

## XIII\*—HYPER-RELIABILITY AND APRIORITY

by James Pryor

ABSTRACT I argue that beliefs that are true whenever held—like *I exist, I am thinking about myself*, and (in an object-dependent framework) *Jack = Jack*—needn't on that account be a priori. It does however seem possible to remove the existential commitment from the last example, to get a belief that *is* knowable a priori. I discuss some difficulties concerning how to do that.

### I

A thought-type is *hyper-reliable* just in case necessarily, whoever thinks a thought of that type thereby thinks a true thought. Many of our a priori justified beliefs are hyper-reliable; my question is whether hyper-reliable thoughts must always be a priori justified.

Put like that, the answer should surely be 'No'. Take any mathematical truth *P* for which you currently lack justification; the thought that *P* will be hyper-reliable, but it's not one for which you have *any* justification, much less a priori justification. In that case, however, you don't merely lack justification to believe *P*; you also lack justification to believe your thought-type is hyper-reliable. (You'd only be in a position to know it's hyper-reliable *if true*.) More interesting are cases where you *do* know in advance—and know a priori—that a given thought-type is hyper-reliable. I will argue that even *those* hyper-reliable thoughts can fail to be a priori justified.

I'll specify thought-types using sentences. By doing so, I'll mean the (most specific) type of thought you'd occurrently have *in* or *by* rehearsing the sentence to yourself.<sup>1</sup> Here are

1. I help myself to the fact that you understand the sentence; this procedure will only pick out the thoughts I intend when that's so. My notion of thinking a thought *in* rehearsing a sentence is an intimate one. If you rehearse to yourself 'Mount Everest is 8850 metres tall' and simultaneously think metalinguistic thoughts *about* that sentence, or think about Hillary's cresting the summit, these further thoughts don't count as thoughts you have *in* rehearsing 'Mount Everest is 8850 metres tall'.

I *don't* assume that all thinking proceeds by rehearsing sentences. I'm only assuming that rehearsing sentences to yourself is one way of having occurrent thoughts. Those very same thoughts may (or may not) be thinkable otherwise than by rehearsing sentences.

\*Meeting of the Aristotelian Society, held in Stewart House, University of London, on Monday 22 May 2006 at 4.15 pm.

some examples:

- (1) I exist.<sup>2</sup>
- (2) I am thinking (i.e., occurrently entertaining thoughts).
- (3) I am thinking about myself.
- (4) Jack exists.
- (5) Jack is self-identical.
- (6)  $Jack = me \vee Jack \neq me$ .

Consider (1)–(3) first. Understanding these thought-types should put you in a position to know a priori that anyone who (occurrently) thinks them must thereby be thinking truly. Nonetheless, I will argue that your *justification* to believe any such thought must be introspective; and I count introspection as an *empirical* source of justification. Hence, I claim these thoughts can only be justified a posteriori. You can know a priori that *anyone who has* a thought of that type is thinking truly; but your justification to believe *you're* thinking a thought of that type, and your justification to believe that particular thought, are only a posteriori. So I will argue.

For (4)–(6), let's adopt a McDowell/Evans-style view<sup>3</sup> on which these thoughts are object-dependent; so when Jack doesn't exist, no thought of the relevant type is available to be had. This view is quite controversial, and I'm not myself attracted to it. However, my aim here is hypothetical. I want to discuss what justification should be available to us a priori, should that McDowell/Evans-style view be correct. For on that view it looks like thoughts of form (4)–(6) will be hyper-reliable. Whoever successfully has such a thought is thinking truly. Moreover, if we can know a priori that the McDowell/Evans-style view is correct (at least about these thought-types), then the thoughts will not merely *be* hyper-reliable; we can *know them a priori* to

2. To keep our discussion manageable, I assume that the thought-type 'I exist' differs from the thought-type 'Jim Pryor exists'. Some philosophers will welcome this because they regard my thoughts as having *different contents*. Other philosophers will assign my thoughts *the same content*; but they'll allow that in each case I grasp that content 'in a different way'. These philosophers can understand me as taking the way in which one grasps a content to be essential to the most specific type of thought one is having. In other words, I'm individuating types of thinking by more than just the thought's content.

3. See McDowell 1977, 1984 and 1986; and Evans 1981 and 1982.

be hyper-reliable. Nonetheless, I will argue that in these cases too your justification to believe *you're* thinking a thought of that type, and your justification to believe that particular thought, are only a posteriori.

## II

I'll begin by clarifying the notion of apriority that I'm employing.

