## Generality and Modularity

Having found all existing solutions to the generality problem unsuccessful, Conee and Feldman conclude,

> In the absence of a brand new idea about relevant types, the problem looks insoluble. Consequently, process reliability theories of justification and knowledge look hopeless. (1998: 24)

I don't have a brand new idea, so it's lucky I don't want to defend a process reliability theory of justification. But I do have someone else's middle-aged idea, and it can be used to provide a halfway decent solution to the generality problem. And solution in turn gives me a theory of justification in which process reliability is a significant part of the story. The idea is Jerry Fodor's modularity of mind thesis. I propose that the justificatory status of a belief that is produced by an innate module is a function of the reliability of that module. Not all beliefs are produced by innate modules. Countlessly many beliefs are non-modular; if Fodor is right they include many of ours. The story about these beliefs should be much more familiar to internalists; non-modular beliefs are justified to the extent that they are supported by the believer's evidence. But even here there's a modular twist: evidence consists of the outputs of reliable modules.

The upshot is yet another epistemology that draws on elements of internalism and externalism. There's been so much cross-breeding between internalist and externalist theories recently that we should be the subject of protests against genetic engineering. But there's still space for a few more variants. My variant has two major virtues. First, by selectively picking elements from both internalist and externalist theories, I avoid the objections to each of the pure theories. Secondly, even though the theory itself is complicated, the complications are grounded in distinctions that carve cognitive nature at its joints, so the theory is not objectionably *ad hoc*.

Crude versions of reliabilism about justification face three pressing objections. (Reliabilism about knowledge faces other objections, and that won't be our focus here.) For clarity, I'll formulate these as objections to a particular version of the theory.

(R) A belief is justified iff most of the beliefs produced by the same process as it was produced are true.

First, there is no plausible way to state when two beliefs are of the same process. This is an old objection, but Conee and Feldman's paper reminds us that it's never really been solved. Secondly, (R) leads to excessive amounts of epistemic *good* luck. BonJour's discussion of the clairvoyant and Plantinga's discussion of the brain lesion constitute objections of this kind, though we shall see that they don't exhaust the range of possible cases of epistemic good luck. The third objection is that (R) leads to excessive amounts of epistemic *bad* luck. There could be different ways of putting this problem, but the "new evil demon problem"[1] seems to be the most general version. For the moderate, who wants to use (R) as a part of the true theory rather than all of it, the generality problem is the most pressing. For reflection on the generality problem suggests that (R) is meaningless until we specify identity conditions for processes, and there will be no way to do this such that distinctions between processes track epistemically important distinctions. The other two objections merely suggest that (R) is false. But a false proposition can, when used correctly, be a part of the true theory, perhaps as the consequent of a carefully framed conditional. On the other hand, a meaningless theory is useless. So the generality problem must be dealt with before anything else is considered, and our answer to it should guide our answer to the other two objections.

I will use Fodor's modular theory of mind to respond to all three objections. Beliefs are produced by the same process iff they are outputs of the same module. I will deal with the problem of good luck by putting some qualifications on (R) that are suggested by Fodor's theory. Finally, I will attack the problem of bad luck by drawing on features of Fodor's views about central processes. Fodor argues, and I agree, that central processes are not modular. I think[2] this suggests that if we are in the business of evaluating

---

[1] (Cohen 1984), (Sosa 1991).

[2] At this point, and I think at this point *only*, I'm in explicit disagreement with Fodor. He says that the location of then boundary between the modular and non-modular parts of the mind is not epistemologically interesting (1983: 88), whereas it will be crucial to my story that it is rather interesting. But this might be a terminological difference, rather than a theoretical difference, for it seems that in saying this Fodor merely meant that the boundary in question is not in any way related to the observation/theory distinction as this is construed in (many versions of) empiricist epistemology. I agree, but I think the boundary might be epistemologically interesting in other ways. None of this is to say that Fodor would agree with much of what I say here were he to comment upon it, just that everything else I say is consistent with Fodor's explicit views.

beliefs by the quality of their production, we should allow that quite different standards apply to beliefs produced by modules and beliefs produced by central processes. Suggestion noted, and adopted. The upshot will be that victims of an evil demon do *not* have justified beliefs, but they are not thereby epistemically blameworthy in the way that many holders of unjustified beliefs are blameworthy, and this explains our intuitions about victims of demons.

