Detecting Design: Fast and Frugal or All Things Considered?

Graham Wood

School of Philosophy
University of Tasmania
Launceston, Tasmania, Australia, 7250
Graham.Wood@utas.edu.au

Abstract. Within the Cognitive Science of Religion, Justin Barrett has proposed that humans possess a *hyperactive agency detection device* that was selected for in our evolutionary past because 'over detecting' (as opposed to 'under detecting') the existence of a predator conferred a survival advantage. Within the Intelligent Design debate, William Dembski has proposed the *law of small probability* that states that specified events of small probability do not occur by chance. Within the Fine-Tuning debate, John Leslie has asserted a *tidiness principle* such that, if we can think of a good explanation for some state of affairs, then an explanation is needed for that state of affairs. In this paper I examine similarities between these three proposals and suggest that they can all be explained with reference to the existence of an *explanation attribution* module in the human mind. The forgoing analysis is considered with reference to a contrast between *classical* rationality and what Gerd Gigerenzer and others have called *ecological* rationality.

The *hyperactive agency detection device* and types of rationality

Justin Barrett has suggested that humans are predisposed to believe in the existence of supernatural agents because we have a "hyperactive agent detection device" (Barrett 2000: 31). The idea, first suggested by Guthrie (1980), is that in our evolutionary past it would have been much better to 'over-detect' the presence of predators than under-detect predators. In Guthrie's words, "one real enemy justifies a hundred false alarms" (1980: 190). Thus, the suggestion is that our minds evolved in such a way as to err on the side of caution when presented with ambiguous sensory data. Put simply, when individuals were presented with ambiguous sensory data, individuals who (at least initially) interpreted that data as indicating the presence of a predator produced more offspring than those that did not. This is a reasonable theory. After all, if you treat ambiguous sensory data, data that is actually associated with a predator, as just a trick of the light and you are wrong, you then pay a high price. Furthermore, it is suggested that it is this disposition to 'over detect' agency in the natural world that facilitates belief in supernatural agents:

> A key feature of the supernatural agent concepts common to all religions is the triggering of an "Innate Releasing Mechanism", or "agency detector" whose proper (naturally selected) domain encompasses animate objects relevant to hominid survival—such as predators, protectors, and prey—but which actually extends to moving dots on computer screens, voices in the wind, and faces in clouds. (Atran & Norenzayan 2004: abstract)

This suggestion raises two important themes. Firstly, it embraces evolutionary psychology, which holds that:

> [Cognitive] systems, like other systems in the body, have been shaped by natural selection to perform specific functions or to solve information processing problems that were important in the environment in which our hominid ancestors evolved. (Samuels *et al.* 1999: section 4)

And, from the perspective of evolutionary psychology, it is assumed that understanding the evolutionary origins of our cognitive systems can help us understand the functioning of those systems.

Secondly, the suggestion highlights the role of false beliefs in a belief system. There seems to be a clear survival advantage in holding (at least initially) false beliefs about the presence of predators in the form of 'false positives' when contrasted with 'false negatives'. So, drawing these two themes together, there seem to be 'good' reasons to hold false beliefs, and the 'goodness' of these false beliefs is explained with reference to evolutionary psychology.

But the suggestion that there can be good reasons to hold false beliefs seems to be in some tension with what might be called classical rationality. Let us say that a human is classically rational if his or her beliefs accord with logic and the probability calculus. Let us also assume that classical rationality is aiming at believing truths and avoiding believing falsehoods. But evolutionary psychology poses significant challenges to the assumption that humans are rational in this classical sense. If our cognitive architecture and psychological processes are the products of evolution, then why assume that we are rational in this classical sense? Indeed, there are many studies that suggest we are not rational in this sense (Kahneman, Slovic & Tversky 1982). If the only driving force here is natural selection, then as long as our cognitive processes lead to the production of more offspring, it seems that these processes need not be rational in the classical sense. Of course, natural selection does not preclude classical rationality, but it does not ensure it either.

If evolution is the process responsible for generating our cognitive processes, what sort of rationality should we expect to have? We should expect to have a 'rationality' that increases our fitness, and that may not be simply *classical* rationality. Believing truths and avoiding believing falsehoods, and the application of logic and the probability calculus may go some way to increasing the fitness of humans, but that might not be the whole story. To illustrate the fact that there may be more than one type of rationality at work in human reasoning I want to contrast what I am calling classical rationality with what Gerd Gigerenzer and others have called 'ecological rationality'. They write that:

> [R]ationality can be found in the use of fast and frugal heuristics, inference mechanisms that can be simple and smart. The laws of logic and probability play little if any role in performance of these components of the mind's adaptive tool box—these heuristics are successful to the degree to which they are ecologically rational, that is, adapted to the structure of the information in the environment in which they are used to make decisions. (Gigerenzer *et al.* 1999: vii)

