

A New Anti-Expertise Dilemma

Thomas Raleigh, University of Luxembourg, thomas.raleigh@uni.lu

ABSTRACT: Instability occurs when the very fact of choosing one particular possible option rather than another affects the expected values of those possible options. In decision theory: An act is stable iff given that it is actually performed, its expected utility is maximal. When there is no stable choice available, the resulting instability can seem to pose a dilemma of practical rationality. A structurally very similar kind of instability, which occurs in cases of *anti-expertise*, can likewise seem to create dilemmas of epistemic rationality. One possible line of response to such cases of instability, suggested by both Jeffrey (1983) and Sorensen (1987), is to insist that a rational agent can simply refuse to accept that such instability applies to herself in the first place. According to this line of thought it can be rational for a subject to discount even very strong empirical evidence that the anti-expertise condition obtains. I present a new variety of anti-expertise condition where no particular empirical stage-setting is required, since the subject can deduce *a priori* that an anti-expertise condition obtains. This kind of anti-expertise case is therefore not amenable to the line of response that Jeffrey and Sorensen recommend.

1. Introduction: Instability and Anti-Expertise

Instability is the general phenomenon that occurs when the very fact of choosing one particular possible option rather than another affects the expected values of those possible options. Such instability can crop up both when we are considering which action to perform and also when we are considering what to believe – so the affected values in question could either be utilities or truth-values (accuracy). The aim for this paper is to present a new case of instability of the latter kind, concerning whether or not to believe a proposition, one which creates, I will suggest, a novel kind of ‘anti-expertise’ dilemma for epistemic rationality. However, since the former kind of instability, concerning the expected utility of different possible actions, has been much more frequently discussed, let’s begin by very briefly rehearsing one of the best-known examples in the decision theory literature: Gibbard & Harper’s (1978) classic ‘Death in Damascus’ case¹.

¹ “Death works from an appointment book which states time and place; a person dies if and only if the book correctly states in what city he will be at the stated time. The book is made up weeks in advance on the basis of highly reliable predictions. An appointment on the next day has been inscribed for him. Suppose, on this basis, the man would take his being in Damascus the next day as strong evidence that his appointment with Death is in Damascus, and would take his being in Aleppo the next day as strong evidence that his appointment is in Aleppo... If... he decides to go to Aleppo, he then has strong grounds for expecting that Aleppo is where Death already expects him to be, and hence it is rational for him to prefer staying in Damascus. Similarly, deciding to stay in Damascus would give him strong grounds for thinking that he ought to go to Aleppo.” (Gibbard & Harper, 1978, 373)

The subject's dilemma in this case is that: if S actually chooses to go to Aleppo then Death will be in Aleppo and so S prefers to be in Damascus, but, if S actually chooses to go to Damascus then Death will be in Damascus and so S prefers to be in Aleppo. Moreover, S can *know in advance* that this will be the case. Assuming that the subject has normal awareness of her own choices, the very act of choosing one option will provide evidence that that option is not optimal. If we are thinking in terms of utility, we can define stability as follows²:

An act is stably preferred iff its expected utility remains maximal *once the decision to perform that act has actually been taken*.

So Death in Damascus is an *unstable* case as there is no stable option available – the decision to do one thing makes the other preferable. To put this in a slightly different way, in Death in Damascus the subject, S, has extremely strong evidence that the following bi-conditional holds:

- **(DD)** S Chooses $\phi \leftrightarrow \phi$ does not maximize utility for S.

Clearly this bi-conditional will be made true, uninterestingly, whenever a subject happens by mistake to choose some option ϕ that does not maximize utility or fails to choose some option ϕ that does maximize utility – e.g. through poor decision making or inattention, etc. But the point, of course, is that in Death in Damascus this bi-conditional will obtain *no matter what the subject chooses to do*. (And the subject knows this beforehand.)