I understand *a priori justification* to be justification that doesn't derive from occurrent experience. By 'occurrent experience' I include perceptual experience and I also include occurrent mental states and processes. On some views, occurrently having a mental state is enough to justify you in believing you do; on others, your justification comes from an *introspective awareness* of your mental state. In neither case, though, do I count the justification as a priori.<sup>4</sup> To be sure, there are important differences between perception and your awareness of your own occurrent mental life. But it seems more natural to group the kind of justification I have to believe that I have a headache, and that I'm thinking of monkeys, together with the justification I have to believe that I'm awake, and seeing a table—rather than with the justification I have to believe that there are prime numbers, and that whoever thinks must exist.

When thinking about a priori justification, we must be careful to distinguish (i) beliefs where one's justification genuinely does derive from experience from (ii) beliefs which are merely such that certain experiences are necessary *for you to have* the belief, though they are not what justify you in believing it. For instance, it may be argued that the concept CHROMATIC is unavailable to subjects who've never had visual experiences. Hence, merely in order to *have* the belief that chromatic colours are colours, one would need to have visual experiences. Plausibly, though,

4. Bonjour 1998, pp. 7ff., also counts introspective justification as a posteriori. Contrast Kitcher 1980, §V, and many of the essays in Ludlow and Martin 1998.

On some views, we can have knowledge of our own mental states *otherwise* than through our occurrent experience or introspective awareness of them. These views may complain that self-knowledge as *they* conceive it is not well thought of as a posteriori. That complaint may be fair; but I'll have to leave this issue for another time.

those experiences would play no role in justifying the belief. Your justification for it is still *a priori*.<sup>5</sup>

I'll argue later that some thoughts modelled after (5) and (6)—though not (5) and (6) themselves—have the same status. Experiences are needed to have the thoughts; but you have justification for the thoughts which the experiences are no part of.

We don't want to swing to the opposite extreme, though, and assume that the experiences you need to entertain a thought will *always* be epistemically irrelevant. I've said that in some cases they do only play a role in enabling you to entertain the thought, and not in justifying you in believing it. But in other cases they will do *both*. They will both enable you to entertain the thought and *also* give you justification to believe it. Here are some examples.

You look out the window and see a canary, and think 'That canary exists'. You hear someone's cellphone ring during a movie, and think 'That phone is ringing'. Stumbling about in a dark room, you place out your hand and find a wall; you think 'This wall is perceived by me'. In each of these cases you require certain kinds of experience just to be able to form the demonstrative thought. And in each case the experiences that enable you form the thought also give you some *a posteriori* justification to believe it. Your visual experiences of the canary both enable you to think the demonstrative thought 'That canary exists' and justify you in believing it. And so on. Your justification for these beliefs seems to me to come from the same spring as the perceptual justification you'd have in these cases to believe 'Some canary exists' and 'Some phone is ringing'. This latter justification is clearly *a posteriori*. So I count your justification to believe 'That canary exists' and 'That phone is ringing' as *a posteriori* too.

Hence the experiences that you need in order to think a given thought will in some cases be part of what justifies you in believing the thought and in other cases won't.<sup>6</sup>

5. This is a familiar point. See Kitcher 1980, pp. 4–5; Burge 1993, p. 460; Plantinga 1993, pp. 103–4; BonJour 1995, p. 53; and BonJour 1998, pp. 9–10.

6. This is a much less familiar point. But see BonJour 1998, p. 10; Salmon 1987/8, esp. pp. 201–3; Salmon 1993a, pp. 130–1; Salmon 1993b, pp. 92–4; Soames 1995, pp. 264–5; and Soames 1997, pp. 86–8.

These canary-type cases are ones where in order to form the relevant thought you need *antecedently to have* some a posteriori justification to believe it, and that justification comprises an essential part of what *does* justify your belief when you *have* formed it. A related, but different, kind of case is one where occurrently having a thought (or occurrently being aware of having the thought) *gives you* some introspective justification; and that's all the a posteriori justification you need to be justified in believing the thought. Here are two examples. You entertain the hypothesis 'Some hypotheses irritate me', and in doing so find yourself irritated by it. Or you have an orange after-image, and while attending to it think 'I'm thinking about this after-image'. In these cases, as in the canary case, someone who has the thought may not need *any more* a posteriori justification to have justification to believe it.

We may dispute which category a given case falls under: for example, when you think 'That canary exists', I think that perceptual justification to believe there's a canary there, which you're seeing, needs to be in place *for* you to be able to understandingly think the demonstrative thought. Others will argue that you can be justified in believing 'That canary exists' solely from your introspective awareness of seeming to successfully think it, and your a priori knowledge that demonstrative thoughts need referents. However, I think we can agree that the examples I've cited each fall into one or the other of these categories, and so draw their justification from *some* empirical source. Hence, none of them is a case of a priori justification as I'm understanding that notion.

I claim that (1)–(6) too are at best of that type.