**1. Fodor on Modularity**

It's helpful to formulate Fodor's theory by contrasting it with two of its contemporary rivals. Both of the theories hold that the mind is basically a computational system, so it needs inputs presented in a computationally tractable manner, i.e. syntactically. So all contact with the world must be mediated by analogue to digital *transducers* that convert signals from the world to syntactic forms. The disputes start over where the output of these transducers goes.

The first theory, the no module theory, says that these outputs are fed directly into a giant general information processing system, which we may as well identify with the mind. The way the mind works is that it forms hypotheses, tests these against the outputs of the transducers, and (at least ideally) forms beliefs by applying rules of good scientific practice in light of the test results. So hypotheses get turned into beliefs when they explain all the evidence and are not in serious conflict with any of it, or something like this.

The second theory, the all modules theory, is impressed by the way in which several of our belief-forming processes do not resemble careful scientific inquiry. One kind of difference is that sometimes our beliefs get formed rather quickly. A panther-shaped black region appears in our visual field, and we rather quickly conclude it is a panther. We conclude this much more quickly than we could possibly process all the possible explanations of the black patch, let alone evaluate which of them best fits the evidence. There is a good reason for this: believers who are not this quick quickly become non-believers. Another difference is that once we've formed the belief we often lose the evidence for it, so we do not learn about some rather distinctive aspects of that evidence on repeated exposure.[3] Thirdly, in some areas the

[3] Quick pop quiz: without looking, do the lecture hall clocks at your college have Arabic or Roman numerals? What about your watch? Which direction do the ex-Presidents face on American coins? Even if you'd just looked at a clock, or a coin, and even if you'd done so do get some information from them, like what time it is, you probably

conclusion seems almost compulsory, which is not true in other areas even when only one hypothesis is consistent with your deeply held beliefs. No matter how well you've integrated the orthodox theory of rational choice, you can see some rational display of resource allocation without seeing it *as* a maximisation of marginal utility.[4] On the other hand, if you've integrated a theory of meaning for English you can't hear me say, "I like watching billiards on TV" without hearing it *as* me saying that I like to watch billiards on TV.[5] Finally, the compulsory outputs here are quite insensitive to your prior expectations, and quite sensitive to the inputs. You'll still hear that utterance as my saying that I like to watch billiards on TV even if you firmly believe that I don't like watching TV, and that I never say things I know to be false. On the hypothesis testing model, you'd presumably come to believe that you misheard something. To be sure, in practice you might on reflection conclude that was really what happened, but at first you'll believe that I said I like billiards on TV.

The conclusion drawn from this is that the brain is not a general purpose hypothesis testing machine, but rather a series of special purpose modules. There is a module for recognising medium shaped physical objects, another for recognising and interpreting language, another for recognising faces, another for recognising cheating, and so on. Modules have a few important features. First, they are domain specific: the face recognition module doesn't help with parsing. Secondly, they are subconscious: we are ignorant of most of the data that makes us conclude that there's a nickel on the table. Thirdly, they are automatic, which is why we can't help hearing language as language. Fourthly, they are informationally encapsulated. The modules have access to some information. The language module, for instance, contains theories about the grammar and lexicon of the subject's native language, for instance, and uses this in interpreting utterances. But it does not have access to theories about my boring character traits, or at least it inevitably ignores this information, as illustrated above. Finally, they are innate. The information in the modules may include acquired data, but the existence of the modules is actually innate.

---

couldn't say unless you'd been prepared for the question. This isn't what we'd expect on the hypothesis testing model.

[4] For an amusing illustration of this, see Stephen Landsburg's article in *Slate* "One Small Step For Man…" about how it could be rational to walk up stairs but stand still on escalators. Link at: http://slate.msn.com/?id=2070182.

[5] Incidentally, as Fodor points out this example also illustrates the previous point. Without looking back at the text, in the quoted text did I (hypothetically) say that I like *watching* billiards on TV or that I like *to watch* billiards on TV?

The no module theory cannot explain why some processes, like linguistic interpretation, are not like hypothesis testing. But, as Fodor stresses, all modules theory cannot explain why some processes we actually engage in, like say hypothesis testing, are actually a lot like hypothesis testing. At least some of the time we reason analogically, or we draw together evidence from many disparate sources. Whatever part of the mind does that is not informationally encapsulated, and hence it is not a module. So neither of these theories is entirely correct. On Fodor's view the mind contains transducers that convert analogue signals to digital information, input analysers, modules that quickly draw some salient information out of that signal, *and* central processors that draw conclusions on the basis of (potentially) disparate sources of information[6], including analysed signals from the modules, and (sometimes) raw signals from the transducers that are not fit to be analysed by any module[7]. This implies that there are *fewer* modules than the all modules theory holds; there's no cheater detection module for example. For current purposes, I'm going to assume Fodor is entirely correct. This is probably foolish; everyone makes mistakes somewhere or other. But I'm not the person to find where they are, so this paper will be an attempt to write an epistemology for Fodorian cognisers.