As I have suggested, evolutionary psychology raises important questions about our rationality. For example, does it mean that we are not rational? If we take Gigerenzer and his collaborators seriously, it seems that the question of human rationality is now more complicated. If we were just asking the question with respect to the norms of classical rationality we could simply compare our reason with, say, logic and probability theory. If we reason in accord with these, then we are rational; if not, then we are not rational. But now, it seems, we first have to decide which type of rationality we are talking about: classical rationality or ecological rationality. For example, consider the functioning of the *hyperactive agency detector* and its tendency to err on the side of caution. Is it rational to 'over-detect' the presence of predators? Is it rational to 'detect' agency or design when there may not be any? Well, if we are using the norms of classical rationality, presumably the answer is 'no'. Classical rationality, one would assume, aims for true beliefs, and reasoning that leads systematically to false belief must be bad reasoning. But once we admit of the possibility of assessing reasoning from the perspective of evolution, then the judgment as to whether particular reasoning strategies are good or bad is more complex. Specifically, there may be pragmatic justifications for believing falsehoods (at least initially). Another way of thinking about this is to think in terms of prudential rationality. If an organism systematically errs on the side of caution with respect to ambiguous sensory data, and by so doing survives, this could be considered prudentially rational.

Along with the distinction between classical rationality and ecological rationality, the distinction between unbounded rationality and bounded rationality (Simon 1957) is relevant here. Central to the approach of classical rationality is the notion of an ideal rational agent. The reasoning of an ideally rational agent employing, say, Bayesian confirmation theory is not limited by anything: it is unbounded. Of course, discussions of classical rationality acknowledge the unrealistic nature of an ideal rational agent. But even if the unrealistic nature of such an agent is acknowledged, classical rationality still implicitly, or perhaps even explicitly, assumes such an agent is the reference against which we are to measure our own reasoning. However, we are not unbounded rational agents. So, perhaps, we need to use a different reference when assessing whether or not certain processes of reasoning are rational.

What significance do these points have for the topic at hand? Perhaps, from the perspective of ecological rationality, erring on the side of caution when it comes to 'detecting' the presence of predators is very rational. After all, it is so important to get it right when avoiding being eaten that it may be reasonable to allow a high rate of false positives. And this permissiveness would not only be reasonable in the case of the detection of predators (of the non-human variety). The 'over-detection' of agency *in general* (whether or not the detected agent is friend or foe) may be a good strategy. As pointed out by Guthrie, "Objectively and subjectively, other people are, to virtually everyone, the most important entities in the world" (1980: 188). So it would not be surprising if natural selection produced cognitive systems strongly biased to detect the presence of agents in the environment and that natural selection would not ruthlessly eliminate 'false positives' produced by such systems. Talking to a tree when mistakenly thinking it was a person has a lower social and hence evolutionary cost than not talking to a person when mistakenly thinking he or she was a tree.

What I hope to have highlighted with these comments is a general contrast between the 'all things considered' reasoning associated with classical rationality and what Gigerenzer has called 'fast and frugal' reasoning (2000: 166) associated with ecological rationality. To illustrate the distinction between all things considered reasoning and fast and frugal reasoning, consider the 'recognition heuristic' (Gigerenzer *et al.* 1999: 37-58). If the non-Germans amongst us are asked to choose which of two German cities has the larger population, and we only recognise the name of one of them, then it is a good bet that the one we recognise is the larger city. It is not clear to me that the application of classical rationality could give a helpful all things considered response here, but the fast and frugal recognition heuristic can. And in the real world of limited time and information fast and frugal heuristics have a significant role to play, and it does not seem too unlikely that they have also had a role to play in our evolutionary development.

Now let us return to the question of rationality. Are fast and frugal heuristics, like the recognition heuristic, rational? Should we use them to decide on beliefs? Well, it comes down to the question of what norms of rationality are being used. There is some tension between fast and frugal reasoning and all things considered reasoning. Using all things considered reasoning we may well look down upon the outputs of fast and frugal reasoning. Indeed, all things considered, we may come to the conclusion that fast and frugal reasoning is not to be trusted. And, if we are measuring reasoning with the yard-stick of classical rationality, then perhaps fast and frugal reasoning is indeed not to be trusted. But if we are measuring reasoning with the yard-stick of ecological rationality, fast and frugal reasoning may well be more reasonable than all things considered reasoning.

William Dembski's *law of small probability*

Gigerenzer and his colleagues have identified a number of fast and frugal heuristics (2000), and I have characterised the *hyperactive agency detection device* as a type of fast and frugal reasoning. But are there other examples? I think there are. Consider Dembski's *law of small probability*: "specified events of small probability do not occur by chance" (1998: 5). Dembski has embedded his law in his *explanatory filter* which he claims can determine the type of explanation that will explain a certain event. I do not wish to critically engage with the filter as a whole because this has been done well elsewhere (Fitelson *et al.* 1999). What I am interested in is Dembski's law of small probability. I suggest this 'law' is not a law at all, but a fast and frugal heuristic.