### III

We'll begin with the cogito-type thoughts, (1)–(3).<sup>7</sup> Here there are two lines of reasoning to examine.

On the first line of reasoning, my justification to believe the cogito-type thought derives from *my justification to believe* that thoughts of that type must be true whenever thought. This line

7. Audi 1999, pp. 212–13, also argues that our justification for cogito-type thoughts is a posteriori.

of reasoning is not very promising. For even if I know a priori that thoughts of type *T* are true whenever thought, that won't justify me in believing *my* thought of type *T* unless I'm justified in believing I'm thinking it. And my justification to believe that will be introspective, and so a posteriori.

Compare the following. The thought-type

(7) (If Jack exists) I am distinct from Jack

is plausibly true whenever entertained by someone distinct from Jack. (We will discuss the existence 'hedge' later.) But the mere fact that a thought of that type is entertained by someone distinct from Jack—say by me—isn't enough to make the subject justified in believing it. I may have no idea that I'm distinct from Jack.<sup>8</sup> I only know that when a thought of that type is entertained by a non-Jack, it's true. For that to help justify me in believing it, I'd have to have justification to believe *I am* a non-Jack entertaining it. Why shouldn't the same be true for thought-types that are true whenever entertained by *anyone*? For that to help justify me in believing the thought, wouldn't I need justification to believe *I am someone* entertaining the thought? And wouldn't that justification need to be introspective, as I've said?

There's a second line of reasoning here, though, that's worth examining more closely. On this line of reasoning, one's justification for believing the cogito-type thought is *not* supposed to be inferential. Rather, *the mere fact* that a thought-type is hyper-reliable is supposed to directly give one justification to believe it.<sup>9</sup>

One problem with this proposal comes from examples of unknown (or empirically known) mathematical truths—I mentioned these earlier. Even if that problem can be finessed,

8. Some philosophers claim that my thought 'I am identical to Jim' has the same content as my thought 'Jim is identical to Jim', and that when I grasp that content in the latter way I do have a priori justification to believe it. So they might claim I am in a position to know a priori that I am identical to Jim. Two comments. First, I think apriority attaches to thought-types; and the thoughts 'I am identical to Jim' and 'Jim is identical to Jim' will be of different types even if they have the same content. But second, I've avoided this issue in any case by talking about the thought 'I am distinct from Jack'. As far as I can see, there's no parallel argument that that thought has a content that's knowable a priori.

9. A variant might say that *our merely appreciating* that a thought-type is hyper-reliable directly gives us justification to believe the thought—not by serving as a premiss from which we can infer anything. This variant faces the same objections I give in the text.

several further examples work together to make the proposal unconvincing.

First, consider a practical analogy. The fact that, should you adopt *X* as an end, you'll achieve it, is by itself no reason for adopting end *X*. (Of course, achieving ends you've adopted may tend to have happy consequences, like lifting your mood; if so, then adopting end *X* may thereby acquire some instrumental merit.) If someone said to you, 'The desire to have desires is an especially happy desire, because it's self-fulfilling. The desire to lack desires, on the other hand, is an especially unhappy desire, because it's self-thwarting. So you should have the desire to have desires,' why can't you respond, 'Why? I neither desire to have desires, nor desire to lack them. How does the fact that the desire to have desires is self-fulfilling give me a reason to have it?' Similarly, the fact that should you believe a thought of type *T* your belief will be true is by itself no reason to believe a thought of type *T*.

Some will protest: 'But there's a difference! Having fulfilled desires isn't intrinsically valuable, true, but having true beliefs is.'

It's not clear that having true beliefs is intrinsically valuable. It may be intrinsically *better* to believe the true *than* to believe the false, but what intrinsic demand is there to have a belief at all, rather than withhold judgement? In any case, even if having true beliefs is intrinsically valuable, that would only provide a practical motive to believe hyper-reliable thoughts. It wouldn't be an epistemic reason to believe those thoughts.

A second analogy. Suppose Jiho is especially stubborn. To date, he's never yet changed his mind once he's settled it. In fact, he believes he will never change his mind. You say to Jiho, 'Look Jiho, consider what would happen if you came to believe that you will someday change your mind about something. That would itself constitute a change of mind; hence if you came to believe it, you'd thereby make it true.' Jiho says, 'OK.' You say, 'So don't you see? That's an epistemic reason to believe it. That's a reason for you to think you will change your mind.' Jiho would look at you bemused. And rightly so.

Some will protest that I'm not considering the best form of my opponent's idea. They'll say: 'True, we shouldn't think that a thought-type *T*'s merely being hyper-reliable gives us justification *to* form a belief of type *T*, regardless of whether we do in fact

believe it, or have even thought it. But in a situation where *we have* a thought or belief of type *T*, its mere hyper-reliability *does* then constitute some justification for the belief. Or alternatively: the belief's being *hyper-reliable and thought by you* is what directly gives you (at least some) justification for believing it.'