## 2. A Modular Theory of Justification

We need to make a few final distinctions before the theory is presented. A believer is *intelligent* (for current purposes) iff it has a non-modular (i.e. informationally unencapsulated) central processor that makes judgements about which outputs of the modules shall remain as beliefs. (I may appear to be in dispute with Fodor here. I have said, and will say repeatedly, that the modules produce beliefs. Fodor

---

[6] Note that the dispute with the second theory is not whether there are central processors or not. There must be some kinds of processors if for no other reason than to facilitate communication between various input modules and output modules. The dispute is whether those central processors are themselves modules, which is to say, whether they are informationally encapsulated. If the central processor is little more than a switchboard, it could well be. And I will assume below that this might be true for the (very) young and for (much) lower animals. But Fodor's arguments that it is not true for *us* seem pretty persuasive.

[7] That last bit is a little speculative, because it doesn't seem to be stressed much, but I don't see any reason to suppose that we have an input analyser for any possible analysandum. So sometimes raw (or merely transduced) signals must be sent straight to the central processors. But this is an empirical question and I'm really not qualified to say what its answer might be.

sometimes says that the modules produce hypotheses, and the fixation of belief is the task of the central processor. I don't think that's *always* the case, and I don't *think* Fodor means to suggest it is. If you see a panther-shaped black patch, and your object recognition module outputs, *That's a panther*, and then you run, what's the best explanation for why you ran? I think, boringly enough, that it's that you believed there was a panther there, and you desired pantherlessness. This explanation is not vitiated by your *later* coming to believe that it was merely a visual illusion. And it's not even vitiated if this realisation occurs a matter of seconds, or even milliseconds, later. I take this case to be fairly general - the module produces a belief, and the role of the central processor, in its role as defender from illusions, is to *overturn* that verdict.[8] Until and unless the central processor intervenes, the belief is the responsibility of the module.[9])

A belief is *pre-cognitive* iff an agent acquires it before she becomes intelligent, *cognitive* otherwise. Innate beliefs are pre-cognitive, I leave it as an open question how many other beliefs are pre-cognitive, though I suspect many beliefs about the types of physical objects that are prevalent in the nearby environment *may* be acquired pre-cognitively.[10] Here, at last, is the theory.

---

[8] Some may think that I've got a deeper disagreement with Fodor, because I'm taking the output of modules to be conceptually structured, and he takes them to be shallow, i.e. non-conceptual. Peter Carruthers (2003: sect. 1.2), for instance, says exactly that. But it's hard to square with the textual evidence. Fodor says some of the most obvious candidates to be non-conceptual outputs, such as Marr's '2.5 D' sketches, are too shallow to be the output of the visual module (1983: 94), that the outputs of that module are likely to be in terms of basic categories (1983: 94-7), and that it's literally possible to see that it is raining (1983: 136), which means, in the context of the theory, that the visual module reports that it is raining. Given all that, I take it that there *might* be a disagreement about whether the outputs are really beliefs, but not about whether they are of the right category to be believed.

[9] Note that it's intervention that matters, not the overturning. If the central processor 'considers' the verdict of the module and upholds it, the belief becomes the responsibility of the central processor

[10] The point is not that an infant has to acquire a central processor in any traditional sense of acquire. The nature of the central processor is presumably largely determined by innate properties of the child. But the processor may not be formed, or functional, in neonates, so it is *as if* they acquire a central processor later in life. Again, the arguments for our having a central processor do not *immediately* apply to infants, since it *seems* they do not engage in things like hypothesis testing, though this is an area of extensive psychological dispute.

1. A module is *reliable* iff

   (a) most of the beliefs it produces are true[11]; and

   (b) most of the information it has access to is *certified*.

2. A belief constitutes *certified* information iff

   (a) it is pre-cognitive and true, or

   (b) it is cognitive and has previously been justified.

