Skepticism and Evolution[[1]](#endnote-1)\*

N.A. Pinillos

Arizona State University

ABSTRACT: I develop a cognitive account of how humans make skeptical judgments (of the form “X does not know p”). In my view, these judgments are produced by a special purpose meta-cognitive "skeptical" mechanism which monitors our reasoning for hasty or overly risky assumptions. I argue that this mechanism is modular and shaped by natural selection. The explanation for why the mechanism is adaptive essentially relies on an internalized principle connecting knowledge and action, a principle central to pragmatic encroachment theories. I end the paper by sketching how we can use the account I develop here to respond to the skeptic.

**1. Introduction**

In the Meditations, Descartes raises the possibility of an evil demon massively deceiving us about the source of our perceptions.[[2]](#endnote-2) As we go through the text, dread (perhaps mixed with intellectual exhilaration) sets in as we realize we may not know many ordinary and common-sense propositions. This is in part why the Meditations makes for such a gripping piece of philosophy.

A lot of epistemology since Descartes has been concerned with showing why this pull towards skepticism is mistaken. Less attention has been paid to the question of why we should feel this pull in the first place. In fact, it doesn’t seem at all *useful* to be drawn to skeptical judgments in the face of bizarre evil demon scenarios. It may even be a pernicious disposition insofar as it may cause us to lose confidence about propositions we should be more confident about. And if the disposition to fold (or to nearly fold) under skeptical pressure appears to serve no useful purpose, perhaps there is no deep reason why we should possess it. Perhaps our dispositions to skepticism are just an accidental feature of cognition or, as I will argue, a side effect of some other important cognitive capacity. From the design perspective (on this line of thought), if we were to build an artificial intelligence device with ideal cognitive capacities, nothing important would be left out if we designed it to not be tempted by skepticism.[[3]](#endnote-3)

A number of researchers have sought to provide psychological explanations for the draw of skeptical judgments.[[4]](#endnote-4) Space limitations prevent me from discussing these theories in detail. The perspective I advance is different enough from these accounts in that it focuses on the evolutionary function of denials of knowledge and how these interact with internalized knowledge-action principles of the type defended recently by “pragmatic encroachers” in epistemology.[[5]](#endnote-5)

The approach I explore here is that many skeptical judgements are the output of a special purpose meta-cognitive mechanism.[[6]](#endnote-6) I will argue that this mechanism is an adaptation. Adaptations solve problems for organisms. For example, the adaptation of having legs solves the problem of locomotion and the adaptation of having teeth solves the problem of grinding food for digestion. I submit that the mechanism which leads to skeptical judgments is an adaptation which solves the problem of figuring out what action to take. The mechanism does this by monitoring our reasoning on command and detecting premises which are too hasty and should not be assumed. The system flags them as “not knowledge”.

This “skeptical mechanism” produced judgements in cases that mattered for our ancestors. These cases involved useful everyday denials of knowledge. In sharp contrast, judgments about evil demon and brains and vats are esoteric philosophical concoctions. Although they are genuine deployments of the mechanism, they are outliers and fall far outside the conditions which originally gave rise to the mechanism. An interesting further question is whether we can learn anything about the epistemic status of mental states that are the outputs of adaptive mechanisms by learning about the original design conditions for those mechanisms. I think we can. I will argue that the mechanism was not designed for esoteric philosophical scenarios and as a result, we are not justified in believing what it tells us about those scenarios. If these outputs are what give the skeptical arguments their intuitive appeal, then we may very well be able to respond to the skeptic. In particular, we will be able to say that the following premise in the skeptical argument is unjustified: I do not know I am not a brain in a vat. I end the paper with a brief discussion on this issue.

2. **Knowledge-Action Principles**

I said above that a certain propensity to deny that people know things in skeptical pressure cases helps us figure out how to act. But how can that be? What does knowledge have to do with action? In a project pioneered by Fantl and McGrath (2002), a number of researchers have posited principles explicitly connecting knowledge and appropriateness of action. Hawthorne and Stanley (2008) propose the following:

(KA) Where one’s choice is p-dependent, it is appropriate to treat the proposition that p as a reason for acting iff you know that p[[7]](#endnote-7)

One important motivation for this principle is that it helps explain why ordinary people use ‘knows’ to criticize behavior. For example, one can criticize a job candidate for not wearing formal attire: ‘You knew you would be meeting the dean, why didn’t you wear formal attire?’ or criticize someone for not bringing a hat: ‘you knew it would be sunny, why didn’t you bring a hat?’. These are positive attributions which make sense of the “right-to-left” direction.

The “left-to-right” direction is also supported by folk appraisals. For example, an agent can be criticized by being told that they did not know some relevant proposition. For example, I could reasonably say ‘Why did you throw away the pizza, you didn’t know I wouldn’t be back for dinner again’, or ‘You shouldn’t have bought that expensive bicycle, you don’t know that we won’t need the money for something else this month’. Thus, both directions of (KA) are supported by folk behavior. [[8]](#endnote-8)

The idea that the principle can explain folk behavior suggests that ordinary people accept the principle (or something like it) or are sensitive to it. That is, the principle is not *just* some deep metaphysical fact, an ideal of rationality hidden in Platos’ heaven. Rather it, is (in some sense) endorsed by ordinary people.

Despite the plausibility of the principle, it might be too strong. Consider the following potential counter-example (to right-to-left) due to Jessica Brown (2008). A surgeon about to remove a patient’s kidney double checks to see if it really is the left kidney that needs removing. Intuitively, the doctor must already know before the surgery that it is the left kidney that needs removal. Yet, she should still double check. That is, it is not acceptable for her to use the known proposition that the left kidney needs removal in her practical reasoning (without double checking). Hence (KA) is incorrect. Jessica Brown also objects to left-to-right direction of (KA). “Gettiered” subjects, she argues, could blamelessly use *p* in reasoning even though they fail to know *p*.

Maybe you don’t find these counter-examples convincing. It doesn’t matter. The point I would like to make is that we don’t need a principle that is as strong as (KA) to establish an important connection between knowledge and the normativity of action. A weakened principle that holds ceteris paribus (or in “normal cases”) will do.[[9]](#endnote-9) In fact, the principles of folk psychology, which aim to explain and predict rational human behavior are of this weakened form. It would not be at all surprising if the correct knowledge-action principle also had this structure.[[10]](#endnote-10)

**3. The Skeptical Mechanism**

I said above that the mechanism which normally produces skeptical judgements is an adaptation. Before I argue for this, I have to say how this mechanism is supposed to function. The workings of this mechanism are not transparent to introspection. It is the sort of process we can understand by inspecting people’s judgments across a range of hypothetical cases and noting the patterns that emerge. This can be done informally by theorists surveying their own judgments about cases (as traditional epistemologists do) or more formally using controlled experiments. Not surprisingly, it turns out that work from both traditional and experimental epistemology can give us insight into the process.

What we are interested in here is a cognitive model of the skeptical process. Typically, cognitive models are algorithmic descriptions of information processing mechanisms.[[11]](#endnote-11) Minimally, to construct such a model we need to specify the inputs, outputs and the computation which turns inputs to outputs. In specifying such a model, there is no assumption made that there is a dedicated component of the mind in charge of carrying out such a computation. That is a further claim, a claim that I will in fact be making about skeptical judgements, but further evidence would need to be brought in.

I will be describing what is surely an oversimplified model of the skeptical process. Suppose we are observing Jon Doe is at a restaurant. Jon Doe, who has properly functioning eyes, makes the true judgment that the couch in front of him is crimson red. There is no funny business going on here and so observers would be happy to say that Jon Doe knows the couch in front of him is crimson red. But now consider a “skeptical pressure case” where observers of Jon Doe are reminded that sometimes restaurants have ambiance lighting, so that a non-crimson red couch may turn out to look crimson red. When we are presented with such a possibility, we are more likely to deny that Jon Doe knows the couch in front of him is crimson red.

This example suggests a simple model with two inputs. (Input 1): Doubt about *p*--which involves the representation of some specific possibility (or class of possibilities) where *p* is false. (Input 2): The epistemic position of an agent *S* with respect to some proposition *p* (evidence or justification *S* possesses, how the agent formed the belief that *p* etc.).

Here’s a description of the computation. The agent determines whether S’s epistemic position (Input 2) can *rule out* the specific possibility that *p* could be false (Input 1). If it is determined that it cannot rule it out, then the process returns ‘*S* does not know *p*’ as the output. If it is determined that it does rule it out, then the system is disengaged. For example, in the couch case, input 2 is Jon Doe’s epistemic position (Jon Doe’s eyes are working properly, he is standing in front of the couch etc.). The possibility of error (input 1) is that the restaurant has weird lighting--which is determined to not be ruled out by the agent’s epistemic position. The computation then returns that Jon Doe fails to know the couch is red. The reason why it’s determined to not be ruled out is that it is perceived that Jon Doe has not checked for ambiance lighting. If it was noticed that he had checked for tricky lighting, then Jon Doe’s epistemic positon would be seen to rule out the possibility of error (and the system would disengage).