But before classifying this law, I will highlight some of its central features. Firstly, Dembski characterises this as a *law*: specified events of small probability *do not* occur by chance. Secondly, this law only concerns events of small probability.[1] Thirdly, this law only relates to 'specified' events. And it is the so-called 'specification' of an

[1] We actually need to interpret these as events of 'apparently' small probability, because the whole point of Dembski's law is to conclude that these events of apparently small probability did not occur by chance after all. Once we accept some other explanation for them, presumably, they cease to be events of small probability. To avoid complications that are not relevant to our current discussion, let us assume that we can identify events of (apparently) small probability in the world.

event that is relevant here. On my reading of Dembski, a specified event is an event that can be described independently of any knowledge of the actual occurrence of the event (1998: 136-74). An example will illustrate.

Let's say that the 'single event' defined as the outcome of tossing a coin 100 times is an event of small probability. Each possible series of 100 heads and tails would be one such event, and all would have a very small (equal) probability of occurring (by chance). Now, imagine that I asked you to write down a series of 100 'H's and 'T's to represent the tosses of a coin, and then I produced a coin and began tossing it. As you watch, the coin lands in exactly the series of heads and tails that you specified. What would be your reaction? Would you simply say, "Oh well, each series of heads and tails has an equal probability of occurring, and it just so happened that the series that did occur matched my series of 'H's and 'T's." Or would you say, "Hang on a minute, something is going on here!"? I, for one, would be very reluctant to attribute that event to chance. This illustration, I suggest, captures the essence of the notion of a specified event of small probability.[2] And the rejection of chance, that I imagine would be your response, captures the essence of the *law of small probability*. Finally, Dembski also allows for the event to be specified after the event has occurred, as long as the event can be described independently of any knowledge of the actual occurrence of the event (1998: 138-42).

So, what should we make of Dembski's law? Is it rational to assert such a law? I think there is something right about it. After all, I assume that most people would be uncomfortable attributing the above coin toss to chance. So Dembski is onto something. But, on the other hand, there is something not quite right about it. After all, if there were enough people writing out series of 'H's and 'T's and enough subsequent coin tosses, then the specified event could occur by chance. So in that sense 'law' seems too strong.[3]

Again, the question of interest is this: is it rational to assert this law? The answer seems to depend on what norms of rationality one is using. Using the norms of classical rationality, it is not rational to assert the law. This point is made by Elliot Sober when he asks: "Is there a Law of Improbability that begins with the premise that Pr(O/H) is very low and concludes that H should be rejected? There is no such principle" (2003: 34). Or, put another way, if we employ Bayesian confirmation theory perhaps all we can say is that the 'rigged' hypothesis has a higher posterior probability than the 'chance' hypothesis. So, following Sober and classical rationality (in the form of Bayesian confirmation theory), we cannot simply reject the chance explanation in the coin toss example above. All things considered, we cannot reject the chance explanation. But is that really rational? Perhaps there is another way to look at this whereby it is rational to simply reject the chance explanation.

If we adopt the norms of fast and frugal reasoning, perhaps the law is rational. Perhaps, if an event would otherwise be really improbable and we can specify that event, then it is a good bet that we are onto something and furthermore that chance is not the explanation of that event. To get an idea of how this might be rational in some sense

---

[2] Dembski would say that the probability in this example is not 'small' enough (1998: 175-223), but it will do for illustrative purposes.

[3] According to Dembski, if there were enough people writing out series of 'H's and 'T's and enough subsequent coin tosses, then the event would cease to be an event of small probability and the law would not apply.

consider the process of specification. Specification is the description of an event. But how is the specification produced? Presumably it is produced with reference to some explanatory schema. Perhaps we have some explanatory schema (that is not just 'chance') in mind and we generate a specification with reference to that schema. If we can specify an event in this way, it is a very good bet that chance is not the explanation of that event.

Thus I suggest this law looks a lot more reasonable when assessed with reference to fast and frugal reasoning. Why? Because the law is doing some useful work for us. Of all the many apparently improbable events that we observe, which ones should we attend to? Which ones should we spend time considering? Which ones should we try and explain?

Consider two simple examples. Is the apparently very improbable event of that leaf falling in just that way onto that particular patch of grass worth attending to? Or is the apparently very improbable event of that patch of darkness moving toward me in just the way it would if it was the shadow of a tiger worth attending to? I suggest there are very good pragmatic reasons to attend to the second and not the first. Furthermore, I suggest it is rational (at least initially) to exclude chance in the second scenario and not the first. And that is because I can specify the latter event in a way that I cannot specify the former. When considered with reference to ecological rationality Dembski's law looks quite reasonable. But perhaps 'law' is going a little too far. Perhaps we can think of Dembski's insight not so much as identifying a law, but as identifying a heuristic. So in place of Dembski's law, I offer the following *small probability heuristic*:

> *If an event of apparently small probability occurs that I can specify, then I should not initially assume chance is a reasonable explanation for that event.*

John Leslie's *tidiness principle*

Now is this heuristic at work anywhere else? I think it is. Consider these two passages from Leslie.