3. A belief is *justified* iff

   (a) it is the responsibility of an innate reliable module or a module acquired before the believer became intelligent; or

   (b) it is the responsibility of a module acquired after the believer became intelligent and the information that that module is reliable has previously been certified; or

   (c) it is the responsibility of the central processor, and is supported by the outputs of reliable modules.

There may appear to be some circularity here because each of the three terms I've defined uses some of the other terms. But the circularity vanishes if we remember the properties here are all tensed. The important point is that a belief becomes certified information *because* it is justified in the past, by other means. It can *then* be added to a module's stock of information. So if we trace an agent's doxastic life out in time the above definitions determine unambiguously whether a particular belief is justified. I suspect readers will have many questions about the above theory, so before we get onto how the theory handles the three objections to (R), I'll answer the most pressing ones.

Q: Why so complicated?

A: Because the concept is complicated. Our philosophical notion of justification is an amalgam of two simpler concepts: reliability and groundedness. The debate between externalists and internalists is about which of these is more central to justification. I say they're each central for different kinds of creatures. A creature that has no non-modular processors, who just receives evidence and coverts it to beliefs via

---

[11] The pronoun 'it' here refers to the token module in question, not to modules of that type. The reason for this is that a module in environment $e_1$ may intuitively be unreliable even if most modules of that type occur in environment $e_2$, and in that environment the modules regularly produce true beliefs.

modules, is, intuitively, justified in its beliefs to the extent those modules rely on truths and produce truths. If crows and cockatoos have purely modular minds[12], but the object recognition module in cockatoos produces truths more often than the matching module in crows, or contains more true information, so it's not making lucky guesses, then a cockatoo is, it seems to me, more justified in thinking there's a worm on the ground than a similarly situated crow. At the other extreme is the kind of creature found in the no module theory. This creature only ever forms beliefs as a result of theorising about perceptual inputs. It seems plausible that the right theory of justification for it is some kind of internalist theory, probably a foundationalist theory. Since its beliefs are the product of reasoning, it is the quality of that reasoning that determines the justificatory status of its beliefs. The difficulty arises in applying the concept of justification to creatures that are 'between' our two extremes. (That would be creatures like us.) The active intuition here is that a particular belief of ours is justified iff it is produced by methods like the cockatoo's belief and it is justified by standards appropriate to that, or it is produced by methods like the pure thinker's belief, and it is justified by standards appropriate to that. All the complications flow from explicating that intuition.[13]

Q: But aren't all these qualifications ad hoc?

A: Not once we've recognised the fractured nature of the concept. And this recognition is practically forced upon us. Once we've decided to start judging beliefs by some criteria other than their truth or falsity, we must be strongly tempted to judge them by the quality of their origins. This is common ground, I think, to almost all hands in current debates about justification. And once we accept that beliefs may have radically different *kinds* of origins, once we accept that modular input systems form a natural kind (which we're accepting as part of the universal acceptance of Fodor's theses about the mind) we should expect our explication of justification to say different things about each kind. By my lights, a theory of

---

[12] Note that I don't say that even they have purely modular minds, this is all just a thought experiment.

[13] There are obvious, and intentional, similarities here between my theory and Ernest Sosa's distinction between animal knowledge and human knowledge. There are two differences, one minor one of some importance. The minor difference is that I'm talking about justification and Sosa about knowledge. The major difference is that Sosa thinks that there are two concepts here, one of which is only appropriate to creatures with certain reflective capacities. I think there's one concept with different facets. Some of the reasons for this difference should become apparent in what follows.

justification that wasn't this fragmented would be as plausible as a theory of what makes a good baseball player that didn't have distinct clauses for hitters and pitchers.[14]

Q: Doesn't this assume we know all the kinds of belief forming mechanisms?
A: Yes, and that at least *seems* false. The theory may need to be made more complicated in light of future research showing there are more natural kinds in cognitive science than we've previously accepted.