Is this simple cognitive model realistic? Something like it is assumed in contextualist treatments of knowledge (Cohen 1998, Lewis 1996).[[12]](#endnote-12) And experimental work backs this up. For example, Nagel et al (2013 ) gave participants “skeptical pressure” cases like the sofa scenario above and found that participants who read those scenarios were more likely to deny knowledge than participants in a control condition where they were not reminded that lighting could be tricky. Similar results were verified by Powell et al (2015), and Schaffer and Knobe (2012) deploying distinct experimental paradigms.

In the next sections I put some flesh into this model. Space constraints limit detailed discussions on these important points.

**3.1 Inputs**

(a) Skeptical judgments do not just involve fantastical evil demon type situations. They also involve “local” skeptical scenarios. For example, the “crimson red couch” scenario leads to a denial of knowledge regarding the color of the couch, but not to a denial of all perceptual knowledge (even accepting closure). This is an important point. Cognitively, local skeptical scenarios will be treated in the same way as global skeptical cases. Some of these local skeptical scenarios are known in the literature as “lottery” cases (Kyburg 1961). For example, agents claim to not know they will lose the lottery (even though the probability of losing is extremely high and even though they are willing to say they know a lot of other things which have a higher probability of being false—like knowing the time based on a wrist watch). Another example involves agents claiming to not know that the president hasn’t been assassinated a few seconds ago (supposed they haven’t checked their smart phones). Even though the chances of this occurring are very low, agents tend to judge they fail to know this. The approach I take here is that, cognitively, these cases trade on the possibility of error being made salient plus the agent’s epistemic perceived position not being able to rule out this possibility of error (at least not perceived as being ruled out---see section 3.3). As we will see, all of this leads to denials of knowledge.

Note that the reason why these local skeptical cases are philosophically puzzling is that if we really don’t know these things, then (by epistemic closure[[13]](#endnote-13)) it would seem like we fail to know more mundane things we ordinarily take ourselves to know. For example, if I don’t know that I will lose the lottery this week, then I also don’t know the following: that I will be going grocery shopping this weekend (suppose that I know that if I won the lottery, I would eat out at a restaurant for every meal). But presumably I know that I will be heading to the grocery store this weekend (if someone asks me if I know what I will be doing this weekend, I will mention the trip to the grocery store). Note that I am not saying that epistemically, lottery and global skeptical cases are alike. I am conjecturing that judgments from those domains arise from similar cognitive mechanisms.

The way that skeptical and lottery type cases are discussed in the philosophical literature, it may be thought that they are contrived cases concocted to make a philosophical point with no application to the “real world”. But the cognitive mechanisms that explain these cases are, in my view, also responsible for interesting cases in law.

According to the Wells effect, jurors are less likely to convict on probabilistic evidence that is purely “statistical” as opposed to evidence that seems causally connected to the particular case at hand. In a famous experiment, Wells (1992) described to a mock jury a case where it was alleged that a particular bus company, the Blue Bus Company, harmed a dog. One set of jurors were told that the evidence against the Blue Bus company was that 80% of the buses in town were run by the Blue Bus company (“statistical evidence”) and hence there is an 80% chance the Blue Bus company harmed the dog. Another set of participants were presented with an eye witness who claimed to see a Blue Bus bus harm the dog. The jurors were told that the witness was 80% reliable. In both cases, jurors agree that the chance the Blue Bus company harmed the dog was 80% but nonetheless they were more willing to convict on the eyewitness testimony than on the statistical evidence. In recent work, Friedman and Turri (2014) provided evidence that effect persisted for knowledge attributions. Applied to the Blue Bus case, participants would be less likely to say they know that the blue bus company harmed the dog when the statistical evidence was presented as opposed to the eye witness case. This is so even though participants would readily admit that in both cases there was an 80% chance of guilt. Interestingly, Turri et al (2017) gave evidence that judgments about whether the accused was blameworthy were mediated by their attributions of knowledge, indicating the cognitive centrality of knowledge judgements.

The idea that global, local skeptical, lottery, and Wells judgements have a common cognitive etiology is a substantive claim. I will say something in its favor below but note that if we are on the right track, then skeptical judgments may be of interest not only to epistemologists and cognitive scientists but also to legal scholars and advocates of justice.

(b) Some research indicates that along with salience of error, the practical import of the question (or ‘what is at stake’) could also lead to denials of knowledge.[[14]](#endnote-14) But this research is controversial.[[15]](#endnote-15) There is also the question of how practical interests should be incorporated into our model, if they play a role at all. For example, Buckwalter and Schaffer (2015) propose that the practical interest of an issue has an indirect effect on skeptical judgements. They argue that practical interests affect the salience of a possibility of error which in turn leads to knowledge denials. On another model they attribute to Pinillos, stakes play a direct role in knowledge attributions. For the purposes of this paper, I will leave the issue of practical interests aside.

(c) Not every case where error is raised to salience leads to skeptical denials (even when the hypothesis is not ruled out by the evidence). Consider the following example inspired by John Hawthorne (2004, p 64): Some friends are watching the movie ‘The Matrix’ (which presumably makes the possibility of massive error salient) and one asks the other whether he *knows* if they put real butter in the popcorn they are sharing. The interlocutor would naturally respond in a way that does not reflect a skeptical perspective. This seems to be a counter-example to the simple picture where salience of error (which cannot be ruled out by evidence) leads to denials of knowledge. We need some way to distinguish this type of case from the scenarios where skeptical judgements do arise.

(d) I said that “doubt” (input 1) would be represented as a specific possibility of error. But these are different things. A doubt that perhaps the lights above are tricky is not quite a specific representation of a possibility. A more sophisticated model would need to explain how agents get from doubt to the representation of a possibility. I also said that input 2 specified an agent’s epistemic status, but left it wide open how this is represented by the agent. Clearly, more details are needed.

**3.2 Outputs**

I said above that the skeptical mechanism produces skeptical judgments which are just denials of knowledge. But there is reason to think that the output of the mechanism yields something weaker, like *an inclination* to judge that one fails to know---as opposed to full blown judgments or beliefs. Imagine a philosopher who is a convinced non-skeptic. Convinced by anti-skeptical arguments, they sincerely judge that they know they are not deceived by an evil demon or that they know they will lose the lottery. Nonetheless, they still find themselves feeling the pull to make the skeptical judgement. This is why philosophers often feel the need to explain the psychological pull for skepticism even after all the arguments against skepticism have been given. The idea that the skeptical pull remains after beliefs are otherwise formed suggests that the skeptical mechanism produces an output that is short of belief. In Fodor’s terminology, the output is “shallow” and stops short of a full judgment or belief.[[16]](#endnote-16)

Another piece of evidence that the skeptical pull does not become full belief is that it is forgotten soon after the skeptical episode ends. This general phenomenon was noted by Hume when he wrote “Tis happy, therefore, that nature breaks the force of all skeptical arguments in time, and keeps them from having any considerable influence on the understanding.”[[17]](#endnote-17) Here, Hume highlights the impermanence of skeptical judgments or inclinations. Though Hume’s remarks suggest not only that skeptical inclinations fall short of being full blown beliefs, but that they are all but forgotten “in time”.

**3.3. Algorithm**

I said that when the mechanism determines that X’s epistemic position with respect to *p* (input 2) cannot rule out the possibility of error that *p* (input 1) the mechanism will return an inclination to judge that X fails to know *p*. There is a difference between saying that a system *determines* that X’s epistemic position rules out a possibility of error and saying that in fact X’s epistemic position rules out such a possibility. Perhaps one’s epistemic position in fact rules out the possibility of an evil demon deceiving us, but it doesn’t follow our system will say that it does. In other words, supposing the mechanism represents part of our folk epistemology, it may be that this part of our folk epistemology succumbs to skepticism even though *we* ought not to succumb to it.

Figure 9.1: Simplified Skeptical Mechanism (SM)

Does EP rule out error?

[Is belief-*p* sensitive?]

Possibility that ~p (error)

If Yes: Exit

If No: “S fails to know p”

S’s epistemic position EP

A well-known philosophical approach to deal with skeptical cases appeals to the notion of sensitive belief (a type of “truth-tracking”). An agent’s belief that *p* is sensitive just in case if *p* were false, the agent would not believe *p*. Robert Nozick (1981) proposed that a necessary condition for knowing *p* is that the belief that *p* be sensitive.[[18]](#endnote-18) Take an ordinary case of knowledge, like knowing that you are reading a paper right now. This belief is sensitive since if you weren’t reading a paper right now then you would not believe that you were. For example, if you weren’t reading a paper right now, maybe you would be making coffee and so you would believe you are making coffee right now (not reading a paper). Hence, sensitivity gets it right in at least some of these ordinary cases of knowledge.[[19]](#endnote-19)

Now consider a global skeptical case. Suppose you believe that you are not a brain in a vat (BIV). This belief is not sensitive: if you were a BIV, you would still believe that you were not a BIV. Accordingly, you don’t *know* that you are not a BIV. Similarly, for the lottery cases. Your belief today that you will lose the lottery tomorrow (suppose you own a ticket and have formed the belief that you will lose tomorrow) is not sensitive. If you were to win tomorrow, you would still believe today that you will lose the lottery tomorrow. Enoch et al (2012) have argued that statistical “wells effect” cases can also be explained by sensitivity. Suppose you believe on purely statistical evidence that a Blue Bus bus harmed the dog, then presumably you would still have this belief if it turned out a different bus company harmed the dog. Hence the statistical belief is not sensitive. What about the eyewitness case? Suppose you believe that the blue bus company was responsible on the basis of the eye-witness testimony. Now if the blue bus company weren’t responsible, you wouldn’t believe this anymore. Presumably, the eye-witness would not have testified. At the very least, it less intuitive that your belief would remain the same in the second case.