> You seem to see mere rubbish in your opponent's poker hand of an eight, six, five, four, and three. It is natural to assume that Chance gave it to him. But you then recall that poker has many versions; that you had agreed on one in which his Little Tiger ('eight high, three low, no pair') defeats your seemingly much stronger hand; that a million dollars are at stake; and that card players occasionally cheat. At once your suspicions are aroused. (1989: 9-10)

> Any hand of thirteen cards is in an important sense exactly as unlikely as any other, but our suspicions are aroused when we watch Smith winning a million dollars with a hand of thirteen spades that Smith has dealt to Smith. We do not just say 'Lucky Smith!', disregarding the explanation that stares at us. (1989: 121)

Leslie explains these examples with what might be called his *tidiness principle*: "A chief reason for thinking that something stands in special need of explanation is that we

actually glimpse some tidy way in which it might be explained" (1989: 121). This principle is endorsed by Peter van Inwagen:

> Suppose that there is a certain fact that has no known explanation; suppose that one can think of a possible explanation of that fact, an explanation that (if only it were true) would be a very good explanation; then it is wrong to say that that event stands in no more need of explanation than an otherwise similar event for which no such explanation is available. (2002: 135)

What is going on here? For Leslie, it seems, there is a relationship between our ability to think of an explanation for a state of affairs and the need for that state of affairs to be explained. In other words, Leslie and van Inwagen might be taken to be asserting the following principle:

> *If I can think of a good explanation, then an explanation is needed.*

Is it rational to assert this principle? Again, I think there is something right about this principle and something wrong about it. Surely, my capacity to think of an explanation should not determine whether an explanation is needed or not. That would imply that someone who is less imaginative would *need* to explain less in their world, and someone who is more imaginative would *need* to explain more. That does not sound like it is tracking what *really* needs to be explained. But, on the other hand, perhaps there is something right about Leslie's principle. There does seem to be something very reasonable about choosing to explain the things that you are capable of explaining, rather than choosing to explain the things that you are not capable of explaining!

Now let's look at Leslie's justification for his principle. He claims that his *tidiness principle* is an aspect of the more general principle fundamental to all science formalised in Bayes' rule: "that observations improve your reasons for accepting some hypothesis when its truth would have made those observations more likely" (1989: 121). So, Leslie justifies this principle with reference to the probability calculus. But is this principle really just 'an aspect' of Bayes' rule? To consider this let's look at the basic features of the principle. Firstly, there is the ability to think of a good explanation. And secondly, if such an explanation can be thought of, then the thing being explained *needs* explanation in a way that it would not need if no good explanation can be thought of.

Whatever is going on here does not sound like simply 'an aspect' of the probability calculus. For a start, our ability to think of an explanation is completely external to the formalism of the probability calculus. And furthermore, our ability in this case to formulate good explanations seems to have some normative impact on what things in the world *need* explanation. I suspect that what is going on here has something to do with the difference between unbounded and bounded rational agents and the fact that classical rationality is built on the model of ideal (unbounded) rational agents. Perhaps ideal rational agents do not concern themselves with questions about what *needs* explanation because they are capable of explaining everything. They do not have limited cognitive capacities that need to be used frugally. But, of course, limited rational agents do need to concern themselves with such mundane questions as what to spend time

attending to, what to spend time thinking about, or what to spend time attempting to explain.

So, I don't think Leslie's *tidiness principle* neatly drops out of classical rationality. Nonetheless, I think there is something to it. And I think the central question lurking in the background is this: when is an explanation needed? That is a big question, one that I will not attempt to address in any detail here. So, perhaps the way forward is to consider the following question: is it rational to think an explanation is needed because one can be thought of? In order to address this question we need to ask a further question: what norms of rationality are being applied here? As I have said, Leslie points to rationality in the form of Bayesian confirmation theory, so presumably Leslie is using the norms of classical rationality. So let's start there: using the norms of classical rationality, is it rational to think an explanation is needed because one can be thought of? I doubt whether an affirmative answer is to be found within the norms of classical rationality and I do not think the answer is to be found in the probability calculus itself. So I suggest that Leslie's principle cannot be justified with reference to classical rationality. However, I think a pragmatic justification can be found with reference to ecological rationality. Thus, I suggest the following *explanation attribution heuristic*:

> *If I can think of a good explanation for an event, then I should not initially assume chance is a reasonable explanation for that event.*

And here I am assuming that chance would not be considered a 'good' explanation.

Explanation attribution

Thus far I have discussed three distinct cognitive processes: Barrett's *hyperactive agency detection device*, Dembski's *law of small probability*, and Leslie's *tidiness principle*. I have suggested that it is problematic to justify any of these from the perspective of classical rationality. However, if we take a different perspective, one that acknowledges the cognitive limitations within which humans operate, I think these three cognitive processes look more reasonable. I suggest this perspective is implicit within both ecological and bounded rationality and my central point is that these three cognitive processes look more reasonable when considered as pragmatic solutions for evolved agents operating with limited cognitive capacities in complex informational environments.