Q: Why are acquired modules treated differently?
A: Intelligent creatures have certain epistemic rights. In particular, they have the right to acquire non-perceptual justified beliefs, beliefs based on a variety of different perceptions. But that right gives them epistemic responsibilities, in particular the responsibility to check that the automatic belief generating mechanisms they acquire, i.e. new modules, are reliable. It's no fault of the cockatoo (or the infant human) that it doesn't check whether its vision is reliable before it uses it, because it can't. But if it could have, it should have.[15]

Q: What's 'supported' mean here?
A: Good question! I mean here the kind of relation that evidentialists and foundationalists have in mind when they say that a belief is supported by the foundational evidence. There is a lot of work to be done on explicating this concept, and I have little to add to the existing work by, among others, BonJour, Conee

---

[14] I don't just mean here that a decent theory of what makes a good ballplayer should say that pitchers should get make outs and hitters should avoid them. Current orthodoxy strongly suggests that it matters quite a bit how pitchers get hitters out (lots of strikeouts are a sign of a good pitcher, lots of ground ball and fly ball outs a sign of a lucky one) but it matters very little how hitters make outs (ground ball outs are a sign of poor skill, not poor luck, for a hitter). It doesn't seem *that* surprising that our theory of what makes for a justified belief, as opposed to a lucky guess, should be more complicated than our theory of what makes for a good, as opposed to lucky, ballplayer.

[15] When I presented an ancestor of this paper in New Zealand someone suggested I call this a republican theory of justification because of this feature. It's a nice suggestion, but since I can't remember exactly who made it, I'm not sure I could properly give credit where credit's due were I to use it.

and Feldman, and Vogel[16]. I think the aspects of externalism that my theory adopts makes my theory preferable to each of theirs, but they each have things to say that can be adopted into my theory here.[17]

Q: Does this theory imply that a pure thinker with no innate beliefs has no justified beliefs?
A: Yes. The sceptics who say that such a thinker would have no grounds for choosing her preferred theories of the world to other theories compatible with her evidence are correct. That's a rather nice explanation for why they are so hard to refute. Fortunately *we* are nothing like such thinkers.

**3. The Generality Problem**

Conee and Feldman raise three kinds of concerns for purported solutions to the generality problem.

Some solutions are too vague. Some vagueness is not a problem, the analysandum here is vague so the analysans should be too, but it's a platitude that sometimes beliefs are determinately justified or unjustified. This isn't a problem for my theory.

Some solutions turn on epistemically insignificant factors. Solutions that individuate processes in terms of the *explanations* of the belief they produce face this kind of problem. Which explanation is best depends on our interests. But whether someone else's belief is justified is independent of our interests.[18] Again, this kind of problem doesn't seem to arise for my theory.

The final problem is the most pressing. It's a consequence of any reliabilist theory that all beliefs produced by the same processes are equally justified. Reliabilist theories are indiscriminate in this way.

---

[16] (BonJour 2000), (Conee and Feldman 1985), (Vogel 1990). Of course these are only a sample of the many important foundationalist works that a theory of grounding should drawn from.

[17] And hopefully will be in subsequent work on the theory.

[18] Or so I say. There's actually quite a large debate possible here about whether contextualism about justification is true. If it is, then the analysandum and analysans might contextually vary together. *Very* briefly, I doubt that contextualist theories here, or elsewhere in epistemology, are true because I think that in general there's no contextual variation without this variation being controlled by unpronounced syntactic elements, there are usually ways of testing whether such elements are present, and epistemic ascriptions fail all these tests. See (Stanley 2000) for (Stanley forthcoming) for some salient discussion of these topics.

But intuitively this is not always the case.[19] Smith may be more justified in believing that the object in front of her is a tree than that it is a maple tree, even if she acquires both beliefs by visual inspection.[20] This particular case is not a problem for my theory, for the object recognition module does not proclaim that the object is a maple tree. As Fodor notes, the concepts that feature in its outputs are 'basic categories' (1983: 94ff). Assume that we have a category hierarchy; Fodor uses *poodle, dog, mammal, animal, physical object*. The basic category in the hierarchy is "the least abstract member of the hierarchy that is momomorphemically lexicalised" (95), in this case 'dog'. Basic categories have many other features. Words for them are learned earlier and more frequently, they can be taught be ostensive introduction, and they are usually phenomenologically given.[21] So while we *see* that there's a tree, we *infer* that it's a maple, and that belief is only justified to the extent it is supported by the evidence, and hence can be no more justified than the belief that it's a tree, as required.

Sadly, this is not the only kind of indiscriminability objection we face. Smith glimpses what might be (and actually is) a panther lurking behind the tree. The object recognition module screams *PANTHER!* and she runs. Arguably her belief here that there's a panther nearby is less justified than her belief that there's a tree nearby. It's tempting to retreat and amend the theory here, perhaps to say that (in effect) beliefs are produced by the same process iff they are outputs of the same module and the same information is *used* in generating the two beliefs.[22] But one analogy suggests that we should hold firm provided we think that true beliefs are justified when they are true *because* they were produced by a

---

[19] This point is quite familiar, as it appears in (Goldman 1979), but that doesn't mean it has been adequately handled.