When we count the fact that I'm thinking a thought among the considerations that justify me in believing it, it's starting to sound like my justification is partly introspective. As I said earlier, some theories of introspective justification will say it's my merely *having* a given mental state—my *experiencing* or *undergoing* a given mental process—that gives me justification to believe I am. The current proposal sounds more like introspective justification of that sort than it does a priori justification. However, the line between these can be difficult to discern.

My third example will help us decide what to say. It weighs against both of the views I'm now opposing: neither mere hyper-reliability, nor hyper-reliability by a thought you actually have, will make a thought a priori justified. Consider this sentence:

(8) I am uttering a sentence.

Given what this sentence means, it follows that whenever it's used to think a thought, that thought is true. (I count rehearsing a sentence to yourself privately as a kind of utterance.) And anyone who understands the sentence is in a position to know this. However, suppose you do utter the sentence (either privately or aloud). What then justifies you in believing that you are uttering it, or any sentence? It can't be your understanding of the sentence. That would only justify you in having beliefs about what's true *whenever the sentence is uttered*. It doesn't help you determine when that condition is fulfilled. The natural thing to say is that what justifies you in believing you are uttering the sentence is your introspective or perceptual awareness of uttering it. Hence, your justification for believing the thought you have by rehearsing (8) is a posteriori—despite the fact that you know, just by virtue of understanding (8), that whenever it's used to think a thought, that thought is true.

I think the same holds for sentences like 'I exist', 'I am thinking', and so on. Your understanding of these sentences is enough to justify you in believing that whenever they're used to think thoughts those thoughts are true. *That* justification may

be a priori. But your justification to believe *the thought you have* by rehearsing ‘I exist’ is not itself a priori. What justifies you in believing that you exist would be your experience or introspective awareness of your own thought processes—including perhaps the process of entertaining the thought that you exist.

#### IV

Let’s turn now to case (4), where you have an object-dependent thought in rehearsing to yourself ‘Jack exists’. As we’ll see later, I have no *general* objection to the idea that object-dependent thoughts might be justified a priori. My qualms about this case are based entirely on the fact that it seems counter-intuitive that you could know of Jack’s existence a priori. The thought ‘Jack exists’ seems to group more naturally with the canary-type thoughts I mentioned in §II. You need experiences of certain sorts to be able to entertain these thoughts, and those experiences seem intuitively to constitute part of what justifies you in believing the thought.

I acknowledge this may be reasonably disputed. And on other construals of what apriority amounts to, it may be less strange to claim that you can have ‘a priori’ justification to believe that Jack exists. But on the construal I’m operating with, the most natural view is surely that you don’t have a priori justification to believe that Jack exists. I want to explore how far we can pursue this natural thought.

One problem with counting (4) as a posteriori is that, on the one hand, (5) and (6) look like *much better* candidates to be justifiable a priori. They seem like claims that you can be justified in believing just through your understanding of identity, distinctness, disjunction, and so on. On the other hand, though, (5) and (6) seem to entail (4). So if we allow them to be a priori, that puts pressure on us to count (4) as a priori too. It’s very puzzling what we should say. As Stephen Schiffer pointed out to me, we’re also intuitively tempted to assign degree of belief *one* to (5) and (6), but a degree of belief of *less than one* to (4). Yet this is hard to sustain if (5) and (6) do entail (4).

There’s a lot to sort out here. Some epistemologists will deny that having a priori justification to believe one thing guarantees you’ll thereby have a priori justification to believe what it

entails.<sup>10</sup> Some free logics will deny that (5) and (6) *do* entail (4). We cannot explore all of these options. I will just present my own preferred way of sorting things out.

I want to count (4) as a posteriori. An opposing line of thought says: ‘Logic is our paradigm of what’s knowable a priori. And the sentence “Jack exists” (along with the other examples you’re considering) is a logical truth. So—at least when you do successfully have a thought by rehearsing “Jack exists” to yourself—that thought should be knowable a priori.’

We might resist this in different ways. We might just say: well, (4) through (6) must be *counter-examples* to the claim that logical truths are always knowable a priori. Or we might move to a free logic on which ‘Jack exists’ is no longer a logical truth. I think doing that will be illuminating.