Furthermore, I suggest that there is a common cognitive system that underlies these three processes. I call it the *explanation attribution* system. Before I explain how I think it works I will briefly explain what I take to be a broadly similar system: *face recognition*. While I am not a neuroscientist, let me describe what I take to be the essence of the face recognition process. Once a human has experienced a face, some features of that face (perhaps particular ratios between certain points on the face) are remembered and linked in some way with knowledge of whom that face belongs to. As more faces are experienced the library of facial features, linked to knowledge of whom those features belong to, grows. Then, on future occasions, when a face is experienced, this library is searched and if the facial features in the library match the facial features currently being

experienced, then the face is recognised, and this recognition is then brought to the awareness of central cognition with the thought: "I recognise that person".

Here is how I think the *explanation attribution* system works. Central to the system is a library of pairings of representations of previously experienced states of affairs and explanations of those states of affairs. Here 'previously experienced' needs to be interpreted rather loosely, because it may include genetically hard-wired 'previous experience'. The library does not contain representations of all previously experienced states of affairs, but only representations that are linked to some relevant explanations, or perhaps a relevant explanatory type, e.g., 'agency' or 'design' explanations. When a new state of affairs is experienced the system compares it with the library of pairings. This library of pairings is ordered in some way, and the new state of affairs is compared with each pair in the library. At each comparison two outcomes are possible. Either the comparison process results in no match or it results in a match. If there is no match the next comparison is made with the next pair in the library. If there is a match, then the system links the new state of affairs with the explanation of the previously experienced state of affairs and the process stops. In other words, an explanation is attributed to the new state of affairs. This pairing of the new state of affairs and the attributed explanation is then made available as information to central cognition. The successful functioning of this process could be represented as follows:

Input:	Newly experienced state of affairs (NESA)
Process:	NESA matched with a previously explained state of affairs in the library
Output:	Explanation of previously experienced state of affairs attributed to NESA

I suggest that this basic pattern is common to the three processes we have been considering. So, let's consider the *hyperactive agency detection device*, Dembski's *law* and Leslie's *tidiness principle* to see if they fit this pattern.

Firstly, the *hyperactive agency detection device* can be represented as follows:

Input:	Rustling in the bushes (NESA)
Process:	NESA matched with previously explained state of affairs in the library
Output:	'Predator' explanation of previously experienced state of affairs attributed to the rustling

The output in this situation is the attribution of the explanation 'predator', and I suggest this explanation could have been identified by comparing the newly experienced state of affairs 'rustling in the bushes' with a library of representations of previously experienced states of affairs linked to 'predator' explanations.

Secondly, Dembski's *law of small probability* can be represented as follows:

Input:	Event of small probability (NESA)
Process:	NESA matched (specified) with a previously explained state of affairs in the library
Output:	Non-Chance Explanation is attributed to the NESA (and hence chance is eliminated)

This situation is a little more complex, because of the relationship between specification and explanation, but I think it has the same fundamental form as the other processes. The basic common pattern is a library of pairings of representations of previously experienced states of affairs and explanations of those states of affairs. But in this case we need to incorporate the notion of specification. As mentioned earlier, specifications might be generated using an explanatory schema, and so one way to think of a specification in this context is as an additional feature of the pairing of representations of previously experienced states of affairs and explanations of those states of affairs. So here I am suggesting that there is a library of representations/specifications of previously experienced states of affairs each of which is paired with an explanation. If there is a match between newly experienced states of affairs (the event of small probability) and a pairing in the library, then the specification and the corresponding explanation of that pairing are attributed to the newly experienced state of affairs. Due to the way that Dembski has set up his *explanatory filter*, the *law of small probability* only applies in situations that can be explained either by chance or by design (1998: 36-66). So if chance is eliminated, design is inferred. Thus, the output 'chance is eliminated' means, for Dembski, that design is inferred. This can be interpreted as a 'design' explanation being attributed to the event.

And finally, Leslie's *tidiness principle* can be represented as follows:

Input:	Newly experienced state of affairs (NESA)
Process:	NESA matched with a previously (well) explained state of affairs
Output:	An explanation is needed because the NESA has been matched with a good explanation

Here the output, that an explanation is needed, is produced because the newly experienced state of affairs is matched with a previously experienced state of affairs that has been explained. Thus, an explanation (or explanatory type) is attributed to the new data. And given that an explanation has been attributed to the new data the conclusion is drawn that an explanation is needed. (Here is it assumed that 'chance' is not a good explanation.)

I think the individual form of these three processes is sufficiently similar to the general pattern of *explanation attribution* to suggest that they are all instances of the operation of this cognitive process. To continue, I now speculate on the mechanism of this process.

Is *explanation attribution* modular?