One possible response to a possible scenario in which DD is supposed to hold is simply to accept that it is a genuine dilemma of practical rationality – whatever the subject chooses, she will end up being rationally sub-optimal. A different kind of response is to deny that it is genuinely possible for such a situation to even arise (for a rational agent). Or at least, to deny that a rational agent could *rationally believe* that they themselves are in such a situation where DD applies to them. This was the suggestion made, in passing, by Richard Jeffrey in the second edition of his 'The Logic of Decision' (1983) where he claims that such a situation could only *appear* to arise due to the subject's faulty beliefs or incoherent desires, advising

² One might alternatively define stability in terms of *conditional* expected value: an option is stably preferred iff its expected utility *conditional on its actually being chosen* is maximal. I am very grateful to an anonymous referee for helpful comments on defining stability.

‘you do well to reassess your beliefs and desires before choosing’ (p19)³. We will return to this sort of response in section 2, below.

Death in Damascus and similar instability cases have generated an enormous decision-theoretic literature. But the focus for this paper is a structurally similar kind of ‘unstable’ situation, which occurs in so-called ‘Anti-Expertise’ cases (see, e.g., Conee 1982, Sorensen 1987, Egan & Elga 2005, Christensen 2007, Bommarito 2010).

“I’m a neurologist, and know there’s a device that has been shown to induce the following state in people: they believe that their brains are in state S iff their brains are not in state S. I watch many trials with the device, and become extremely confident that it’s extremely reliable. I’m also confident that my brain is not in state S. Then the device is placed on my head and switched on. My confidence that my brain is not in state S.... well, it’s not clear here what should happen here.” (Christensen 2010, drawn from Conee 1982)

In this anti-expertise case, the subject, S, has extremely good *empirical* evidence that, for some particular proposition, p, the following bi-conditional holds⁴:

- (AE) S Believes that $p \leftrightarrow \neg p$

Again, this bi-conditional will, trivially, happen to be made true whenever the subject happens to believe a falsehood or happens not to believe some truth. But in an anti-expertise case, the bi-conditional will hold *whether or not* the subject believes the proposition in question (i.e. *whatever* doxastic state the subject chooses with respect to p). Moreover, as with Death in Damascus, the subject in such an Anti-Expertise case can know in advance that whether she believes p or not, having formed one doxastic state, the other doxastic state is bound to be more accurate/less inaccurate. Anti-Expertise cases might thus be thought to pose a genuine dilemma of epistemic rationality. I trust that the similarity between the bi-conditionals AE and DD is clear enough. But let’s now briefly consider three points of dissimilarity between the sort of Anti-Expertise case that Conee and Christensen are considering and the Death in Damascus case.

³ Notice that strictly speaking, according to Jeffrey’s own official view, there are no such things as outright or full beliefs – we should eliminate such talk and replace it with subjective probabilities.

⁴ A different condition that is also sometimes labeled ‘anti-expertise’ is when: $(Bp \rightarrow \neg p) \& (B\neg p \rightarrow p)$. I.e. if you believe it, it’s false, if you disbelieve it, it’s true. Notice that this condition creates less threat of a dilemma insofar as it says nothing about suspending judgement or simply neither believing nor disbelieving that p. The bi-conditional AE, in contrast, states that: $(Bp \rightarrow \neg p) \& (\neg Bp \rightarrow p)$.

(i) As Conee sets his anti-expertise case up there *is* causal influence involved. The fictional device is supposed to induce the anti-expertise condition: i.e. there is some kind of causal relation between the subject forming a certain belief and the operations of the machine. In contrast, the figure of Death, in *Death in Damascus*, is neither causally influencing nor causally responsive to the subject's actual choice. The idea is that Death is just a fantastic *predictor* of what the subject's decision will (very probably) be.

(ii) A second point of difference is that it is plausible that one cannot simply *choose* what to believe in the way that one can just choose to go to Damascus or to Aleppo – this is the thesis of doxastic involuntarism. The precise sense in which we can and cannot voluntarily influence our own beliefs is very tricky to spell out precisely – but there is clearly some important sense in which belief formation is very different from forming a decision to act, and this difference turns on the fact that our beliefs (unlike our intentions and desires etc) are somehow essentially constrained and determined by our evidence. As it is sometimes put: 'belief is transparent to the truth'.

