed ideas externally in symbolic media (De Smedt & De Cruz, 2011). Kim Sterelny (2003) incorporates theories of niche construction in human cognitive evolution, arguing that, just as termites build their own environment, humans construct their own cognitive niche using artifacts to suit epistemic purposes. Niche construction is the process in which organisms change their own selective environments, thereby influencing their evolutionary history. Common examples include termites, beavers and ants. Members of these species build complex artifacts to ameliorate their environment in terms of temperature and humidity, and to provide an optimal nursery setting for offspring (Laland & Brown, 2006). Similarly, humans improve their environment by building houses, boats and other artifacts; they also improve their epistemic environment by designing tools to suit their cognitive purposes. If the extended mind is indeed a key feature of human cognitive evolution, it might be possible to justify non-commonsense beliefs through EAs, namely as those beliefs that reliably arise through the judicious use of external

media which can counter cognitive biases. Of course, the use of external tools does not guarantee that the beliefs acquired or stored in this way will always be reliable. It does, however, provide an evolutionary explanation for why humans can have knowledge that goes beyond their commonsense intuitions.

Temperature provides an interesting case study of the biases of external perception, and the use of external media in surmounting these. Peripheral thermoreception is the system that reacts to surface skin temperature. As Kathleen Akins (1996) compellingly argues, this system does not provide us with a reliable representation of the external world. Rather, it tends to produce representations that are especially conducive to an organism's fitness. There are four kinds of thermoreceptors in the skin: "warm spots", "cold spots" and two types of pain receptors that fire under conditions of extreme heat or cold[9]. The ratio between warm spots and cold spots is not evenly distributed across the body, for example, the lips are almost exclusively sensitive to warmth, and the scalp is mainly sensitive to cold. Also, the sensitivity of these receptors is not a linear function, and their response critically depends upon the starting temperature. The change from tepid to warm water, for example, elicits less response than that from warm to hot water. These curious properties of thermoreception can be explained in evolutionary terms—what is important to warm-blooded creatures like us is that our perception of temperature helps us to act adaptively in the world, to avoid injury, overheating and hypothermia. Thus, it makes sense that the human scalp is sensitive to cold, since excessive cooling of the brain is potentially life-threatening. The fact that thermoreception is concerned with an organism's fitness, rather than with an accurate, disinterested rendition of temperature, makes it vulnerable to various biases. For example, if you place your left hand in hot water and your right hand in cold water, they will register different temperatures when put in the same lukewarm water. Seen from the perspective of an organism that has evolved mechanisms to maintain a constant body temperature, such illusions make sense.

Humans, being used to rely on external media, have come up with an elegant solution to the inherent biases of thermoreception: find fixed points outside of our own thermoreceptive experience and use these as a scale of reference for external devices that register temperature. Already since the 1600s, scientists who attempted to make reliable thermometers had a preference for observer-independent points, such as the melting point of butter or the boiling point of wine. Hasok Chang (2004) provides a detailed account of the human quest for standardized temperature measurement and of the establishment of the (to some extent idealized) fixed points of freezing and boiling water on the Celsius scale. To be sure, the fact that we place epistemic trust in such external devices does require the reliability of other intuitions and cognitive faculties, such as transitivity (if x is warmer than y, and y is warmer than z, then x is warmer than z), and a trust in the existence of general laws of nature. Nevertheless, relying on external measuring devices provides a solution for at least one set of cognitive biases, namely those of human thermoreception.

Once humans were able to outsource their measurement of temperature to the outside world, they were not only able to measure temperature in a more objective way, but also in a much more precise fashion than if they relied on their evolved capacities alone. Interestingly, the practical advances in thermometry also contributed to the conceptual understanding of thermodynamics, such as the understanding that cold is not on an ontological par with heat, and the formulation of an absolute zero point. The folk intuition that cold and heat are both equally real phenomena may be derived from our tactile perception of temperature (the occurrence of cold spots and warm spots), an intuition that was countered by experiments that

[9] These data on thermoreception are derived from Akins (1996).

indicated that cold is just the absence of heat. The fact that current thermodynamics flies in the face of our evolved capacities for thermoreception provides at least *prima facie* evidence against a sweeping applicability of EDAs. Humans are able to surmount the biases of their evolved thermoreception through external tools. To put it differently: we have an intuition that cold really exists and there is a good evolutionary rationale for this. In spite of this, we can come to believe that cold is merely absence of heat—for this too, there is a good evolutionary explanation, namely our evolved ability to use external media in epistemic contexts.