[20] The example is from Conee and Feldman. As they note, if it's a known law that all maple trees are trees, Smith can hardly be *less* justified in believing it's a tree than that it's a maple tree, and it would be incredible if in *every* such case she was equally justified in the two beliefs.

[21] Since he doesn't like definitions, Fodor is not trying to use any of these properties to *define* basic property, as it might appear I'm doing here, but just trying to suggest a classification that you'll latch onto once you see it. I do hope he's at least succeeded at that.

[22] Of course in Smith's case the same information is *available* to the module in producing her panther belief and her tree belief, but unless the module is rather holistic in its operation, which is unlikely given its impressive speed, different pieces of information will be used in the two inferences.

reliable module, rather than because the believer was lucky. And if we're trying to respond to objections on behalf of the reliabilists, that's exactly what we should think.[23]

New cars start reliably. Some old cars are unreliable; they start, when they do, because of good luck rather than because they are reliable. If that's right, we can determine whether the car is reliable by determining whether its (not) starting on a given occasion is due to its (un)reliability. Now consider two old cars, René and Henri. Both of them start frequently in warm weather, and infrequently in cold weather. In Henri's case, that should say he would start infrequently in cold weather, for he lives in LA where it's almost never cold. Poor René lives in Boston where it's almost never warm. So in practice, Henri starts reliably and René doesn't. Today is rather unseasonal for February: it's warm in Boston and cold in LA. And René did start first time, but Henri had to be push-started. Did René start because he's reliable. No; it was good luck that he started, though the luck involved the weather rather than a fortuitous cold weather start. On the other hand it was just bad luck that Henri didn't start, even if it could have been predicted once the weather was known.

The lessons from these examples carry across to believers. Imagine Jones routinely forms false beliefs about nearby objects because he often sees panther shaped colour patches. In that case his object recognition is unreliable, and occasional successes are due to luck, even if on this occasion he's merely looking at a tree. The analogy here is to René starting on the unseasonably warm day. So Jones's tree belief is unjustified. Brown almost always forms true beliefs about the objects around her. Even if she's currently being tricked by a panther like sense impression, she is reliable (like Henri), and it's bad luck not unreliability that causes this error. So Brown's panther belief is justified. For similar reasons, Smith's panther belief is just as justified as her tree belief.

**4. Luck**

One day Norman develops two distinctive characteristics. He acquires a brain tumour that announces its own presence: it takes any kind of input and outputs to the central processor that there is a brain tumour (very) nearby.[24] And he becomes a clairvoyant. He somehow gets visual impressions that cause (or

---

[23] The idea that we should draw out features of our folk concept of reliability by considering cars is from (Heller 1995).

[24] (Plantinga 1988).

perhaps constitute) beliefs about the distant external world, and somehow these are usually true.[25] Both of these characteristics could, with a bit of massaging, be seen as modules, and by definition they usually produce true beliefs. But intuitively they don't produce *justified* beliefs.

One response to these examples is to deny the intuitions, and claim that their force comes from our not having reliable brain tumours or clairvoyance modules.[26] But this doesn't account for the full force of the intuitions. Imagine a blind intelligent agent who doesn't know there's such a faculty as vision. One day she miraculously acquires vision. It's hard to know what this would be like, but I'd guess that she'd involuntarily acquire some beliefs about the external world through vision, and until she could confirm these through familiar senses (like touch) she'd regard these beliefs as unjustified. And it feels that she'd be right to do so, even though *we* are perfectly familiar with sight.

The real force of these cases is that we can't get justified beliefs from modules acquired after the onset of intelligence until we have verified that these modules are reliable. As this is just what my theory says, for the reasons given above, Norman is no counterexample to my theory. But maybe there are counterexamples lurking.

Surprisingly, Lamarkism is true in Norman's world. His daughter Norma has a similar tumour and a similar 'clairvoyance module' innately. Is she justified in believing the outputs of them? Some may think not, though it's worth reflecting on whether there are epistemologically salient differences between Norma's justification for believing that she has a brain tumour or for believing (on the basis of a clairvoyant image) that the cat is on the mat and our justification for believing we have a headache or for believing (on the basis of vision) that the cat is on the mat. Even if Norma's modules regularly produce truths, if they do so by drawing on a tacitly held false theory of the world, such as any theory that says there's a reason why her clairvoyant visions are veridical, her beliefs still aren't justified. But if her clairvoyance module is innate, and it relies on a basically true theory of the world, and it usually produces truths, it seems right to say that beliefs thus generated are justified.