(iii) Thirdly, and most importantly for present purposes, unlike *Death in Damascus* where it is stipulated that there are only two possible choices – Aleppo or Damascus – in the Anti-Expertise case there is a third option: one can believe, one can disbelieve *or* one can *suspend judgement*. In other words the option of not believing that *p* sub-divides into two importantly different options – outright disbelief or mere suspension (lack of belief).

The reason that this last dissimilarity is especially important is that Conee himself thought that the rational (or least irrational) option for a subject in the anti-expertise case who accepts (AE) is to suspend judgement. Just to spell this out a little more explicitly: Conee assumes that the subject can suspend whether *p* and then also *know* that she does not currently believe that *p*:

$K(\neg Bp)$

But of course given that she accepts the biconditional (AE), she is in a position to know that her lack of belief entails that *p* is true:

$K(\neg Bp \rightarrow p)$

So she knows, or justifiably believes, all the premises she needs in order to go ahead and perform the simplest of modus ponens inferences to the conclusion that p. But of course if she *did* go ahead and conclude that p, then she would believe that p which would mean, by (AE), that p is false and so we would be set back off on our vicious circle.

Conee suggests that at this point the subject can permissibly *refrain* from drawing this obvious, modus ponens of an inference (despite knowing both premises) and instead just continue to suspend judgement whether p⁵. So Conee here denies that Knowledge (or even justified Belief) is closed under known entailment.

2. Sorensen's response to Anti-Expertise

Roy Sorensen (1987) was not satisfied with this sort of response to the anti-expertise case, levelling two related criticisms at Conee's proposal:

(i) Sorensen claims that it would be a 'low account of rationality', which allows a subject to suspend judgement on a proposition even when she grasps that her own evidence/beliefs *entail* that the proposition is true. In other words, it cannot *always* be rational to suspend judgement no matter how strong your evidence; there must come a point when such epistemic caution becomes irrational and surely this kind of transparently obvious modus ponens inference is about as clear a case as one could possibly have where the evidence rationally calls for belief rather than suspension.

(ii) Sorensen also suggests that when a *perfectly rational* and attentive subject fully believes and explicitly grasps the premises of a simple modus ponens like this, the subject will thereby automatically accept the conclusion (unless of course she *already disbelieves* the conclusion.) In other words, Sorensen claims that at least for an ideally rational, attentive etc. thinker, belief is closed under known entailment: $[Kp \ \& \ K(p \rightarrow q)] \rightarrow Bq$

Now of course this second criticism will be controversial given that a number of epistemologists have denied such closure principles. And I suspect that there may be some philosophers who would want to resist the first criticism. After all, the ancient Pyrrhonians apparently advised people to suspend judgement about everything – though of course if they were to positively believe or judge that: one ought to suspend judgement about everything,

⁵ Reed Richter (1990) likewise argues that it can be rationally permitted to not believe the deductive consequences of your beliefs, *even if* you know and grasp that they have these consequences.

then they would have failed to follow their own advice to suspend judgement about everything! But for present purposes we need not evaluate how telling Sorensen's criticisms of Conee are.

Instead I want to focus on Sorensen's positive alternative suggestion, which is that the rational thing for the subject to do in a putative case of anti-expertise is to refuse to believe that the anti-expertise condition really does apply in the first place, no matter how strong the empirical evidence that it does. In particular, Sorensen makes the claim that *no* conceivable empirical evidence could be strong enough to make belief in (AE) rational, as believing (AE) would then either snare one in a vicious loop/regress, or require one to flout basic principles of rationality such as following through to the conclusion in manifest cases of Modus Ponens. Likewise Sorensen claims that in Death in Damascus no amount of 'empirical evidence' (note empirical!) could rationally convince the subject that if she chooses Aleppo then death will be in Aleppo and if she chooses Damascus then death will be in Damascus. Sorensen then is effectively endorsing Jeffrey's suggestion for the Death in Damascus case – that we should reject that we really are in a case where the bi-conditional holds – and extending the suggestion to the Anti-Expertise case as well. Sorensen comments:

'Since we are warranted in making costly revisions to our background assumptions to escape acceptance of an inconsistent proposition, we are also justified in paying a high price to avoid positions which cannot be consistently accepted.' (Sorensen, 1987, 312)

According to Sorensen then we have here something like what Titelbaum (2015) has more recently called a "fixed point of rationality" – a proposition which no amount of supporting evidence can make it rational to believe⁶. Sorensen himself calls these 'over-shadowed hypotheses' as no matter what the evidence, these hypotheses will always be over-shadowed by some *other* hypothesis – i.e. the evidence will always make it more rational to believe something else. Sorensen provides examples of other supposedly over-shadowed hypotheses:

I don't currently exist.