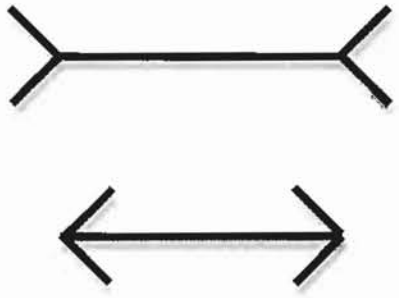
Perhaps all this *explanation attribution* is completely conscious. Perhaps it is operating in the way that the recognition heuristic is operating when we choose the German city that we recognise as being the larger of the two cities. Perhaps a fully conscious process is what Leslie is trying to characterise with his *tidiness principle*. But, while acknowledging this possibility, I think that at least some of this process is working at a 'deeper' level, and this leads me to believe it is modular.

Consider the parts of Leslie's process that do not seem to reduce to a simple application of the probability calculus. Talk of what situations 'need' explanation

implies, to me at least, that there are intuitions at work here as well as conscious processes. Note the language Leslie uses: "suspicions are aroused" and "the explanation that stares at us". This sounds like intuition at work. Another way to think about this is to ask for the 'reason' that an explanation is 'needed'. Of course, Leslie claims it is because we can think of an explanation. But, surely, that is not a reason that is implied by classical rationality. If classical rationality implied anything it would be that everything needed explanation. So, I suggest that a fully conscious classically rational process will not capture the process of explanation attribution.

Consider, then, Dembski's *law of small probability*. Sober has correctly pointed out that it is not (classically) rational to unconditionally reject a hypotheses simply because a certain event would be improbable given the truth of that hypothesis. So we cannot justify rejecting chance using classical rationality alone. But, on the other hand, there does seem to be something intuitively right about rejecting chance in the coin toss example. And, again, I think more than just conscious (classically rational) deliberation is at work here. So, while acknowledging that there could be a completely conscious story to tell here, I think that there are good reasons for considering a modular scenario. I now therefore make some tentative suggestions in that direction.

Fodor (1983) introduced the idea of a mental module. He suggested the existence of peripheral modules that work with incoming sensory data and output interpretations of that data to central cognition. The Müller-Lyer illusion is a good example of a Fodorian module at work.



The Müller-Lyer illusion

Although the lines are the same length, they appear to us to be different lengths. This appearance is said to be the output of a module. The visual system has provided an input to the module that presumably involves information about two lines of equal length, but the module manipulates this information and outputs information representing two lines of unequal length. Fodorian modules, such as that which is said to produce the Müller-Lyer illusion, act on the periphery of the mind and simply give central cognition an interpretation of sensory data that, among other things, is not revisable. Even when we measure the lines and consciously determine they are the same length, we cannot consciously override the output of the module and *see* the lines as the same length. While this is an interesting starting point, I do not see the explanation attribution process as the functioning of a Fodorian module. However, I do see something of the 'non-revisable' nature of the output of a Fodorian module in the output of the explanation attribution

module.[4] Consider the coin toss example. Even when I consciously tell myself it is not reasonable to straightforwardly reject the chance hypothesis, another part of me is saying "Come on, this just can't be chance!", and it is this intuition that I think might be the output of a module.

If the explanation attribution process is not the functioning of a Fodorian module, what type of module might it be? I will consider two types of module: computational modules and Chomskian modules. Samuels *et al.* characterise a Chomskian module as "a domain specific body of mentally represented knowledge or information that accounts for a cognitive capacity" (1999: section 4.1.1). In other words, Chomskian modules are databases of domain specific information that can be accessed either by domain general cognition or by computational modules. In contrast to Chomskian modules, which do not process information, computational modules are processing devices: they manipulate information (1999: section 4.1.2). A computational module may manipulate information that is sourced from within its own library, or sourced from a linked Chomskian module.

I will start with Chomskian modules, given that we began considering explanation attribution as a fully conscious heuristic and the functioning of these modules is closer to a fully conscious heuristic than a computational module. If a Chomskian module is the correct way to characterise the explanation attribution module then it might work something like this. There is a library of representations of previously experience states of affairs (perhaps with explanations attached) in the module. Central cognition can compare the newly experienced state of affairs with the module's library. If central cognition can find a match between the newly experienced state of affairs and a previously experienced state of affairs, then it attributes the explanation (that is paired with the previously experienced state of affairs from the library) to the newly experienced state of affairs.

Now let's consider explanation attribution as a process involving a Chomskian module with respect to the *hyperactive agency detection device*, Dembski's *law of small probability* and Leslie's *tidiness principle*. Whether the module is Chomskian seems to be related to two things in particular: firstly, how much of the process of attribution is consciously deliberated, and secondly, how fast the process is. I imagine that the more the process is consciously deliberated and the slower it is, the more likely the module is Chomskian, while the less conscious deliberation that is involved and the faster the process is, the more likely the module is computational.

Here, I think it is fair to say, there is some variation among the three processes under consideration. Leslie's *tidiness principle* seems to involve the most conscious deliberation. The way Leslie describes the process suggests conscious deliberation, particularly with his reference to Bayes' rule. But I wonder how much of this description is Leslie's attempt to (classically) rationalise the process. His use of phrases like "suspicions are aroused" and an "explanation that stares at us" suggests that more than conscious calculation is at work here. So, while acknowledging that the *tidiness principle* is the most cognitively accessible, I still think intuition is a work here. Finally, Leslie's *tidiness principle* seems to be the slowest of the three processes. So, of the three, it is the one most likely to be related to a Chomskian module.