Many philosophers doubt that sensitivity is a necessary condition for knowledge. The classic problem with that approach is that (given some plausible assumptions) it requires denying closure for knowledge. As we saw, the account predicts we don’t know we are brains in vats (since the corresponding belief is insensitive), but unless we go for global skepticism, we should accept that that we know we have hands (which corresponds to a sensitive belief). But the problem is that by closure if I know that I have hands, then I should be able to deduce and come to know that I am not a brain in a vat.

My approach does not say that sensitivity is a necessary condition for knowledge. Rather, I am making the psychological claim that the skeptical judgments are produced by testing for sensitivity. I am not saying anything about what knowledge itself requires.

As a first approximation, what it means for X’s epistemic position with respect to *p* to rule out the possibility of error E is that the following is true: Given X’s epistemic position, if E were the case, X would not believe P. For example, let us suppose the possibility of error *E\** at issue concerns a deceptive evil demon. Given my epistemic position, if E\* were the case, I would continue to believe that I live in a demon-free world. Hence, my belief that I live in a demon-free world is not sensitive, and so I have the inclination to judge that I fail to know that I live in a demon-free world.

I say that this is an “approximation” because the algorithm would likely involve testing a sample of several nearby-counterfactuals as opposed to relying on a single make or break counterfactual. But for the purposes of this paper, we will assume a single counter-factual test.

In what follows I will continue to talk about doubt which is not ruled out by an agent’s epistemic position or “unresolved doubt”. I will leave this at an intuitive level. Readers who find this notion too imprecise may replace it by the notion of insensitivity as described in the previous paragraphs.[[20]](#endnote-20)

**4. The skeptical mechanism as an adaptation**

We are now ready to consider the hypothesis that the skeptical mechanism (SM) is an adaptation. To say that the SM is an adaptation is to say that it is an adaptation for some function or purpose. I mean to be using teleological terms ‘function’, ‘for’, ‘purpose’ and the like in the sense familiar from the theory of natural selection. This is the sense in which evolutionary biologists say that wings are an adaptation for flying or that hearts are an adaptation for pumping blood.

The notion of an adaptation is a historical-explanatory notion (Williams 1966). Elliot Sober provides the following analysis (1984 208): ‘A trait T is an adaptation for X in a lineage just in case T evolved in the lineage because there was selection for T and there was selection for T because having T promoted doing X.’ As such, my claim that ‘SM is an adaptation for X’ or ‘the function of SM is for X’ is really a claim about the historical-explanatory role of SM.[[21]](#endnote-21)

There are other uses of teleological terms not essentially tied with natural selection. For example, the physiologists sense of ‘function’ concerns the role an organ or capacity plays in a complex organism. This is a synchronic ahistorical notion. The synchronic notion often coincides with the historical one. But they can come apart. The appendix presumably has no synchronic function but it has a selection function, for example. To clarify, I will be using “function” and other teleological terms in the historical sense essentially tied to natural selection not the synchronic one.

It is of course, difficult to find direct evidence that SM is an adaptation for X. We can’t visually observe the SM or its components, or rely on fossil records. Our evidence will be indirect and highly speculative. As a first step, we need to identify the ‘X’ and give some reason to think it promoted fitness. And here we have our work cut out for us. In many cases, it is easy to at least come up with a plausible story of how a trait helped us survive and reproduce. It doesn’t take much imagination to note how having hearts or wings aids fitness. But in the case of SM, it is not at all obvious how it could help us survive or reproduce (Recall that we started the paper noting how skeptical judgements seem like a useless feature of our epistemic capacities). So what is the use of SM?

**4.1 The Adaptive Function of SM**

Let’s begin by thinking about a real-world version of the crimson red couch example. Suppose we are walking around the restaurant. A fellow customer says to us: ‘There is a chance that the lighting here is weird’. What should our response be? If it is an otherwise ordinary case and I haven’t checked for tricky lighting, surely the fact that this utterance was made raises the likelihood for me that there is in fact tricky lightning and so now I would fail to know the couch is crimson red. After all, why would someone say this if they didn’t have good reason to think there is tricky lighting? Similar remarks apply in the third person. If I am observing someone shopping at the store and this person has not checked for tricky lighting, it would also be reasonable to deny he knows the couch is crimson red (when doubt is raised).

Consider a different case. Suppose I am about to eat some berries I found in the field (I believe they are safe) and a friend tells me that he found similar looking berries that were poisonous. Suppose further that I didn’t rule out the possibility of poisonous berries (I didn’t check for poison). My friend’s utterance would likewise raise the likelihood for me that the berries I am about to eat are poisonous. Perhaps I now do not know that the berries I was about to eat are safe.

What these examples suggest is that there are certain cases where a rational response to somebody raising unresolved doubt *is* to deny knowledge. This is because the raising of unresolved doubt (in these cases) is itself a good indicator that there is evidence against the proposition in question. Let’s call these the ‘good cases’. I call these cases ‘good’ because the denials of knowledge seem true.

Now, there are also ‘bad’ cases. One is the brain in a vat scenario. You are a freshman in college and your philosophy professor raises the doubt that you may be a brain in a vat. You run back to your roommate with the bad news that it is all a big con. The possibility of error is unresolved leading you to judge you fail to know ordinary things. Another example is the lottery case where you deny you know you will lose the lottery when considering the possibility that you may win (which is not ruled out). These are ‘bad’ cases because they are problematic in the sense that they lead to epistemic paradoxes (lottery and skepticism). I do not want to assume here that these judgments are necessarily wrong however. By ‘bad’ I do not mean to say that the judgements are false.

The philosophical preoccupation with paradoxical cases has perhaps prevented us from seeing the value of skeptic-type judgments. Focusing on the “good” cases, it is plausible that they are produced by SM (since these are cases in which the raising of uneliminated doubt leads to denials of knowledge). And so we can see the function of having the SM: it helps us to correctly detect cases in which an agent fails to know. In other words, the SM is an ignorance detector:

(Ignorance Detector): The adaptive function of the SM is that it detects failures to know.

Two points of clarification. First, the SM need not get every case right. Wings are for flying but it doesn’t follow that birds with wings are able to fly in all weather conditions including heavy rain. The SM gets it right in the sorts of cases that we labeled ‘good’ above. I will leave it open for now whether it gets it right in skeptical scenarios (are skeptical cases for SM analogous to heavy rain for wings?). Second, the SM need not be the only way an organism comes to judge that someone fails to know. Opposable thumbs are adaptations for picking up objects, but it doesn’t follow that there aren’t other ways of picking up objects (we can pick up objects with our mouths, for example).

**4.2 Explaining Adaptiveness**

We now need to see how detecting ignorance is supposed to be adaptive. Saying that the function of SM is to detect failures to know is not quite to say how it can aid fitness. We need one more step. We need an explanation of how detecting ignorance aids fitness (the analogous case for wings is explaining how flying aids fitness). I will rely on the knowledge-action principles I mentioned earlier. But first let us look at some of the literature on the purpose of knowledge. Edward Craig (1991), Steven Reynolds (2017) and others have argued that knowledge attributions are *for* identifying reliable informants. [[22]](#endnote-22) If I say Fred knows whether the car needs repairing, I am identifying him as a reliable source of information regarding car repairs. If that’s the benefit of knowledge attributions, then since SM detects those who fail to know, the benefit of possessing SM is presumably that it helps us detect those individuals that are *not* reliable informants. The idea here would be that early on in evolutionary history, those individuals who possessed SM were better able to weed out bad informants and, as a result, were more likely to survive and pass on their genes.