I have no evidence about anything.

The successor of largest number ever conceived is 9874534244523453.

⁶ Sorensen does allow that in light of new evidence one could increase one's confidence that one is an anti-expert from a low credence to a less low credence, so long as one does not actually (outright) believe that one is an anti-expert. Likewise he allows that one can permissibly believe that one *might* be an anti-expert. So Sorensen does allow for some limited sensitivity to evidence in favour of one's being an anti-expert. Many thanks to an anonymous referee for this journal for helpful comments on this point.

Its raining and I don't believe that its raining⁷.

These are what Sorensen (1988) memorably called "Blindspots". A blindspot is a proposition that can be true but which cannot be both true and believed. Again the idea is that no matter how much empirical evidence we have that such a proposition is true, it cannot be rational for me to go ahead and believe it. It will always be more rational to believe something else. E.g. even if I can both see and feel the rain *and* see a brain scanning device with an amazing track record of reliability which says that I do *not* believe that its raining, still this would not rationalise my believing the blindspot. It would, according to Sorensen, always be more rational to either believe that I am hallucinating the rain, or that the brain-scanner is malfunctioning, etc.

Notice: with some blindspot propositions, just by considering their meaning we can deduce that the following condition holds:

$$SB(p) \rightarrow \neg p$$

E.g. we can tell *a priori* just from the content of the blindspot proposition: 'It's raining and I don't believe that it's raining' that one direction – the left-hand to right-hand direction – of (AE) holds. If I believe that: [Its raining and I don't believe that its raining], then this belief is bound to be false. But of course the other direction of the bi-conditional is *not* guaranteed to hold just because of the content of such a Blindspot:

$$\neg p \rightarrow SB(p)$$

Or contraposing:

$$\neg SB(p) \rightarrow p$$

Just because such a blindspot proposition is false, that would not entail that I believe it! Nor, equivalently, does my lack of belief entail that the blindspot is true.

⁷ This last sentence is, of course, an instance what we now call (following Wittgenstein 1953, part II, section x) 'Moore's paradox'. Moore's original sentence was: 'I went to the pictures last Tuesday, but I don't believe that I did.' (Moore, 1942, 543), which is of the form: $p \ \& \ \neg Bp$. Sentences of the form: $p \ \& \ B\neg p$, or: $Bp \ \& \ \neg p$, are also standardly counted as 'Moorean'.

3. Deducing that an anti-expertise condition obtains

In this section I will argue that (what we might call) the Sorensen-Jeffrey line of response to (putative) cases of Anti-Expertise instability – that one can always rationally reject the evidence that an anti-expertise condition really does apply to oneself – will not always work. The problem is that whilst the sort of cases Sorensen considers, and which have mostly been discussed⁸, are situations in which we get merely very strong *empirical evidence* that we are subject to an anti-expertise condition, it is also possible to *deductively infer* that a kind of anti-expertise condition holds, simply by considering a certain kind of possible thought or judgement. As we will see, this anti-expertise condition does not quite fit the simple schema of (AE): $S \text{ Believes that } p \leftrightarrow \neg p$. Rather, by considering a particular kind of possible demonstrative judgement, the subject can deduce that *in her current situation* she is in an unstable position concerning whether to believe a particular proposition. And whilst that specific proposition might be truly believable *in a different situation*, there would then be a different proposition for which she is in just the same kind of unstable position.

Before we consider the kind of possible judgement that leads to this instability, it will be helpful if we first consider a different but related possible thought:

q: I am currently in *this* total mental state.