5. Conclusion: a defeasible evolutionary account

What are the consequences of evolutionary theory for the epistemic standing of our beliefs? The development of EAs is faced with several challenges. Advocates of EDAs point out that there is no connection between fitness and truth-value (premise 4*) or that the force and scope of natural selection may be constrained for several reasons (premise 6*). In addition, EAs are open to the charge of circularity. We have explored venues for setting up EAs: the externalist solution, which uses bootstrapping by way of benign as opposed to vicious circularity and the use of observations that are independent from evolutionary hypotheses. Next to this, we have proposed that incorporating external media in human cognition provides us with a defeasible EA that indicates that knowledge outside of the domain of common sense is at least a *prima facie* possibility. Truth-approximating or instrumentally useful knowledge can be attained by reliance on external tools that are independent of human cognitive biases. Of course, emphasizing the role of external media in human cognition does not provide us with a universal, all-encompassing epistemic justification for human knowledge. What it does provide is a plausible reason, from an evolutionary perspective, of why humans can have reliable knowledge outside of domains where accurate knowledge matters for survival and reproduction, like, for example, scientific knowledge. As we have seen, EAs that do not appeal to the extended mind seem limited to the domain of commonsense knowledge. The role of external media in human cognition can provide an epistemic justification for some non-commonsense beliefs that humans entertain[*].

REFERENCES

ADAM, M. 2004, "Why worry about theory-dependence? Circularity, minimal empiricality and reliability", *International Studies in the Philosophy of Science* 18, pp. 117-132

AKINS, K. 1996, "Of sensory systems and the "aboutness" of mental states", *Journal of Philosophy* 93, pp. 337-372

BERING, J.M. 2006, The folk psychology of souls. *Behavioral and Brain Sciences* 29, pp. 453-462

BERING, J.M. 2011, *The God instinct. The psychology of souls, destiny and the meaning of life*, London: Nicholas Brealy

[*] Acknowledgments: We express our gratitude to Anne Meylan, Igor Douven, and an anonymous reviewer for their elucidating comments on and suggestions to an earlier version of this paper. This research is supported by the Research Foundation Flanders and grants COM07/PWM/001 and BOF08/24J/041 from Ghent University.

BLOOM, P. 2009, "Religious belief as an evolutionary accident", in: J. Schloss & M.J. Murray, eds., *The believing primate. Scientific, philosophical, and theological reflections on the origin of religion*, Oxford: Oxford University Press, pp. 118-127

BOUDRY, M., BLANCKE, S. and BRAECKMAN, J. 2010, "How not to attack Intelligent Design Creationism: Philosophical misconceptions about methodological naturalism", *Foundations of Science* 15, pp. 227-244

BOULTER, S.J. 2007, "The "evolutionary argument" and the metaphilosophy of commonsense", *Biology & Philosophy* 22, pp. 369-382

BUSHNELL, I.W.R. 2001, "Mother's face recognition in newborn infants: Learning and memory", *Infant and Child Development* 10, pp. 67-74

CHANG, H. 2004, *Inventing temperature: Measurement and scientific progress*, Oxford: Oxford University Press

CLARK, A. 1996, "Linguistic anchors in the sea of thought?", *Pragmatics & Cognition* 4, pp. 93-103

CLENDINNEN, F.J. 1989, "Evolutionary explanation and the justification of belief", in: K. Hahlweg and C.A. Hooker, eds., *Issues in evolutionary epistemology*, Albany, NY: State University of New York Press, pp. 458-474

DARWIN, C. 1871, *The descent of man, and selection in relation to sex*, London: John Murray

DARWIN, C. 1881, Letter 3230—Charles Darwin to William Graham, 3 July 1881, retrieved from http://www.darwinproject.ac.uk/entry-13230 on 20 May 2010

DAWKINS, R. 2006, *The God delusion*, Boston: Houghton Mifflin

DE CRUZ, H. and DE SMEDT, J. in press, "Mathematical symbols as epistemic actions", *Synthese*

DENNETT, D.C. 1995, *Darwin's dangerous idea. Evolution and the meanings of life*, London: Allen Lane

DE SMEDT, J. and DE CRUZ, H. 2011, "The role of material culture in human time representation. Calendrical systems as extensions of mental time travel", *Adaptive Behavior* 19, pp. 63-76

DUNBAR, R.I.M. and BARRETT, L. (eds.) 2007, *Oxford handbook of evolutionary psychology*, Oxford: Oxford University Press