I propose a *neutral free logic*: that is, a free logic on which atomic sentences containing non-referring terms are truth-valueless. The sentential logic for such systems is a weak Kleene logic: the connectives work like classical connectives when their operands have classical truth-values, and yield no truth-value when any of their operands are truth-valueless. On these systems, it’s standard for quantifiers to be bivalent. For example,  $\exists x: \phi$  is counted true on an interpretation  $I$  and assignment  $V$  if there’s any assignment  $V^*$  (differing from  $V$  at most with respect to what it assigns to  $x$ ) such that  $\phi$  is true on  $I$  and  $V^*$ ; *otherwise* (including the case when  $\phi$  is truth-valueless for every  $I$  and  $V^*$ )  $\exists x: \phi$  is counted false on  $I$  and  $V$ . ‘ $\alpha$  exists’ is usually defined as  $\exists x: x = \alpha$  and so will be false when  $\alpha$  does not refer.

One can define various notions of logical consequence for multi-valued logics. I will understand logical consequence to be a matter of never going from truth to non-truth. That is:  $\{A, \dots\} \models B$  means there is no interpretation  $I$  and assignment  $V$  on which every sentence in  $\{A, \dots\}$  is true but  $B$  is false or truth-valueless. Logical truths will be sentences that are consequences of every other sentence; this requires them to be true on every interpretation and assignment.

(Of course, these semantics will need to be complexified to handle the indexical ‘me’ in sentence (6). I’ll take those complexities as understood.)

10. See Wright 2000 and 2003; and Davies 1998, 2000 and 2003.

I propose that it's only the logical truths of a system *like this* that we should expect to be knowable a priori.<sup>11</sup>

My argument for this is just that it gives us the intuitively desired results. Sentence (4), 'Jack exists', is not a logical truth of this system: it's false on interpretations where 'Jack' doesn't refer. (4) *is* a logical consequence in this system of (5) and of (6); however, neither of those is a logical truth either.

As with the cogito-type thoughts we examined earlier, you may know a priori that whoever (successfully) thinks a thought of types (4) through (6) thereby thinks truly. But as I've already argued, that doesn't mean the thoughts are justified a priori when you do think them.

## V

I'm not the first philosopher to shy away from letting thoughts of type (5) and (6) be justified a priori because of their existential commitment. A standard move in the literature is to retreat to '*hedged*' versions of these thoughts, as follows:

- (5\*) (If Jack exists) Jack is self-identical.
- (6\*) (If Jack exists) Jack = me  $\vee$  Jack  $\neq$  me.

The hope is that these thoughts, unlike (5) and (6), will *not* entail that Jack exists, and hence can safely be justified or known a priori.

This move demands more delicacy than its proponents tend to acknowledge. It requires care in our choice of a free logic, and in the interpretation of (5\*) and (6\*). On the neutral free logic I proposed, for example, the connective  $\supset$  yields no truth-value if either of its operands is truth-valueless. So if (5\*) is read as

- (5\* $\supset$ ) Jack exists  $\supset$  Jack is self-identical

then when 'Jack' refers, (5\* $\supset$ ) will be true; and when 'Jack' does not refer, the consequent of (5\* $\supset$ ), and so (5\* $\supset$ ) itself, will be

11. I'm *not* claiming that nothing is knowable a priori except the logical truths of this system. I'm claiming that we shouldn't expect being a logical truth of another system, like classical logic, *to make for* a priori knowability. It's compatible with that that other things might make for a priori knowability, and hence that there can be non-logical a priori truths.

truth-valueless. These are precisely the same truth-values that (5) has in the respective cases. In other words,  $(5^* \supset)$  and (5) never diverge in truth-value. So we won't find one of them having logical consequences that the other lacks.

What philosophers who make the 'hedging' move seem to be after is an operation on (5) that will yield truth when 'Jack' does not refer, and the regular truth-value of (5) when 'Jack' does refer. Nothing stops us from introducing an operation of that sort. We can do it by brute force. Or, recalling that it's standard in these systems for quantifiers to be bivalent, we can observe that  $\neg\exists x:\phi$  is counted *false* on an interpretation  $I$  and assignment  $V$  if there's any assignment  $V^*$  (differing from  $V$  at most with respect to what it assigns to  $x$ ) such that  $\phi$  is true on  $I$  and  $V^*$ ; *otherwise* (including the case when  $\phi$  is truth-valueless for every  $I$  and  $V^*$ )  $\neg\exists x:\phi$  is counted *true* on  $I$  and  $V$ . Now we can interpret  $(5^*)$  like this:

$$(5^* \neg\exists) \quad \neg\exists x: (x = \text{Jack and NOT: Jack is self-identical}).$$

When 'Jack' doesn't refer, the embedded expression ' $x = \text{Jack and NOT: Jack is self-identical}$ ' will be truth-valueless; hence  $(5^* \neg\exists)$  will be true.