---

[4] The non-revisability of the output of a Fodorian module relates to what Fodor discusses as the mandatory nature of modules and the fact that there is only limited central access to the workings of modules (1983: 52-60).

Consider also Dembski's *law of small probability*. Again, the language Dembski uses to characterise the process suggests fully conscious deliberation. But, again, I wonder how much of this language is Dembski's attempt to rationalise the process. In fact, I think Dembski's *explanatory filter* project, as a whole, is an attempt to rationalise what may be a largely intuitive process. When I consider my imagined reaction to watching the coin land in front of me following exactly the series I have specified, I do not imagine myself being quietly reminded of Sober's point. Rather, I imagine myself experiencing rising amazement that something that simply *could not* happen is happening in front of me. So, again, while the intellectual presentation by Demski of his *law* (and *explanatory filter*) is couched in rational terms, I think what he is trying to explain is actually based on much less (classical) rationality than he would have us believe. Finally, I feel that the rejection of chance in the case of the *law of small probability* is a quicker process than the identification of the 'need' to explain a state of affairs, à la Leslie. So, while there could be a Chomskian module at work here, I think it looks more like a computational module.

Consider finally the *hyperactive agency detection device*. Arguably this process involves much less conscious deliberation when compared to the other two processes, and it is also much faster. And given its claimed evolutionary origins, this comes as no surprise. The *hyperactive agency detection device* needs to act very quickly and with the minimum of conscious deliberation. Of the three processes this one is most likely a computational module.

So if any of these processes is a Chomskian module the best candidate is the *tidiness principle*. Here central cognition would draw upon data within the Chomskian module to come to the conscious conclusion that an explanation is needed in the situation under consideration. A Chomskian module could also be the best way to characterise the rejection of chance in cases of specified events of small probability, but I think this is less convincing. The *law of small probability* has a more definitive feel about it than the *tidiness principle*. This suggests that it is not conscious deliberation that rejects chance here, but rather the rejection of chance is presented to central cognition.

The *presentation* of a position to central cognition, rather than the *deliberation* of positions by central cognition, brings us to computational modules. While Chomskian modules are databases of domain specific information available for use by conscious processes, computational modules are more self-contained. Computational modules process information and generate outputs, in ways that Chomskian modules do not. And I think a computational modular story is the most interesting to consider with respect to these three processes.

If these three processes are run by computational modules then the whole of each process is occurring within the modules. A representation of the new state of affairs is compared to the library of previously experienced (and explained) states of affairs. If a match occurs then the explanation (associated with the previously experience state of affairs) is attributed to the new state of affairs. Then the module 'outputs' the new state of affairs *together with* the attributed explanation to central cognition. So, perhaps, the first time central cognition is aware of the new state of affairs it comes with an attributed explanation already attached.

Let's consider how this computational module scenario compares to the three processes. I think it fits very nicely with the *hyperactive agency detection device*. The

awareness of the moving dark area in the forest comes with the idea of a predator already attached. And, although it might be more of a stretch, I also think that it is a reasonable way of characterising the *law of small probability* and the *tidiness principle*. For it is the definitiveness of the *law of small probability* that suggests that a computational module is at work. You just 'know' that chance cannot explain the coin toss, and this knowledge comes to you as you watch the process. It does not come to you after the fact. Furthermore, it is an intuitive response, and I think this is evidence that some processing has occurred before you come to consciously think about the situation. Finally, the "suspicion" that Leslie acknowledges and the fact that "explanations stare at us" also seem to be evidence that something is going on *before* we start consciously comparing probabilities in the case of the *tidiness principle*. All this suggests that a computational module is at work comparing representations of new states of affairs with representations of previously experienced states of affairs. When there is a match, the module then presents to central cognition the new state of affairs together with an explanation that the module has attributed to it.

I also think there is another reason that makes the computational module scenario more compelling than the Chomskian module scenario. This relates to a pragmatic concern. We simply cannot consciously attend to all the information that is being received by our cognitive system. If we did attempt to attend to all information, our central cognitive processes would collapse. So we need some mechanism to select certain parts of the incoming stream of information and bring those parts to our conscious attention. We need some way of selecting what to devote our limited central cognitive capacities to. A computational explanation attribution module seems like a good way of making those selections. There are very good evolutionary reasons for getting it right when it comes to selecting where to invest our cognitive resources. In the case of predator detection, we want a quick way of deciding what parts of our environment might be predators. Domain general processes are probably not the best way to survey large amounts of incoming sensory information. Perhaps it is the job for a computational module. If there were a system that compared all incoming sensory information with information that corresponded to previous predator scenarios, and when it found a match presented that information to central cognition, then this might increase the fitness of the organism that possessed such a system.