FABRE-THORPE, M., RICHARD, G. andTHORPE, S.J. 1998, "Rapid categorization of natural images by rhesus monkeys", *NeuroReport* 9, pp. 303-308

FALES, E. 1996, "Plantinga's case against naturalistic epistemology", *Philosophy of Science* 63, pp. 432-451

FARRONI, T., JOHNSON, M.H., MENON, E., ZULIAN, L., FARAGUNA, D. and CSIBRA, G. 2005, "Newborns' preference for face-relevant stimuli: Effects of contrast polarity", *Proceedings of the National Academy of Sciences of the USA* 102, pp. 17245-17250

GOLDMAN, A. 1990, "Natural selection, justification, and inference to the best explanation", in: N. RESCHER (ed.), *Evolution, cognition and realism. Studies in evolutionary epistemology*, Lanham: University Press of America, pp. 39-46

GUTHRIE, S. 1993, *Faces in the clouds: A new theory of religion*, New York: Oxford University Press

KAHANE, G. 2011, "Evolutionary debunking arguments", *Noûs* 45, pp. 103-125

KAHNEMAN, D., SLOVIC, P. and TVERSKY, A. 1982, *Judgment under uncertainty: Heuristics and biases*, Cambridge: Cambridge University Press

KAUFMANN, E. 2010, *Shall the religious inherit the earth? Demography and politics in the twenty-first century*, London: Profile Books

KELEMEN, D. 2004, "Are children "intuitive theists"?", *Psychological Science* 15, pp. 295-301

LALAND, K.N. and BROWN, G.R. 2006, "Niche construction, human behavior, and the adaptive-lag hypothesis", *Evolutionary Anthropology* 15, pp. 95-104

LOFTUS, E.F. 2003, "Make-believe memories", *American Psychologist* 58, pp. 867-873

MCKAY, R.T. and DENNETT, D.C. 2009, "The evolution of misbelieve", *Behavioral and Brain Sciences* 32, pp. 493-510

MILLER, G. 2000, *The mating mind. How sexual choice shaped the evolution of human nature*, London: William Heineman

MILLIKAN, R. 1984, *Language, thought, and other biological categories*, Cambridge, Ma.: MIT Press

MOREWEDGE, C.K., GILBERT, D.T. and WILSON, T.D. 2005, "The least likely of times: How remembering the past biases forecasts of the future", *Psychological Science* 16, pp. 626-630

PAPINEAU, D. 2000, "The evolution of knowledge", in: P. CARRUTHERS and A. CHAMBERLAIN, eds., *Evolution and the human mind. Modularity, language and meta-cognition*, Cambridge: Cambridge University Press, pp. 170-206

PINKER, S. 2005, "So how *does* the mind work?", *Mind and Language* 20, pp. 1-24

PLANTINGA, A. 1993, *Warrant and proper function*, Oxford: Oxford University Press

QUINE, W.V.O. 1969, *Ontological relativity and other essays*, New York: Columbia University Press

QUINE, W.V.O. 1975, "The nature of natural knowledge", in: S. GUTTENPLAN, ed., *Mind and language: Wolfson College lectures*, Oxford: Clarendon Press, pp. 67-81

RAMSEY, G. 2006, "Block fitness", *Studies in the History and Philosophy of Biology & Biomedical Sciences* 37, pp. 484-498

REID, T. 1764, *An inquiry into the human mind, on the principles of common sense*, Edinburgh: Millar, Kincaid & Bell

SHOGENJI, T. 2000, "Self-dependent justification without circularity", *British Journal for the Philosophy of Science* 51, pp. 287-298

STEPHENS, C.L. 2001, "When is it selectively advantageous to have true beliefs? Sandwiching the better safe than sorry argument", *Philosophical Studies* 105, pp. 161-189

STERELNY, K. 2003, *Thought in a hostile world: The evolution of human cognition*, Oxford: Blackwell

STEWART-WILLIAMS, S. 2005, "Innate ideas as a naturalistic source of metaphysical knowledge", *Biology & Philosophy* 20, pp. 791-814

STICH, S.P. 1990, *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*, Cambridge, Ma.: MIT Press

WALSH, D.M., LEWENS, T. and ARIEW, A. 2002, "The trials of life: Natural selection and random drift", *Philosophy of Science* 69, pp. 452-473

WENGER, A. and FOWERS, B. J. 2008, "Positive illusions in parenting: Every child is above average", *Journal of Applied Social Psychology* 3, pp. 611-634