There are other techniques that will deliver the same result. The important point is that we're not treating the 'hedge' in  $(5^*)$  as an ordinary material conditional.<sup>12</sup>

If we understand  $(5^*)$  and  $(6^*)$  in the way I'm proposing, then *they will be* logical truths in the system of §IV; and I think we *should* regard them as thought-types for which one can have a priori justification. There is no danger of that a priori justification leaking through to (4) 'Jack exists', because (4) is not a logical consequence of  $(5^*)$  or  $(6^*)$ .

## VI

Donnellan made the 'hedging' move in 'The Contingent A Priori and Rigid Designators'. Evans then complained that in the object-dependent framework Donnellan was employing, the

12. On a 'negative' free logic, on which atomic sentences containing non-referring terms are always false, the 'hedge' in  $(5^*)$  can be interpreted as a material conditional.

‘hedging’ move doesn’t do anything to help.<sup>13</sup> The thought you have by rehearsing (5\*) to yourself is still an object-dependent thought about Jack. Evans says that to avoid existential commitment, we need to move to a free logic *and also* to move away from object-dependent thoughts.

Apparently, then, he thought that so long as we *continue* dealing with object-dependent thoughts—no matter how they’re hedged—allowing the thoughts to be a priori will make correlative existential claims a priori too. And even if Evans didn’t think that, you might. So I’ll try to persuade you otherwise. I’ll try to spell out and defend my proposal to count (5\*) and (6\*) as a priori, even though we withhold that status from (4) and the ‘unhedged’ (5) and (6).

On the object-dependent framework we’re working with, when ‘Jack’ doesn’t refer, you don’t think any contentful thought *in* rehearsing ‘Jack exists’, You may think that you *do* have some thought in rehearsing that sentence. And you may think other thoughts; for example, you may think that someone *named* ‘Jack’ exists. But these won’t be thoughts you have *in* rehearsing ‘Jack exists’ to yourself. On our present framework, nothing will be.

So let’s restrict our attention to cases where the subject’s use of ‘Jack’ *does* in fact refer. However, we *don’t* assume that the subject knows yet that this is the case. The subject may know a priori that, in rehearsing ‘Jack exists’ to himself, he’s either having a true thought or no thought at all. But he won’t know a priori which of those alternatives obtains.

There are different things we can say about how a subject *gets* justification to believe he’s in the one situation rather than the other. One view says that there’s an a priori *presumption* in favour of one’s seeming-thoughts being genuine thoughts. A subject would then be in a position, through introspection, to believe he really is thinking a thought by rehearsing ‘Jack exists’. He may even be in a position to conclude, on the basis of what he knows a priori and through introspection, that Jack does exist. A different view says that the subject needs ordinary *perceptual* justification to believe that Jack exists before he’ll be

13. See Evans 1979, pp. 195–6; and Oppy 1994, n. 14. Ray 1994 is a good critical discussion.

justified in taking his experience of seeming to think a thought by rehearsing 'Jack exists' to be a genuine thinking. Various more complicated views are also possible.<sup>14</sup>

No matter which of these views we adopt, we can agree that a subject won't have *a priori* justification to believe he's thinking a contentful thought, when he does rehearse 'Jack exists'. At best his justification will be partly introspective.

It's compatible with that, though, that the subject *is* in fact thinking a thought. And we can inquire into what justification the subject has for the thought. I see no obstacle to allowing this justification sometimes to be *a priori*. That is: a subject can sometimes in fact be in a position to think a certain type of thought, and *have a priori justification* for that thought, without being in a position to tell *a priori* that all this is so.

This is what seems to me to be going on with thoughts (5\*) and (6\*). If you are in fact able to entertain thoughts of those types, your understanding of identity, disjunction, and so on, seem to be enough to justify you in believing them. Those are *a priori* sources of justification. Unlike (5) and (6), these thoughts don't *logically entail*, in the relevant system, that Jack exists. It's just that, as with (5) and (6), you wouldn't be able to have the thoughts unless Jack did exist.

Summing up, on the view I'm proposing:

You have a *a priori* justification to believe that each of these thought-types—(4) through (6), and (5\*) and (6\*)—is hyper-reliable, and can only successfully be had when Jack exists.

But that doesn't make the thoughts justified *a priori*. Some of the thoughts *are* justified *a priori*; but the case for that has to be made independently.

Your *a priori* justification for those thoughts sits alongside a *lack* of a *a priori* justification to believe you're successfully thinking them. From the premiss *that* you're thinking the thoughts, you could infer *a priori* that Jack exists; but from the thoughts themselves, you cannot. It takes more

14. This takes us into the literature surrounding 'McKinsey's argument'. For references and my own complicated view about what to say, see Pryor forthcoming.

to justify you in believing *you have* the thoughts than it takes to justify the thoughts themselves.<sup>15</sup>

Experiences are necessary to be able to have *any* of the thoughts. In some of these cases, like the thought ‘Jack exists’, the experiences play a justifying role. In others, like (5\*) and (6\*), they do not.