The English sentence I have written here to express the thought in question contains a demonstrative term ‘this’, the temporal indexical ‘currently’ and also the 1st-person indexical ‘I’, so of course this same English sentence type can be used to express indefinitely many different propositions on different occasions of use. But the same proposition [at least on one way of individuating propositions] that would get expressed or believed on some particular occasion of stating or judging this sentence type – e.g. that Selma is in a particular total mental state at a particular time – might alternatively be stated or judged using some other, non-indexical form. We could eliminate indexical expressions if we stipulate that ‘M’ picks out the total mental state that Selma is in at time *t*. So if, at time *t*, Selma believes “Selma is in M at time *t*”, this total mental state M thus *includes*, amongst all of Selma’s other beliefs and desires etc., the very belief that ‘Selma is in M at *t*’. This would be an alternative, albeit much less natural, way for a subject to believe the same proposition that would be more

⁸ Compare for example, Koons (1990), who presents a pair of ‘Doxastic Paradoxes’ that do not rely on self-reference. But the set-up of these situations does still rely on the subject somehow having extremely strong empirical evidence that a bi-conditional of the form: $p \leftrightarrow S$ cannot justifiably Bp, really does apply to her.

naturally expressed using the sentence ‘I am currently in *this* total mental state’. Either way it would be a belief about one’s own total current mental state including *all* one’s beliefs and judgements. And so the total mental state that I am thereby thinking about when I form the belief that q *includes* that very mental attitude of believing q: ‘I am in *this* total mental state’.

To judge that q, in this demonstrative format, is bound to be true – it is a bit like judging: ‘I am *this* tall’ whilst demonstrating the top of one’s own head⁹, or “I am currently located *here*’, or “it is currently *now*”, etc. Of course, the proposition that the subject judges in these kinds of cases – e.g. that Selma is currently 1.7 metres tall – is not a necessary truth. There are possible worlds where Selma is some other height and so where Selma would judge falsely that she is 1.7 metres tall. But given that Selma formed her judgement in a distinctively demonstrative way that is bound to refer to her actual height, or to her actual total mental state, the judgement is bound to be true in the actual world – she is bound to be in *this* total mental state, just as she is bound to be *this* tall.

Now that we are familiar with making this kind of demonstrative judgement about one’s own current total mental state, let’s turn to considering the following different possible kind of demonstrative thought.

p: I am not in *this* total current mental state.

In contrast to q, if you form a belief that p, you believe that you are *not* in this current total mental state, where the demonstrative ‘this’ is now picking out a total mental state which includes that very belief p: ‘I am not in this total current mental state.’ Notice therefore that what would be believed if one formed a belief that p is not the negation of what would be

⁹ Wittgenstein (1953) very briefly mentions this sort of demonstrative judgement of one’s own height in section §279 of the *Investigations*, in the course of what is generally considered to be his ‘private language argument’: ‘Imagine someone saying: "But I know how tall I am!" and laying his hand on top of his head to prove it.’ One might naturally read Wittgenstein here as dismissing such a claim and/or demonstration as meaningless. However, in the immediately preceding section §278 he writes: “I know how the colour green looks to me”—surely that makes sense!—Certainly: what use of the proposition are you thinking of?’ Assuming that in the last sentence here Wittgenstein is speaking in *propria persona*, his point then seems to be not so much that such statements are simply meaningless, but rather just that we should ask what the point or usage of such statements is supposed to be in any given context. But setting aside matters of Wittgensteinian interpretation, one possible objection would be to claim that these kinds of demonstrative judgements are meaningless, hence fail to be true. Let me just state that this strikes me as highly implausible. If someone were to point at the ground beneath their feet and assert “I am located *here*!” we may wonder what the point of the assertion is, but there seems no basis whatsoever for denying that they have uttered a perfectly well-formed English sentence, expressing a perfectly meaningful claim that is, trivially, true. Many thanks to an anonymous referee for interesting and helpful discussion on this point.

believed if one instead formed the belief that *q*, because the demonstrative ‘this’ would be picking out a *different* total mental state in each case.