---

[25] (BonJour 1985).

[26] See, for example, (Goldman 1992).

**5. New Evil Demon**

Poor Morpheus. All his sensations are caused by a computer, not the expected operations of the external world. Hence all his modules are unreliable.[27] Hence he has no justified beliefs. But this seems unreasonable. For one thing, Morpheus's mundane beliefs about the locus of political power are much more reasonable than his workmate who believes every conspiracy theory that floats by. For another, we could have been in Morpheus's shoes, and if we had we'd like to think that our beliefs would have been justified.

One solution to this problem is to introduce rigidity operators into the definition of justification. A module is reliable iff it *actually* produces mostly true beliefs.[28] Because Morpheus's modules are actually reliable, they produce justified beliefs[29]. This solves the problem as stated, but doesn't get to the heart of the matter. The intuitions about justification remain even if we express the objection in contexts that are insensitive to the presence of rigidifying operators. Stephen Yablo (2002) argues that the consequent position in *If it had turned out that p, it would have turned out that q* is such a position[30], and several people have suggested that the consequent position in indicative conditionals is such a context.[31] If either claim is right, Morpheus poses a problem for rigidified reliabilists too, for both (1) and (2) are just as plausible as the claim that Morpheus's beliefs are justified.

---

[27] I set aside here semantic arguments that Morpheus really has mostly true beliefs, but, like us in some ways, he'd be quite surprised to learn the intrinsic nature of the constituents of his beliefs. See (Chalmers forthcoming) for a modern presentation of such a view. If these arguments are successful, then the problem presented in the text for my theory does not arise, so I don't tilt the playing field in my favour by this move.

[28] For a defence of this position, see Comesaña (2002), and, less directly, Sosa (1993).

[29] This requires an assessment of reliability by module types not module tokens, so I couldn't accept it even if the objection below failed.

[30] So, for example, while it is false that *If there the oceans had been filled with an atomic substance, water would have been atomic*, for in that case the fact that water is necessarily molecular implies that there would have been no water in the oceans, it is true that *If it had turned out that the oceans had been filled with an atomic substance, water would have been atomic*. I should note that Yablo doesn't unconditionally endorse the view that this context neuters hidden rigidifiers, but merely suggests that those who believe in the kind of framework Comesaña adopts should hold that it does.

[31] See, for instance, (Chalmers 2000), (Weatherson 2001), (Nolan forthcoming).

(1) If it had turned out that we were the victims of an evil demon, our beliefs would be justified.

(2) If we are the victims of an evil demon, our beliefs are justified.

Reliabilists will have to face Morpheus head on.[32] The key is to note that we *can* say that Morpheus's beliefs are *better* than his workmates' without admitting they are justified. Morpheus's non-modular beliefs are grounded in the outputs of his modules, his workmate's are not. Say that a non-modular belief is *virtuous* iff it is so grounded.[33] I claim that (a) a believer is epistemically praiseworthy to the extent her non-modular beliefs are grounded and (b) this explains our reactions to Morpheus. It seems fairly clear that if (a) is true, (b) could follow. Since Morpheus's beliefs are so disconnected from reality, it seems reasonable that we should withhold some epistemic honours from him, but it also seems he's not doing anything wrong. Saying he is praiseworthy, but his beliefs are unjustified, seems a fair compromise, if it can work.

First, note that praiseworthiness cannot track justification, despite perhaps it being *prima facie* plausible that it should do so, for a reason due to William Alston. Modules act automatically, they are not under voluntary control, so we should not praise or blame their hosts for their actions. Hence we should not praise or blame people for modular beliefs. Alston thinks this argument generalises for no beliefs are under voluntary control. But this is not true of non-modular beliefs. Consider Morpheus's conspiratorial workmate. He *could* have remained sceptical when presented with 'evidence' that lizards rule the world. In general we can *choose* to view any evidence sceptically. And making that choice blocks the formation

---

[32] Well, they will have to face it head on if they want to stay being reliabilists in the strict sense I have in mind here. I think that, for instance, Alvin Goldman's position in *Epistemology and Cognition*, which deals with the new evil demon problem by only requiring methods to be reliable in presumptively normal worlds, isn't strictly speaking reliabilism. Of course, 'reliabilism' is just a technical term, so we shouldn't quibble over its application, but I think a few too many *internalists* (as they'd traditionally be classified) would think that all and only belief forming methods that produce justified beliefs are reliable in presumptively normal worlds for this to be usefully classified with more *radically* reliabilist theories.