## VII

What will all of this seem like from the inside?

Well, on the one hand, you may have no view about whether you’re in a position to think any thought by rehearsing (5\*). Or you may just (truly) take it for granted that you are, without having any justification for believing so. It takes a good amount of conceptual and philosophical sophistication to think about one’s sentences, and whether one manages to think a thought in rehearsing them. One can certainly use sentences to think thoughts like (5\*) long before one has achieved that level of sophistication. And one can also have the understanding of self-identity that seems to me to justify belief in (5\*) before one has achieved that level of sophistication. So I think it is possible to be justified in believing (5\*) even if one has no view about, and lacks any justification for believing, the claim that in rehearsing (5\*) one is successfully thinking a thought.

But now, on the other hand, suppose you’re rehearsing (5\*) and you’re *unsure whether* you’re thereby successfully thinking a thought. I’m still assuming that as a matter of fact Jack does exist and you are successfully thinking. But you don’t know this to be the case. You may even be all-things-considered justified in believing it’s *not* the case.<sup>16</sup> In such situations, it may seem

15. Here I must part company with Davies 2000. As I interpret his view (see Pryor 2004), it requires that to be justified in believing *P*, one have antecedent entitlement to believe *there is* such a belief as *P*. Hence, if—as I believe—we can only have a *posteriori* justification that there is a belief to be had by rehearsing (5\*), it’d follow that in order to be justified in believing (5\*), we’d need antecedent a posteriori justification to believe something else. That would preclude us from ever having a priori justification to believe (5\*).

16. Perhaps perversely, I think you can be in a position of successfully thinking ‘Jack does *not* exist’, have all-things-considered *justification* to believe that thought, but be justified in believing that your attempts to think it are unsuccessful. What’s more,

peculiar to go ahead and *try to form* a thought of type (5\*), since you know you won't succeed unless Jack does exist. However, you don't really have a good alternative. Should you be *agnostic* towards (5\*)? If agnosticism is assigning a middling degree of belief to a thought, then your ability to have *that* attitude towards (5\*) is just as threatened by the prospect of Jack's not existing as full belief towards (5\*) would be. There *is* such a state as *having no degree of belief at all* towards a thought. That may be the state you're in if you've never considered the thought, or are conceptually incapable of thinking it. But I don't think you will occupy that state in the cases we're considering. In these cases, you do in fact succeed in considering the thought (5\*); you just don't know a priori that you do. So you will in fact have some degree of belief towards (5\*).

Where are we then? You suspect that Jack may not exist, and that in rehearsing (5\*) you may not be thinking any thought. But you are as a matter of fact thinking (5\*), and will as a matter of fact have some degree of belief towards it. On my free logical semantics for (5\*), it doesn't entail (4) 'Jack exists'. So in, say, believing (5\*) to degree 1 and (4) and (5) only to degree  $\frac{1}{4}$ , you won't *contradict* yourself. Still, it remains somewhat puzzling what it amounts to, to *think* or *believe* these free logical sentences.

This may help. Let Sue be thinking a 'hedged' thought of the form '(If Jack exists) Jack is *F*', and 'believe' it to degree  $\frac{2}{3}$ . In classical terms, that'd be like her taking a doxastic stance of *being prepared* to attribute *F*-ness to Jack, on the assumption that Jack exists. That is, she's *ready* to believe (in the classical sense) to degree  $\frac{2}{3}$  that Jack is *F*. But she's not committing herself to Jack's actually existing. In the object-dependent framework we're working in, it's not possible for Sue to take even this 'hedged' attitude towards Jack when Jack doesn't exist. But she can go through the motions. When Jack *does* exist, and she is in a position to think about him, she'll thereby have succeeded in taking the 'hedged' attitude—without yet knowing that she has. What 'hedging' gets her is not an attitude to have even when Jack fails to exist. What it gets her is the ability, when Jack

you can know a priori that if your attempts to think it *were* successful, they'd have to be false.

does in fact exist, to avoid *committing herself* prematurely about whether he does.