I don’t find this line of reasoning promising. Here’s the problem. The SM judgments we discussed above involved peculiarities concerning specific cases that don’t seem to generalize to claims about reliable informants. When I judge (through SM) that my friend doesn’t know that the berries are safe to eat, it is not correct to think that he is an unreliable informant about the safety of berries in general. The only thing I can reasonably infer in this regard is that he did not check for poison in these particular berries. And when I judge (through SM) that an agent does not know the couch is crimson red (in the tricky lighting case), it would not be correct for me to infer that the agent is unreliable with respect couch colors. All I can say in this regard is that he failed to check for tricky lighting and so fails to know the color of the couch in this specific instance. More generally, SM judgments involve the processing of specific error scenarios that concern what might have gone wrong in a particular case. The inference to a general reliability claim seems unwarranted (even if we understand this inference as defeasible).[[23]](#endnote-23)

Relatedly, Sinan Dogramaci (2012) has argued that the function of rational assessments of beliefs is to get others to adopt (or reject) the epistemic method which lead to those beliefs-- which in turn makes testimony (reliance on the words of others) trustworthy. Dogramaci is explicit that he is concerned with a *sustaining* function of the practice of rational assessments, as opposed to a selection function (which is our present concern). In addition, he is interested in the function of a practice as opposed to the function of a cognitive mechanism. Nonetheless, we might try to extend his ideas to our present concerns (though he might not agree with this extension) by investigating the evolutionary roots of social stigmatization. Let us agree with Dogramaci that when we can call someone ‘irrational’ we are pushing the hearer to reject a certain belief forming practice. But how does this “pushing” work? What happens if one doesn’t go along with the speaker? One possible answer is that there is a threat of stigmatization in cases where the hearer fails to reject the belief forming practice in question. Now, Kurzban and Leary (2001) have argued that humans possess adaptive mental mechanisms for stigmatizing (as one of a suite of specialized adaptive mechanism for social interaction). The evolutionary benefit of the stigmatization mechanism involves avoiding social partners that would hinder our reproductive or social success (including avoiding uncooperative or pathogen bearing partners). These ideas applied to the SM would be that denials of knowledge are essentially threats. They are threats to reject certain epistemic practices on penalty of being ostracized from the group.

James Beebe (2012) has proposed another social role for knowledge attributions (or the ‘knowledge’ concept). He argues that knowledge is for distinguishing agents who are most blameworthy due to their un-cooperating behavior. For example, it is useful to detect the agents who are *knowingly* breaking the rules (as opposed to those who are doing so without knowledge that they are breaking the rules). In support of his thesis, Beebe brings up evidence from evolutionary game theory as well as work on the connection between knowledge attributions and blame.

These two social approaches to the function of knowledge are plausible. But they seem to suggest certain peculiar predictions which I am not sure are borne out. If knowledge attributions (or denials) involved threats to ostracize, we would expect to find that agents who make a knowledge attribution will monitor their audience to ensure they end up endorsing the relevant belief-forming method. In case they don’t, the speaker should be ready to punish their interlocutor by ostracizing or some other means. Although, we would need to get empirical evidence, it certainly doesn’t seem like we monitor our audience in this way today (though maybe things were different in the communities of our ancestors). Regarding Beebe’s account, his view seems to predict that knowledge attributions are more reliable in cases that involve cheaters and social norms (in the way that logical reasoning is thought to be more reliable in cheater domains). Again, empirical evidence would need to be brought in, but this is not something that has been noted by epistemologists who are ever so vigilant regarding subtle uses of ‘knows’.

A different and more direct approach appeals to the connection between knowledge and action. We noted that there is an intuitive connection between knowledge and appropriateness of action. According to (KA) (left to right), it is acceptable for X to use *p* in reasoning only if X knows *p*. And here we have a possible explanation for how SM can aid fitness. If the principle just mentioned is true and agents accept it, then when SM correctly detects failure to know *p*, one will be able to correctly deduce that it is not acceptable to use *p* in reasoning. For example, if my SM correctly detects that Fred fails to know the berries are safe to eat, I will be in a position to correctly judge that it is not acceptable for Fred to rely on the proposition that the berries are safe to eat. That is, he shouldn’t assume the berries are safe to eat. Possessing this type of information would give me an evolutionary advantage. If I am Fred, then learning that I shouldn’t eat the berries could save my life. If I am considering getting the berries from Fred, then unless I have some further information about the berries, I will also fail to know the berries are safe. And so by the principle, I can judge that I shouldn’t assume that the berries are safe (which could save my life). If Fred is my kin, then passing on the information that he shouldn’t assume the berries are safe could save his life and help preserve my genes. And in yet other cases, informing others (including Fred) that Fred shouldn’t assume the berries are safe can play a beneficial role cooperative strategies.

In the reasoning just rehearsed we used the (KA-left to right) principle. But we could also employ the other direction (right to left) of the principle: If X knows *p* (and *p* is relevant), then it is acceptable to use *p* in reasoning. On this conception, when the SM delivers that X fails to know *p*, we can’t directly infer that it is not acceptable for X to use *p*. But suppose that prior to activating the SM, you thought that X knew *p* and then (via KA-right to left) you judged that it is acceptable for X to rely on *p*. Now when the SM tells you that X doesn’t know *p* after all, you no longer have a reason to think it is acceptable for X to rely on *p*. So then you stop relying on *p*. This would be another way in which the SM could be useful.

The argument here depended on an internalized principle connecting knowledge and action. But as discussed earlier, the argument goes through even on a weaker connection between those concepts (principles involving *ceteris paribus* clauses). In addition, our proposal may over-intellectualize meta-cognitive processes. We just imagined our agents helping themselves to an abstract principle connecting the concepts and *deducing* a proposition concerning the appropriateness of action. But the process is likely more reflexive and automatic. What we require is that knowledge denials (or attributes) are deployed to appraise behavior. But we can be neutral about how this is realized in the mind.

Let us put all these strands together. In the “berries” case above, we said that the SM could help us avoid eating poisonous berries. The SM tells us X fails to know the berries are safe and so X shouldn’t assume the berries are safe. This suggests that the fitness benefits of SM is (excessive) risk avoidance. Whenever a possibility of error concerning *p* is made salient in a way that cannot be met by evidence, the mechanism tells us to stop relying on *p*.

I propose the following explanation for why detecting ignorance aids fitness:

(Risk Avoidance) The explanation for why the SM is adaptive is that it helps the agent avoid excessive risks.

Ignorance Detector tells us the adaptive function of the SM (to detect failures to know), but Risk Avoidance goes a step further. It gives us an explanation for why detecting ignorance should be adaptive in the first place. The hypothesis proposed is then that the SM a meta-cognitive mechanism evolved to help us avoid excessively risky decisions. It monitors reasoning and tells us to not assume *p* when there is a warning in the environment (i.e. when doubt that *p* is made salient) and this doubt cannot be ruled out by the agent’s evidence or epistemic position.

The SM gives us an advantage but it also comes at a cost. There will be cases in which the SM would needlessly slow us down. Suppose the meat you are preparing is fine to eat but un-eliminated doubt tells you, via the SM, that this assumption should not be made. This would lead to an unfortunate delay of your meal (because, for example, you may waste time double checking the quality of your food). The account I defend assumes, therefore, that the fitness costs of using of SM would be outweighed by the fitness benefits. This assumption is highly speculative. But we can reasonably speculate about the steepness of the costs. First, note that when the SM says that an assumption *p* should not be made, it does not mean that this cannot be reversed by central processing. Second, an agent may revert to reasoning with probabilities. For example, when the SM says you fail to know *p*, you could still assume that it is probable that *p* instead of assuming simply *p*. These considerations suggest that the costs of the SM are mainly loss of speed in decision making as opposed to missing out entirely on a useful type of reasoning. And so the costs of the SM are somewhat mitigated.

The SM is a sophisticated mechanism. We can imagine a more primitive system (a proto SM) which works by telling us to stop assuming *p* whenever there is a warning about the correctness of *p* in the environment (without checking to see if the doubt can be eliminated like in SM). The inability to weed out doubt signals that “should” be ruled out by the agent’s evidence would put possessors of this system at a disadvantage. Possessors of this proto-system would presumably be *overly* risk averse. In addition, clever competitors could easily exploit this weakness to gain advantage.

**5. Is the Skeptical Mechanism Modular?**

In the previous section I gave some reason to think that SM may have afforded our ancestors with an evolutionary advantage. But we want to go beyond a just-so story. What further evidence can we get that the SM was in fact an adaptation?

Natural selection works in a piecemeal fashion. The human body, for instance, did not evolve all at once. Rather, the human body contains organs each with a distinct evolutionary path. What goes for the human body goes for the mind. The visual system, for instance, has a different evolutionary history from the language processing components which is yet (arguably) distinct from our numerical cognitive capacities.

The idea that the mind might have a rich structure populated by mental organs or faculties was all but lost during the reign of behaviorism. It gained currency with Jerry Fodor’s influential ‘The Modularity of Mind’.[[24]](#endnote-24) According to Fodor, an important difference between these mental organs or “modules” and bodily organs is that the former process or compute information. As such, many of the features associated with modules are those that help explain the tractability of these computations.

We will begin by looking at some of the features Jerry Fodor proposed were central to modularity. As we will see, this list is perhaps too stringent. Evolutionary psychologists have wanted to relax some of these conditions. A main point of contention between Jerry Fodor and some evolutionary psychologists, for example, is that Fodor thought modules concerned peripheral processing and provided inputs to central processing systems. [[25]](#endnote-25) But this central processing system, with its open-endedness and holistic character, is itself non-modular. Some evolutionary psychologists, on the other hand, hold that central processing is itself populated by multiple modules.[[26]](#endnote-26) The Skeptical Mechanism, if modular, is a part of central processing. It is a meta-cognitive module in charge of monitoring reasoning (which tells us if an assumption is too hasty).[[27]](#endnote-27)

**5.1 Domain Specificity**

This feature is central to Fodor’s conception of modules. Linguistic modules concern language processing and vision module(s) concern visual perception. There some imprecision about this notion however. Jerry Fodor writes that “domain specificity has to do with the range of questions for which a device provides answers (the range of inputs for which it computes analyses)” (p.103). These ideas in the quote can be separated, however. The range of questions a mechanism answers could be small while the range of inputs could be large. For example, the theory of mind module (if there is one) will answer a narrow range of questions just concerning the mental states of others, but the range of inputs could be quite large if it could include any belief and any desire no matter what the content. And the same could be said about the cheater detector module (if there is one).[[28]](#endnote-28) It is domain specific in the sense that it is about cheating. Although the range of inputs could be about almost anything (one could cheat “about” anything).