[33] The choice of a loaded, and familiar, term here is deliberate, for in many respects my response to Morpheus's plight is modelled on that found in Sosa's virtue epistemology, though it is not one that Sosa would endorse in all of its detail.

of some beliefs. Belief at will may be hard, but disbelief at will is much easier. That's enough voluntary control to support praise and blame. And it's the right kind of voluntary control, for we think that being appropriately credulous is praiseworthy. So that's Morpheus's position: he's praiseworthy but his beliefs float too free of the truth to be justified. If we are as good reliabilists to insist on a connection between justification and truth, this assessment sounds about right.

**6. Conclusion**

None of the three major problems facing (R) are difficulties for my theory. There's certainly more that could be done with the theory. For instance, I could (a) defend the Fodorian theory on which it's based against numerous objections, (b) argue that we need a new theory of justification because the extant theories all face serious problems, (c) argue for my assessments of some cases that play a central role in the above defence, and (d) say more about support. But Fodor looks to be doing fine with (a) without me, the community of epistemologists collectively does (b) pretty well, and as long as our intuitions are calibrated (c) will be unnecessary. (As acknowledged, there's much to be done with (d).) Given all that, it seems worthwhile to present a new and invulnerable to extant objections theory of justification, as I've done.

*References*


Alston, William (1988) "The Deontological Conception of Epistemic Justification" *Philosophical Perspectives* 2: 257-99.

BonJour, Laurence (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

BonJour, Laurence (2000) "Toward a Defense of Empirical Foundationalism," in Michael DePaul (ed.), *Resurrecting Old Fashioned Foundationalism*. Lanham, MD: Rowman & Littlefield.

Carruthers, Peter (2003) "Moderately Massive Modularity" in A.O'Hear (ed.), *Mind and Persons*. Cambridge: Cambridge University Press.

Chalmers, David (2000) "The Tyranny of the Subjunctive" unpublished.

Chalmers, David (forthcoming) "Envatment as a Metaphysical Hypothesis" paper presented at *Metaphysical Mayhem* VI, Syracuse, NY, August 2002.

Cohen, Stewart (1984) "Justification and Truth" *Philosophical Studies* 46: 279-96.

Comesaña, Juan (2002) "The Diagonal and the Demon" *Philosophical Studies* 110: 249-66.

Conee, Earl and Richard Feldman (1985) "Evidentialism" *Philosophical Studies* 48: 15-34.

Conee, Earl and Richard Feldman (1998) "The Generality Problem for Reliabilism" *Philosophical Studies* 89: 1-29.

Fodor, Jerry A. (1983) *Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, Jerry A. (2000) *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press.

Goldman, Alvin (1979) "What is Justified Belief" in G. Pappas (ed.) *Justification and Knowledge* Dordrecht: Kluwer.

Goldman, Alvin (1986) *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.

Goldman, Alvin (1992) *Liaisons*. Cambridge, MA: MIT Press.

Heller, Mark (1995) "The Simple Solution to the Generality Problem" *Noûs* 29: 501-15.

Landsburg, Stephen (2002) "One Small Step For Man…" *Slate* August 29, 2002.

Nolan, Daniel (forthcoming) "A Possible Worlds Theory of Conditionals"

Plantinga, Alvin (1988) "Positive Epistemic Status and Proper Function" *Philosophical Perspectives* 2: 1-50.

Sosa, Ernest (1991) *Knowledge in Perspective*. Cambridge: Cambridge University Press.

Sosa, Ernest (1993) "Proper Functionalism and Virtue Epistemology" *Noûs* 27: 51-65.

Stanley, Jason (2000) "Context and Logical Form" *Linguistics and Philosophy* 23: 391-434.

Stanley, Jason (forthcoming) "On the Case for Contextualism"

Vogel, Jonathan (1990) "Cartesian Scepticism and Inference to the Best Explanation" *Journal of Philosophy* 90: 658-66.

Weatherson, Brian (2001) "Indicative and Subjunctive Conditionals" *Philosophical Quarterly* 51: 200-16.