My 'translation' of Sue's attitude into classical terms is just a heuristic. What we really need to do is have degrees of belief defined towards the free logical sentences, themselves—or, better, to some interpreted structures that are suited to play the proposition role for them. And we should understand justification to be a matter of having epistemic reason *to* assign degrees of belief in that way. Clearly there's much work yet to be done here. I hope my suggestions will put us on the right track.<sup>17</sup>

*Department of Philosophy*  
*NYU*  
*New York, NY 10003*  
*USA*

#### REFERENCES

- Audi, R. 1999: 'Self-evidence'. *Philosophical Perspectives*, 13, pp. 205–28.
- BonJour, L. 1995: 'Towards a Moderate Rationalism'. *Philosophical Topics*, 23, pp. 47–78.
- BonJour, L. 1998: *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Burge, T. 1993: 'Content Preservation'. *Philosophical Review*, 102, pp. 457–88.
- Davies, M. 1998: 'Externalism, Architecturalism, and Epistemic Warrant'. In C. Wright, B. Smith and C. Macdonald (eds), *Knowing Our Own Minds: Essays in Self-Knowledge*. Oxford: Oxford University Press.
- Davies, M. 2000: 'Externalism and Armchair Knowledge'. In P. Boghossian and C. Peacocke (eds), *New Essays on the A Priori*. Oxford: Oxford University Press.
- Davies, M. 2003: 'The Problem of Armchair Knowledge'. In S. Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge*. Cambridge, MA: MIT Press.
- Donnellan, K. 1977: 'The Contingent A Priori and Rigid Designators'. *Midwest Studies in Philosophy*, 2, pp. 12–27.
- Evans, G. 1979: 'Reference and Contingency'. *Monist*, 62, pp. 161–89. Reprinted in Evans 1985.
- Evans, G. 1981: 'Understanding Demonstratives'. In H. Parret and J. Bouveresse (eds), *Meaning and Understanding*. New York: Walter de Gruyter. Reprinted in Evans 1985.
- Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.
- Evans, G. 1985: *Collected Papers*. Oxford: Oxford University Press.
- Hawthorne, J. 2002: 'Deeply Contingent A Priori Knowledge'. *Philosophy and Phenomenological Research*, 65, pp. 247–69.
- Kitcher, P. 1980: 'A Priori Knowledge'. *Philosophical Review*, 89, pp. 3–23.

17. Thanks are due to Mark Schroeder and members of my spring 2003 grad seminar at Princeton; to Hartry Field and Stephen Schiffer; to audiences at Bigorio, Syracuse, Kingston, Barcelona, Kentucky, and Vermont; and to the many friends who've given me feedback on the larger project from which this is drawn.

- Ludlow, P. and N. Martin (eds) 1998: *Externalism and Self-Knowledge*. Stanford, CA: CSLI Publications.
- McDowell, J. 1977: 'On the Sense and Reference of a Proper Name', *Mind*, 86, pp. 159–85. Reprinted in McDowell 1998.
- McDowell, J. 1984: 'De re Senses'. In C. Wright (ed.), *Frege: Tradition and Influence*. Oxford: Blackwell, pp. 98–109. Reprinted in McDowell 1998.
- McDowell, J. 1986: 'Singular Thought and the Extent of Inner Space'. In P. Pettit and J. McDowell (eds), *Subject, Thought and Context*. Oxford: Oxford University Press, pp. 137–68. Reprinted in McDowell 1998.
- McDowell, J. 1998: *Meaning, Knowledge and Reality*. Cambridge, MA: Harvard University Press.
- Oppy, G. 1994: 'Salmon on the Contingent A Priori and the Necessary A Posteriori'. *Philosophical Studies*, 73, pp. 5–33.
- Plantinga, A. 1993: *Warrant and Proper Function*. New York: Oxford University Press.
- Pryor, J. 2004: 'What's Wrong With Moore's Argument?'. *Philosophical Issues*, 14, pp. 349–78.
- Pryor, J. forthcoming: 'Externalism About Content and McKinsey-style Reasoning'. To appear in Sanford Goldberg (ed.), *Internalism and Externalism in Semantics and Epistemology*, Oxford: Oxford University Press.
- Ray, G. 1994: 'Kripke and the Existential Complaint'. *Philosophical Studies*, 74, pp. 121–36.
- Salmon, N. 1987/8: 'How to Measure the Standard Metre'. *Proceedings of the Aristotelian Society*, 88, pp. 193–217.
- Salmon, N. 1993a: 'Analyticity and Apriority'. *Philosophical Perspectives*, 7, pp. 125–33.
- Salmon, N. 1993b: 'Relative and Absolute Apriority'. *Philosophical Studies*, 69, pp. 83–100.
- Soames, S. 1995: 'T-sentences'. In W. Sinnott-Armstrong, D. Raffman and N. Asher (eds), *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*. New York: Cambridge University Press.
- Soames, S. 1997: 'Reply to Garcia-Carpintero and Richard'. *Philosophical Issues*, 8, pp. 79–93.
- Wright, C. 2000: 'Cogency and Question-begging: Some Reflections on McKinsey's Paradox and Putnam's Proof'. *Philosophical Issues*, 10, pp. 140–63.
- Wright, C. 2003: 'Some Reflections on the Acquisition of Warrant by Inference'. In S. Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge*. Cambridge, MA: MIT Press.