What goes for theory of mind and cheating also goes for the skeptical mechanism. It is certainly domain specific in that it is just about *knowledge*, but the range of inputs could be quite unconstrained since, for example, doubt (input 1) could take on any subject matter. So we should count the SM as domain specific understood in the manner just discussed.

**5.2 Information Encapsulation.**

Modules do not have unfettered access to the agent’s beliefs or desires. This is part of what makes computation tractable and fast. A snake detector mechanism (if there is one) should not be free to survey all your random beliefs about reptiles before warning you that you are about to step on a cobra.

In the well-known Müller-Lyer illusion (figure 2), the visual system continues to deliver the representation (and the inclination to judge) that the lines presented are different lengths even after the agent knows fully well that they are the same length (Figure 2). This suggests that a perception module does not have access to the belief that the lines are the same length.

Figure 9.2: Müller-Lyer illusion



Similarly, the SM exhibits information encapsulation. Suppose you believe that skepticism is false. As noted earlier, you still feel the pull to deny you know you are not a brain a in a vat. You still feel the pull to deny you know you will lose the lottery. Many philosophers believe that skepticism is false. But they still feel like they have to explain the psychological pull of skeptical judgments. They try to provide psychological accounts for this pull (see citations in footnote 2). Why should this be? Encapsulation explains this.

Consider other philosophical discoveries. When Saul Kripke (1980) convinced others that there could be a posteriori but necessary truths, there wasn’t further work to explain away the psychological pull towards thinking that necessary truths must be a priori. This is because there is no left over pull towards that thesis after the arguments convince you. Reason is enough to banish any such inclination. Not so for skepticism. Reason is not enough to extinguish the allure of the skeptic. Encapsulation explains this.

**5.3 Inaccessibility to central processing.**

This is certainly correct for the grammar module(s). We cannot introspect the descriptive rules of grammar. The typical way to discover those rules is by looking for patterns in our intuitive judgments of acceptability.

This is similar to how we got that sensitivity is the algorithm of the SM. When philosophers introduced sensitivity as requirement for knowledge, it wasn’t an introspective report (otherwise, it would have happened a lot sooner). Rather, part of what we do is look at a pattern of judgments and note that sensitivity can explain and predict responses to those cases. Relatedly, there is a debate about whether practical interests play a role in ordinary attributions of knowledge. As we mentioned earlier, some theorists (Buckwalter and Schaffer 2015) hold that it does play a role but only in the following indirect way: when the stakes are high, the possibility of error is salient and this leads to a denial of knowledge. Grafting this idea onto our model, practical interests play a role in determining the input to SM, but that’s the end of it. Practical interests play no role in the internal algorithm of the system. I do not know whether this is the correct view. But the point is that we cannot settle this by introspection. This is evidence that the inner workings of SM are inaccessible to central processing.

**5.4 Mandatory processing.**

Modules are mandatorily triggered. An agent does not have to decide to recognize a face or process color. These things just happen. When you hear someone speak in your language, you immediately understand them. You don’t have to first *decide* to understand what they are saying. The evolutionary advantage of this feature is clear. Some reactions are too important and time-sensitive to be left up to the agent’s central process deliberations.

Unlike language comprehension, which can happen even if one is not attending to someone’s speech, SM seems to be only turned on when the agent is attending to practical or theoretical deliberation. Whether or not this means the SM is mandatory, it clearly reflects a good design. If doubt about *p* is raised about some issue we are not at all interested in, it would be a wasteful use of cognitive resources to execute a function which resolves that we fail to know *p*. Notice that other cognitive modules share the same feature. Numerical cognition is arguably modular, but we don’t just immediately add or subtract numbers when pairs of quantities are presented to us (Mandelbaum 2013). Rather, a precondition for mathematical reasoning is focus on a specific problem.

Eric Mandelbaum (2015) has pointed out that the mandatory feature of modules can be understood in two ways. First it could be understood as meaning “automatic” where this mean that the module will process an input when it is present (regardless of what the agent might be attending to). Second it could be understood as “ballistic”, meaning that when the process has started it cannot be stopped. These notions can come apart. Mandelbaum cites a proposal (Crain and Steedman 1985) for a linguistic parser that is automatic but not ballistic. It is not ballistic because the parser halts and goes to check for contextual cues when it is stumped by a garden path sentence. But it is automatic in that it processes speech even when agents are attending to other tasks.

Mandelbaum suggest that the cognitive modules (like our numerical capacities and cheater detectors) are best thought of as ballistic but not automatic. This seems right for SM. SM is only activated when attention is focused on some bit of reasoning. But once the system engages, it delivers the output. So the SM counts as mandatory if the notion is understood in the way just explained.

**5.5 Evolutionary Explanation and Innateness.**

Modules are generally thought to be evolved. As such, they should have an evolutionary function. We saw above that the evolutionary function of SM is to detect ignorance. In addition, theorists often assume modules have an innate component. This, of course, does not mean that the module does not need to develop and it does not mean that important components of the module are not fixed by the environment. The language modules, for example, develop through childhood and require critical input from the environment.

If the SM is modular, we would expect, for example, characteristic ontogenetic development of the mechanism. Although there is a great deal of research on how children attribute ‘knows’ and ‘thinks’, very little has focused on skeptical judgements. So I will say nothing further about this here.[[29]](#endnote-29)

What about innateness? If the module had an innate component, we would expect some cross-cultural convergence on skeptical judgements. Recent cross-cultural work indicates universality on Gettier judgments, but skeptical judgments haven’t been tested.[[30]](#endnote-30) It is plausible that at least some Gettier judgments fail the sensitivity test. I which case, at least some Gettier judgments may be the products of the skeptical module. If so, then the universality of these judgments would add plausibility to the innateness hypothesis.

Turning now to skeptical judgments. If skepticism had an innate component, we would expect it to naturally arise when humans turn to philosophical reflections. Indeed, we find skeptical traditions not just among western thinkers but also in Chinese (Zhunagzi) and Indian thought (Cārvāka).[[31]](#endnote-31) Further textual evidence would need to be uncovered to demonstrate that the way that skepticism is raised in these texts is consistent with the SM posited here.

Finally, I noted that SM is a meta-cognitive process in charge of detecting ignorance. There is now a great deal of evidence that non-human primates possess meta-cognitive mechanisms in charge of detecting uncertainty as well as ignorance. That is, they may be able to detect failures to know.[[32]](#endnote-32) There is then some reason to be optimistic about the existence of further evidence that the SM is an evolved mechanism.

**5.6 Functional Dissociability.** A key piece of evidence for modularity is the separability of the relevant capacity from the rest of the system. If there is really a module for X, we would expect to have breakdowns of the capacity to X while everything else remains more or less intact. Theorists have argued for the modularity of theory of mind on precisely this ground, for example.[[33]](#endnote-33) Do we have any evidence of an SM deficit? Obsessive Compulsive Disorder is a good candidate. Dubbed as ‘Foile du doute’ or ‘Madness of Doubt’ by the 19th century French psychiatrist Jean Etienne Esquirol, OCD looks like it involves a breakdown of the SM.

Recent cognitive accounts of OCD suggest that OCD is a type of epistemic disorder. It is posited that the failure of a patient to stop checking, washing their hands and other obsessive behavior is due to “Elevated Evidence Requirements” connected to their appraisal of how personally responsible patients felt for the outcome (including what is at stake). In obsessional cases, the criteria to stop the behavior is more stringent. The criteria includes “objective” epistemic markers like making sure one gathers “more evidence” (checking the door is locked one more time) as well as reaching a subjective “feeling of rightness”.[[34]](#endnote-34)

OCD might then involve a dysfunction of the SM mechanism. Possibilities of error that are normally ruled out by evidence are not. According to the SM, this leads to an inclination to judge that you don’t know the relevant proposition (i.e. that you locked the door). And via the connection to action, we can no longer help ourselves to that proposition. If the question is important to you, or seems important to you (as it is in virtually all obsessive cases), you will be prone to go out gather get more evidence by rechecking or washing your hands one more time. The SM will then be reactivated.

**5.7 Recap**

The skeptical mechanism may well be modular. Arguably, it possesses 6 key characteristics associated with modules.[[35]](#endnote-35) This gives us reason to think that the mechanism, with its stated function, is a product of natural selection.