To recap, I postulate the existence of an explanation attribution process. Furthermore, I suggest that the *hyperactive agency detection device*, the *law of small probability* and the *tidiness principle* can all be understood as instances of the functioning of the explanation attribution process. This process might function entirely consciously in the way that other fully conscious heuristics, such as the recognition heuristic, function. However, it might function fully within, or using, a module. That module might be a Chomskian module or a computational module.

The modular version of the explanation attribution process is particularly attractive because it goes some way towards explaining the intuitive dimension of all these processes. I suggest that there are intuitions at work in all three processes, and attempts to (classically) rationalise these intuitions are not successful. If we attempt to give an all things considered justification of these intuitions we fail. However, rather than attempt to explain these intuitions with classical rationality, I think there is a better option. These intuitions are not the product of some classically rational process. Rather,

the intuitions are the output of a fast and frugal modular process. If a new state of affairs matches a previously experienced state of affairs, then the new state of affairs is 'intuitively' explained by attributing the same explanation as was used to explain the previously experienced state of affairs. If we take this approach, we have avoided trying to identify a classically rational justification where none exists. Furthermore, if we embrace the notion of ecological rationality, we can find a justification for the process.

Is the 'detection' of supernatural design the output of a mental module?

In conclusion, I will draw the discussion back to where it began. I started by describing how the possible existence of the *hyperactive agency detection device* has been used to explain humanity's predisposition to believe in supernatural agency. Having described a possible mechanism for the device, in the form of the explanation attribution module, we can say more about how this all fits in with belief in supernatural agency.

Imagine that a computational module has linked a design explanation to some new and ambiguous state of affairs. The module then makes this attributed explanation salient to central cognition. But what happens next? Implicit in my story thus far was the assumption that the output of the module was only the beginning of the story. After the module provides an output, presumably central cognition reviews the output and then makes an all things considered judgement about the attributed explanation. After such deliberation, central cognition presumably confirms the explanation or chooses an alternate explanation. But while that is the implicit story, there is another possibility. Perhaps the module presents the attributed explanation to central cognition and central cognition reviews it, but neither confirms nor replaces the attributed explanation. Or perhaps central cognition does not have time to review the attributed explanation, or it gets distracted and never gets back to it. What happens then? All we have is the attributed explanation as the output of the module. Imagine that a module has attributed an explanation to a state of affairs and that is the end of the story. What would this be like? Perhaps it would be like recognising a face, but then not being able to place that face. You would be left with the knowledge that you know that person, but you just do not know where you know them from.

Imagine that we do possess an explanation attribution module that can attribute 'agency' or 'design' explanations to ambiguous information. And furthermore, imagine that central cognition does not always do anything to overrule such attributed explanations. To me this sounds very much like the 'intuition' that there really is some design or agency explanation lurking just out of reach of our conscious deliberative processes, the intuition that suggests that there are supernatural agents out there. Thus, 'design' or 'agency' explanations of ambiguous information, generated by a computational module, could help explain our propensity for such beliefs.[5]

[5] I thank James Chase for a number of discussions that helped clarify my thoughts and for pointing me in the right direction within the evolutionary psychology literature. Also, I thank Graham Oppy and Peter Forrest for comments on the version of this paper presented at the APRA Conference in 2008, and an anonymous referee for comments on the penultimate draft of this paper.

References

Atran, S. and A. Norenzayan, A. (2004). "Religion's Evolutionary Landscape: Counterintuition, Commitment, Compassion, Communion." *Behavioural and Brain Sciences* 27: 713-70.

Barrett, J. (2000). "Exploring the Natural Foundations of Religion." *Trends in Cognitive Science* 4 (1): 29-34.

Dembski, W.A. (1998). *The Design Inference: Eliminating Chance Through Small Probabilities*. Cambridge: Cambridge University Press.

Fitelson, B., Stephens, C., and Sober, E. (1999). "How Not To Detect Design – Critical Notice: William A. Dembski, *The Design Inference*." *Philosophy of Science* 66 (3): 472-88.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Gigerenzer, G. (2000). *Adaptive Thinking: Rationality and the Real World*. Oxford: Oxford University Press.

Gigerenzer, G., Todd, P.M., and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. Oxford: Oxford University Press.

Kahneman, D., Slovic, P. and Tversky, A. (eds) (1982). *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Leslie, J. (1989). *Universes*. New York: Routledge.

Samuels, R., Stich, S., and Tremoulet, P.D. (1999). "Rethinking Rationality: From Bleak Implications to Darwinian Modules." http://ruccs.rutgers.edu/ArchiveFolder/Research%20Group/Publications/Rethink/rethink.html Accessed 13/06/08.

Simon, H.A. (1957). *Models of Man*. New York: Wiley.

Sober, E. (2003). "The Design Argument." In N.A. Manson (ed.), *God and Design: The Teleological Argument and Modern Science*. London: Routledge, 27-54.

Van Inwagen. P. (2002). *Metaphysics*. Boulder: Westview Press.