**6. Debunking Skepticism? (Perhaps, but not in the way you might think)**

Debunking arguments begin with the claim that some of set of beliefs (the ones to be debunked) are caused (or explained) by factors significantly distinct from what the beliefs are supposedly about. This claim, if accepted, constitutes a defeater for those beliefs. That is, agents who uncover the suspect etiology are no longer justified in having these beliefs. For example, if Joe learns that his belief that the kitchen lights are flickering is caused by a powerful blow to his head, then he is no longer justified in having that belief (assuming he has no other reason to believe the kitchen lights are flickering).

A prominent class of debunking arguments appeal to evolution.[[36]](#endnote-36) The idea here is that some of our beliefs are caused by evolutionary forces instead of their purported subject matter. For example, our moral beliefs are shaped to help us survive and not by some objective moral reality. But if moral beliefs are tied up to commitments to some objective moral reality, then one loses justification for those beliefs upon learning that they are not explained by any objective moral realm (or so the account goes).

**6.1 First Attempt**

In this paper, I argued that skeptical judgements are shaped by natural selection. A tempting thought is that we have all the ingredients for a debunking argument against those judgments. Insofar as those intuitive judgments are what underwrite the premises in skeptical arguments, then we may have a way to block the skeptical advance.

To construct a debunking argument following the blueprint above, we need to first show that skeptical judgements are a product of evolution *and* not explained by their purported subject matter. But I do not believe this can be shown. I am in complete agreement that skeptical judgements are a product of evolution, but they are *also* explained by their subject matter. We saw that the Skeptical Mechanism (SM) takes as inputs an agent’s epistemic position together with salient doubt. It then tests to see whether the agent’s evidence can rule out the raised doubt (by testing for sensitivity). But “checking to see if doubt that *p* is ruled out by S’s evidence” (The SM algorithm) is pretty much the subject matter of judgments of the form “S does not know *p*”. So we cannot say that skeptical judgments are shaped by factors distinct from their subject matter. Now, it may turn out that this isn’t a very good algorithm for detecting failures to know. But this would require a more sophisticated argument going beyond the standard debunking move.

The lesson here is that it can be both true that a belief is shaped by evolution and explained by its purported subject matter even when the subject matter is not directly about fitness (this is well known in the debunking literature). Some beliefs formed through visual perception are like this. They are certainly shaped by evolution but are also appropriately caused by the things the agent is looking at. So just as we do not expect a debunking argument to be successful against perception, we should not expect one to be successful against skeptical judgments—at least not the version just given.

**6.2 Towards a Second Attempt**

Here’s a sketch of a better idea. We begin with the thought that the skeptical mechanism was “designed to function” in real world conditions where making hasty assumptions would hinder survival or reproduction. Global skeptical cases (involving brains in vats and evil demons), on the other hand, are products of philosophical explorations far removed from those real-world conditions.

From these facts, together with some background conditions which I will try to make explicit in a second, we can infer that we are not justified in believing the outputs of the skeptical mechanism *when it considers global skeptical cases*. That is, we can infer that we are not justified in accepting the skeptical intuitions which underwrite the key premise in the skeptical argument (or at least a leading version of the skeptical argument).

To see this, let us begin with an example from an artificial device. Suppose you buy an ordinary body temperature thermometer and you read in the specs that it is designed to work in the temperature range between 90F and 110F. You read that it was tested to work in that range (you know it wasn’t tested in other temperature ranges, suppose) and that it was the intention of the designers to work in that range. Later on you are in need of a meat thermometer. You look everywhere but can only find your body temperature thermometer. You stick it in the roast, wait a few minutes, and take the reading. It says 380F. I think that intuitions are clear that in this case you would not be justified in believing that the meat is 380F or even approximately 380F. This is so even if it turns out that the thermometer reading turned out to be correct and even if the thermometer turned out to be reliable at those high temperatures.

Another example is this. Suppose you take the thermometer to a far away planet where air pressure, temperature etc. are radically different from the original design conditions of the thermometer. You would also be unjustified in believing the outputs of the thermometer in those conditions.

Why should this be? The following principle seems plausible:

DEFEAT-ARTIFICIAL With some exceptions E\*, knowingly using an artificial device *far outside* its intended conditions renders beliefs in its outputs unjustified.

What are some of these exceptions E\*? I can think of at least two. If you have some further reason to think that the device is reliable in those conditions (suppose you tested the thermometer in extreme temperatures) or if you have some other independent way of corroborating the output (suppose you found an additional thermometer designed for those conditions), then you may be justified in believing the outputs after all. But absent these conditions, then I think you would be unjustified in believing the outputs of the device.

I think the general principle extends to natural devices. Of course, here we can’t appeal to the *intended conditions* since natural devices were not brought about by anyone’s intentions. But intentions are not really doing the explanatory work in DEFEAT-ARTIFICIAL anyways. The intentions of the makers are relevant because they tell us something about the reliability of a device. And they do so by telling us about how the device was made. For example, if the makers of a thermometer intended it to work in a certain temperature range, then they likely tested the thermometer to work in that range. But we can recover reliability information for natural devices by learning about how they were shaped by natural selection. If all we know about thermoreceptor cells on our skin is that they are selected for detecting temperature differences, then we would not be justified in thinking that they would work in conditions far removed from the sorts of conditions under which the receptors evolved. For example, if a scientist were to transplant those cells to a device in the lab, they would (absent some further information) not be justified in thinking that the cell receptors would continue to work under extreme temperature conditions, for example. This suggests the following natural analogue of our principle:

DEFEAT-NATURAL With some exceptions E\*\*, knowingly using a natural device *far outside* its natural design conditions renders beliefs in its outputs unjustified.

The exceptions E\*\* are similar to the exceptions E\*. If the agent had some independent reason to think that the device was reliable under the unusual conditions or if the agent had some independent reason to think that the specific output of the device was correct, then they would be justified in believing the output of the device in those conditions.

We can apply DEFEAT-NATURAL to the skeptical mechanism (SM). The key assumption (which can certainly be legitimately questioned) is that the global skeptical scenarios (brain in a vat, evil demon etc.) involve situations that are *far outside* the natural design conditions for SM. So we can conclude that, unless the exceptions E\*\* are operative, we are not justified in believing the outputs of the SM applied to global skepticism.

Are the two exceptions E\*\* operative? The first one does not seem operative. We do not seem to have independent reason to think that SM is reliable outside its design range. What about the second exception? Do we have independent reason to think that its output ‘I do not know that I am not a brain in vat’ is true beyond just relying on the intuition that it is true? If there is a further plausible argument for such a proposition (say, the “under-determination” skeptical argument), then we would need to look carefully at this argument to determine that that its plausibility does not still depend on the outputs of the SM. If it did, then this further argument would not count as an “independent” reason to believe the output of the SM. Hence, the exceptions E\*\* would not be operative and the agent would not be justified in believing the output of SM.

Let us then put these thoughts together. Consider the classical skeptical argument: (Premise 1) I do not know I am not a brain in a vat. (Premise 2) If I know I have hands, I know I am not a brain in a vat. (Conclusion) I do not know I have hands. I find this argument compelling in part because I find Premise 1 intuitive. But now I learn that my intuition for Premise 1 was produced by the SM and that it is operating far outside its design conditions. According to DEFEAT NATURAL, unless the exceptions E\*\* are operative, I am not justified in trusting this intuition. I am not justified in believing Premise 1. But if the justification in believing the conclusion of the skeptical argument is transmitted from the justification of the premises, then I am not justified in believing the conclusion of the skeptical argument after all. The skeptic is defeated!

Two points about this argument. First, it is an argument that I am not justified in believing the conclusion (and a premise) in the skeptical argument. It is not an argument that the conclusion is false. I have not argued that we know we have hands, for example. Second, the lack of justification for believing the conclusion (and Premise 1) does not apply to everyone. It just applies to people who *knowingly* use SM far outside its natural design conditions. But does anyone in fact knowingly use SM far outside its natural conditions? Not at the time of this writing. Before you read this paper, you probably didn’t have any beliefs about SM. The person who is not aware of the claims I made here, for example, will likely still be justified in accepting the conclusion and premises of the skeptical argument. There is nothing too strange about this. This just means that I have argued for something which, if true, is a type of *defeater* for the belief that we do not know we are not brains in vats (Premise 1).[[37]](#endnote-37)

**7. Summary**

In this paper, I argued that there is a mechanism SM which produces inclinations to make skeptical judgements. This mechanism takes as inputs (a) doubt concerning *p* and (b) *S*’s epistemic position with respect to *p*. The mechanism tests to see if S’s belief that *p* is sensitive. If it is not sensitive, then SM returns an inclination to judge that S does not know *p*. If it is sensitive, then it disengages. The range of skeptical inclinations produced by the SM run the gamut including global and local skeptical representations, as well as lottery, and statistical representations. I argued that the SM is a product of evolution whose “selection” function is to detect cases where agents fail to know some proposition. This function aids fitness because when an agent fails to know a proposition (it usually means) that the proposition should not be used in reasoning. In short, the SM is a meta-linguistic mechanism which monitors our reasoning for assumptions that are too hasty and, consequently, should not be deployed.

At the end, I suggested (given certain tentative assumptions) that global skeptical beliefs or judgments that arise through SM are unjustified. This is because global skeptical cases are unusual and far different from the sorts of cases that shaped the SM early on our evolutionary history. I indicated how we may exploit this information to undercut at least one version of the skeptical argument.

**BIBLIOGRAPHY**

Beebe, J. 2012. Social Function of Knowledge Attributions. Eds Jessica Brown and Mikkel Gerken Oxford University Press

Beran MJ, Smith JD, Redford JS, Washburn DA. 2006 Rhesus macaques (*Macaca mulatta*) monitor uncertainty during numerosity judgments. J. Exp. Psychol. Anim. Behav. Process **32**, 111–119

Beran MG, Smith JD. 2011 Information seeking by rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). Cognition **120**, 90–105

Buckwalter, Wesley & Schaffer, Jonathan (2015). Knowledge, Stakes, and Mistakes. \_Noûs\_ 49 (2):201–234.

Carey, S. and Spelke, E. 1994. Domain-specific knowledge and conceptual change. In L. Hirshfeld and S. Gelman (eds.), *Mapping the Mind*. Cambridge University Press.

Carruthers, Peter (2003). Moderately Massive Modularity. \_Royal Institute of Philosophy Supplement\_ 53:67-89.

Carruthers, Peter (2006). \_The Architecture of the Mind: Massive Modularity and the Flexibility of Thought\_. Oxford University Press UK

Cohen, Stewart (1998). Contextualist solutions to epistemological problems: Scepticism, Gettier, and the lottery. \_Australasian Journal of Philosophy\_ 76 (2):289 – 306.

Cosmides, L. and Tooby, J. 1992. Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind*. Oxford University Press.

Cosmides, L., & Tooby, J. (2005). Neurocognitive adaptations designed for social exchange. In D. M. Buss (Ed.), Handbook of evolutionary psychology (pp. 584–627). Hoboken, NJ: Wiley.

Craig, Edward (1991). \_Knowledge and the State of Nature: An Essay in Conceptual Synthesis\_. Clarendon Press.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), Natural language parsing (pp. 320–358). Cambridge: Cambridge University Press.

Descartes, René (1996). \_Meditations on First Philosophy: With Selections From the Objections and Replies\_. Cambridge University Press

Dogramaci, Sinan (2012). Reverse Engineering Epistemic Evaluations. Philosophy and Phenomenological Reasoning. Vol LXXXIv no. 3

Dretske, F. 1971. “Conclusive Reasons.” *Australasian Journal of Philosophy* 49: 1-22.

Enoch D, Spectre L, & Fisher T., Statistical Evidence, Sensitivity, and the Legal Value of Knowledge**.** Philosophy and Public Affairs. June 2012

Fantl, Jeremy & McGrath, Matthew (2002). Evidence, pragmatics, and justification. \_Philosophical Review\_ 111 (1):67-94.

Ferrigno, S, Kornell, N & Cantlon, J. (2017). A Metacognitive Illusion in Monkeys. Proceedings of the Royal Society B. 284: 20171541.

Fodor, Jerry A. (1983). \_The Modularity of Mind\_. MIT Press.

Friedman, Ori & Turri, John (2014). Is Probabilistic Evidence a Source of Knowledge? \_Cognitive Science\_ 39 (5):1062-1080.

Gerken, Mikkel (2017). On Folk Epistemology. Oxford.

Godfrey-Smith (1994), A modern history theory of functions. Nous 28(3):344–362

Griffiths (1993), Functional analysis and proper functions. Br J Philos Sci 44:409–422

Hampton RR. 2001 Rhesus monkeys know when they remember. Proc. Natl Acad. Sci. USA **98**, 5359–5362.

Hampton, RR, Zivin A, Murray EA. 2004 Rhesus monkeys (*Macaca mulatta*) discriminate between knowing and not knowing and collect information as needed before acting. Anim. Cogn. **7**, 239–246.

Hansen, Chad, "Zhuangzi", *The Stanford Encyclopedia of Philosophy*(Spring 2017 Edition), Edward N. Zalta (ed.)

Hawthorne, J. (2004). Knowledge and lotteries. Oxford: Clarendon.

Hawthorne and Stanley (2008) Knowledge and Action. \_Journal of Philosophy\_ 105 (10):571-590.

Hume, D. (1968). Treatise of Human Nature. Ed. L.A. Selby-Bigge. Oxford.

Jackson, A. (2015). “How you know you are not a brain in a vat,” Philosophical Studies, 172(10):2799-2822.

Joyce, Richard (2006), The Evolution of Morality (Cambridge: MIT Press)

Kim, Brian (2017) Pragmatic Encroachment. Philosophy Compass.

Kornell N, Son LK, &Terrace HS. 2007 Transfer of metacognitive skills and hint seeking in monkeys. Psychol. Sci.**18**, 64–71.

Kyburg, Probability and the Logic of Rational Belief (Middletown, Conn.: Wesleyan University Press, 1961).

Kraemer, D.M. Revisiting Recent Etiological Theories of Functions. Biol Philos (2014) 29: 747

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge: Harvard University Press

Kurzban, R. and Leary, M. (2001) Evolutionary Origins of Stigmatization: The Functions of Social Exclusion. Psychological Bulletin Vol 127 no 2 187-208

Lewis, David (1996). Elusive knowledge. \_Australasian Journal of Philosophy\_ 74 (4):549 – 567.

Mandelbaum, Eric (2013). Numerical Architecture. \_Topics in Cognitive Science\_ 5 (1):367-386.

Mandelbaum, Eric (2015). The automatic and the ballistic: Modularity beyond perceptual processes. \_Philosophical Psychology\_ 28 (8):1147-1156.

Machery, Edouard ; Stich, Stephen ; Rose, David ; Chatterjee, Amita ; Karasawa, Kaori ; Struchiner, Noel ; Sirker, Smita ; Usui, Naoki & Hashimoto, Takaaki (2016). Gettier Across Cultures. \_Noûs\_ 50 (4).

Millikan (1989), In defense of proper functions. Philos Sci 56(2):288–302

Nagel, J. (2010). Knowledge ascriptions and the psychological consequences of thinking about error. The Philosophical Quarterly, 60(239):286-306.

Nagel, J. (2012). Mindreading in Gettier Cases and Skeptical Pressure Cases. In J. Brown & M. Gerken (eds.), Knowledge Ascriptions. Oxford University Press.

Nagel, Jennifer ; San Juan, Valerie & Mar, Raymond A. (2013). Lay Denial of Knowledge for Justified True Beliefs. \_Cognition\_ 129 (3):652-661.

Neander (1991), Functions as selected effects: the conceptual analyst’s defense. Philos Sci 58(2):168–184

Nichols and Pinillos (Forthcoming). ‘Skepticism and The Acquisition of Knowledge’, *Mind and Language.*

Nozick, R. (1981) Philosophical Explanations. Oxford.

Papineau, David (2000). 8 The evolution of knowledge. In Peter Carruthers & A. Chamberlain (eds.), \_Evolution and the Human Mind\_. Cambridge University Press. pp. 170.

Phillips, Stephen, "Epistemology in Classical Indian Philosophy", *The Stanford Encyclopedia of Philosophy*(Spring 2017 Edition), Edward N. Zalta (ed.)

Pinillos, N. Ángel (MS) Mind and Doubt. Manuscript

Pinillos, Ángel (2012). Knowledge, Experiments, and Practical Interests. In Jessica Brown & Mikkel Gerken (eds.), \_Knowledge Ascriptions\_. Oxford University Press. pp. 192.

Pinillos, Ángel & Simpson, Shaun (2014). Experimental Evidence in Support of Anti-Intellectualism About Knowledge. In James Beebe (ed.), \_Advances in Experimental Epistemology\_. Bloomsbury Academic. pp. 9-44.

Pinillos, N. Ángel (forthcoming). Knowledge and the permissibility of action. \_Synthese\_:1-23.

Pinker, S. 1997. *How the Mind Works*. Penguin Press.

Powell, Derek ; Horne, Zachary ; Pinillos, Ángel & Holyoak, Keith (2015). A Bayesian framework for knowledge attribution: Evidence from semantic integration. \_Cognition\_ 139:92-104.

Rachman, S. (2002). A cognitive theory of compulsive checking. Behaviour Research and Therapy, 40, 625–639.

Reynolds, S; 2017. Knowledge as acceptable Testimony. Cambridge.

Rosati AG, Santos LR. 2016 Spontaneous metacognition in rhesus monkeys. Psychol. Sci. **27**, 1181–1191

Ross, Jacob & Schroeder, Mark (2014). Belief, Credence, and Pragmatic Encroachment. \_Philosophy and Phenomenological Research\_ 88 (2):259-288.

Salkovskis, P.M. (1999). Understanding and treating obsessive-compulsive disorder. Behaviour Research and Therapy, 37, S29-S52.

Slakovskis 2017 The Termination of Checking and the Role of Just Right Feelings: A Study of Obsessional Checkers Compared with Anxious and Non-clinical Controls.

Schaffer, J and Knobe, J. (2012) Contrastive Knowledge Surveyed. NOUS 46:4 (2012) 675–708

Scholl, B. J. and Leslie, A. M. (1999), Modularity, Development and ‘Theory of Mind’. Mind & Language, 14: 131–153

Sober, E. 1984. The Nature of Selection: Evolutionary Theory in Philosophical Focus. Bradford ⁄ MIT, Cambridge

Spicer, Finn (2008). Knowledge and the heuristics of folk epistemology. In Vincent Hendricks (ed.), \_New Waves in Epistemology\_. Palgrave-Macmillan.

Sperber, D. 1994. The modularity of thought and the epidemiology of representations. In L. Hirschfeld and S. Gelman (eds.), *Mapping the Mind*. Cambridge University Press.

Street, Sharon (2006), ‘A Darwinian Dilemma for Realist Theories of Value’, Philosophical Studies 127: 109-166.

Turri, John (2014). Skeptical Appeal: The Source‐Content Bias. \_Cognitive Science\_ 38 (5):307-324.

Turri, John & Friedman, Ori (2014). Winners and Losers in the Folk Epistemology of Lotteries. In James Beebe (ed.), \_Advances in Experimental Epistemology\_. Bloomsbury Press

Turri, John, Ori Friedman & Ashley Keefner (2017). Knowledge Central: A Central Role for Knowledge Attributions in Social Evaluations. *Quarterly Journal of Experimental Psychology*. Vol 70, Issue 3, pp. 504 - 515

Wahl, K., Salkovskis, P.M. and Cotter, I. (2008). [`]I wash until it feels right': the phenomenology of stopping criteria in obsessive-compulsive washing. Journal of Anxiety Disorders, 22, 143-161.

Wells, G. L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology, 62*(5), 739-752.

Williams, G.C. 1966. Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought. Princeton University Press, Princeton.

Williamson, T. (2005). Contextualism, Subject‐ Sensitive Invariantism and Knowledge of Knowledge. The Philosophical Quarterly, 55(219), 213-235.

Wright (1973). Functions. Philos Rev 82(2):139–168

1. \* I would like to thank Sinan Dogramaci, Brian Kim, Shyam Nair, Shaun Nichols, Mark Schroeder, and audiences at The University of Buffalo and at Institut Jean Nicod for helpful feedback. [↑](#endnote-ref-1)
2. Descartes (1996). [↑](#endnote-ref-2)
3. There is a sense in which contextualists about ‘knowledge’ are also trying to explain why we find skepticism appealing. But their focus is different from the approach I take here which concerns the psychological mechanism which gives rise to our inclinations as well as issue of why we would possess the mechanism in the first place. Their approach, in contrast, is essentially a semantic approach. Contextualists need not have a theory of why language and ‘knows’ works the way it does. [↑](#endnote-ref-3)
4. Some recent work: Gerken (2017), Jackson (2015), Hawthorne (2004, 164), Nagel (2010), Nichols and Pinillos (forthcoming), Turri (2014), Turri and Friedman (2014), Williamson (2005) and many others.

   [↑](#endnote-ref-4)
5. Of course, these other approaches also focus on denials since they concern skepticism. [↑](#endnote-ref-5)
6. The defense that this system is genuinely “meta-cognitive” is left for another time (Pinillos MS). But readers familiar with that literature will notice the connections. First, the mechanism has essentially a monitor and control flow (monitoring for knowledge following control of our practical reasoning). Second, the algorithm deploys a sensitivity test which is just a counter-factual test directly about mental states (beliefs). Third, the mechanism involves a mix of conscious reflection and implicit inference through an opaque mechanism (think of how “feelings of knowing” arise). [↑](#endnote-ref-6)
7. See also Ross and Schroeder (2014) and Pinillos (forthcoming) [↑](#endnote-ref-7)
8. The Knowledge-Action principles are intimately connected with the thesis of pragmatic encroachment. I will not discuss encroachment per se here. See Kim (2017) for an overview. [↑](#endnote-ref-8)
9. See Pinillos (forthcoming) [↑](#endnote-ref-9)
10. Philosophers who prefer to think in terms of functions (as opposed to universal principles) might take the folk data cited above to tell us that the *function* of *knowledge attributions* is to assess whether a premise is actionable. If this is right, Brown’s counter-examples would be less bothersome. Chairs are to sit, but they can be used for other purposes. In this paper, I will be discussing the evolutionary function of certain cognitive mechanisms which produce inclinations make knowledge-judgments. I will have nothing to say about the function of knowledge *attributions* per se. [↑](#endnote-ref-10)
11. It should be clear that I am interested here in Marr’s algorithmic level of cognitive analysis as opposed to the implementation or computational levels. In other work, (Pinillos MS) I have tried to address the computation or function the agent is trying to approximate when responding in a skeptical pressure case. [↑](#endnote-ref-11)
12. Contextualist theories are semantic/pragmatic not psychological theories. Some of these theories will tell us, for example, that ‘X knows P’ is false in a context in which the possibility of error is salient (or attention is drawn to the possibility). It won’t tell us that in fact people will deny knowledge in those contexts. However, these theorists get their plausibility and motivation from their ability to capture ordinary behavior. [↑](#endnote-ref-12)
13. There are many formulations of “closure”. We need not worry about getting it exactly right here. A plausible enough version says this: If *S* knows *p* and correctly deduces *q* from *p*, then S is in a position to know *q*. [↑](#endnote-ref-13)
14. Pinillos (2012), Sripada and Stanley (2012). [↑](#endnote-ref-14)
15. Buckwalter and Schaffer (2015) and for a response see Pinillos and Simpson (2014) [↑](#endnote-ref-15)
16. Fodor (1983) [↑](#endnote-ref-16)
17. Hume (1968, 187) [↑](#endnote-ref-17)
18. See also Dretske (1971) [↑](#endnote-ref-18)
19. According to Nozick, Sensitivity plus some other conditions plausibly met in the paper-reading case are sufficient for knowledge. [↑](#endnote-ref-19)
20. It may be that the mechanism posited serves other functions besides those resulting in denying knowledge. For example, Brian Kim suggests plausibly that it can help us eliminate alternative hypothesis in deliberation. [↑](#endnote-ref-20)
21. This view is sometimes known as the ‘Selected Effects Theory of Functions’. See, for example, Kraemer (2014), Godfrey-Smith (1994), Griffiths (1993), Millikan (1989a), Neander (1991), Wright (1973). [↑](#endnote-ref-21)
22. I am not attributing to Craig or Reynolds claims about evolutionary (selection) functions. They seem to have some other type of teleology in mind. In addition, Craig and Reynolds both talk about the function of knowledge *attributions* (or concepts). In contrast, I am only making a claim about the selection function of a psychological mechanism. Nonetheless, we can try to extend their insights to our present concerns.

    . [↑](#endnote-ref-22)
23. Nor does it help to say that the agent is unreliable with respect to the specific proposition that *these* berries are safe and *this* crimson couch is red. Whether or not a person is reliable with respect these specific propositions is not very useful information since it is very unlikely that the agents will make a judgment about these particular propositions in the future. What is useful information is whether an agent is reliable with respect to a topic or class of propositions. [↑](#endnote-ref-23)
24. Fodor (1983) [↑](#endnote-ref-24)
25. See Carruthers (2003) and the references therein. [↑](#endnote-ref-25)
26. Carruthers (2006), Carey and Spelke (1994), Cosmides and Tooby (2002), Sperber (1994), Pinker (1997), [↑](#endnote-ref-26)
27. Finn Spicer (2008) and David Papineau (2000) argue that we have a folk epistemology module which produces positive knowledge judgments. It is important to distinguish this claim from the one I am making here. I am making a specific proposal regarding modularity and a type of knowledge denial. I am neutral about whether other “knowledge” judgments are modular or whether folk epistemology is produced by a module. [↑](#endnote-ref-27)
28. Cosmides and Tooby (2005). [↑](#endnote-ref-28)
29. Although developmental work on skepticism is being carried out through the Children’s Museum of Phoenix and the Cognition, Computation and Development Lab at Arizona State University. [↑](#endnote-ref-29)
30. Machery et al 2016 [↑](#endnote-ref-30)
31. See Hansen (2017) for an overview of Zhunagzi and Phillips (2017) for discussion on the Cārvāka school. See the sections on skepticism. [↑](#endnote-ref-31)
32. Beran et al (2006), Beran and Smith (2011), Ferrigno et al (2017), Hampton (2001), Hampton et al (2004), Kornell et al (2007), Rosati and Santos (2016) [↑](#endnote-ref-32)
33. Scholl and Leslie (1999). [↑](#endnote-ref-33)
34. Salkovskis 1999, Salkovskis et al (2017), Rachman (2002), Wahl (2008) [↑](#endnote-ref-34)
35. Jerry Fodor also include “shallow outputs” as part of the criteria for modularity. I argued on 3.2 that the outputs of SM are indeed shallow. [↑](#endnote-ref-35)
36. See, for example, Street (2006) or Joyce (2006). [↑](#endnote-ref-36)
37. Brian Kim suggests that a full explanation of the psychological pull concerns why we would generate skeptical hypotheses in the first place. This is correct and I have said nothing about this here. Though an implicit claim being made here is that local skeptical hypotheses naturally arose in our evolutionary history. [↑](#endnote-ref-37)