As with *q*, above, the same proposition that is expressed, in some specific context, by the sentence *p*: “I am not in this total current mental state”, could also be expressed in non-indexical fashion. Let’s now stipulate that: ‘M*’ refers to Selma’s total mental state at *t*. And let’s also suppose that at *t* Selma judges that “Selma is *not* in M* at *t*”. The total mental state M* would thus include, along with all sorts of other beliefs and desires etc., the belief that ‘Selma is not in M* at *t*’.

A judgement that *p*: “I am not in *this* total current mental state”, formed in this demonstrative fashion, is bound to be false – it’s a bit like judging: ‘I am *not* this height’ whilst demonstrating the top of your own head. Again, that is not to say that the proposition that gets judged is a necessary falsehood – there are possible worlds in which it is true that Selma is not 1.7 m tall just as there are possible worlds in which it is true that Selma is not in total mental state M* at time *t*. But given the demonstrative way in which these judgements are made, they are bound to be (actually) false. Whereas if *S* forms the non-indexical judgement: ‘*S* is not in M* at *t*’ then this judgement is only guaranteed to be false if it is made at *t* – which is not something that a subject can tell just by reflection. Moreover, *S* might not know which total mental state the label ‘M*’ refers to, or even which person the name ‘*S*’ refers to. It is the indexical form of the judgement: I am not in *this* total current mental state, which ensures that the time at which the judgement is made and the time that the judgement is about are the same (and also that the person making the judgement is the same as the person that the judgement is about). And given that the subject mentally attends to her own current mental state in order to fix the reference of the phrase ‘*this* total current mental state’, there can be no doubt which mental state is in question. This illustrates the familiar point that sometimes a subject can know just by reflection (*a priori* as it were) that a proposition is true/false when it is expressed as a certain kind of demonstrative judgement, and yet the same proposition when expressed in non-indexical terms might not be knowable to be true/false¹⁰.

Now consider that: if, at time *t*, *S* does *not* believe the proposition “*S* is not in M* at *t*”, this proposition must be true – since the total mental state ‘M*’ would *include* a belief in the proposition “*S* is not in M* at *t*”. So if, at time *t*, *S* does not believe this proposition then it is

¹⁰ E.g. I can know by reflection that “I exist right now”, whereas I might not be able to know just by reflection that “*N* exists at time *t*”. One might dispute whether the former sort of knowledge should really count as *a priori* since it plausibly relies on the subject’s self-conscious experience of her own mental life, but it seems clear at least that it does not rely on any empirical evidence about the external world and so is something that can be known just ‘by reflection’.

true – S is not in M^* at t . And so we have the following situation: when S is merely contemplating forming a belief “I am not in *this* total mental state” at some point in the future, she is in a position to grasp that the proposition that would thereby be believed must be true at t if S does *not* believe it at t . Notice that this marks a difference from demonstrative judgements such as: “I am not *this* height” (demonstrating the top of one’s own head), or “I am not *here*” (demonstrating one’s current location), for when we merely contemplate making these latter judgements it is clear that the propositions they would express will (typically) be false whether or not one believes them. E.g. as one merely contemplates the judgement ‘I am not currently *this* height’ one can see that the proposition it *would* express would be false if the judgement were actually made but would also be false if the judgement is not actually formed (unless the very act of actually making/not-making this judgement could somehow thereby change one’s height!).

But S is also in a position to grasp just by reflection that if, at time t , she were to go ahead and actually form the judgement “I am not in *this* total mental state”, that proposition judged would then be false. So this is a situation where the subject can *deduce just by reflection* that a kind of instability obtains concerning what to believe – i.e. the threat of instability here is not based on any empirical evidence about a brain scanning device or an omniscient predictor, etc. Notice also that were the subject to actually form the judgement “I am not in *this* total mental state” and so come to believe the (false) proposition “S is not in M^* at t ”, this would not simply be an instance of a subject having a false belief about her own state of mind – as, say, a subject with implicit or unconsciously racist attitudes might wrongly believe “I am not racist”. Forming a belief “I am not racist” does not thereby *make* one racist, nor would the absence of such a belief ensure that one is not racist. This familiar sort of failure to accurately introspect one’s own beliefs does not amount to an instance of *instability*. Whereas, if the subject, S, at time t does not believe: “S is not in M^* at t ”, then this proposition must be true (since mental state M^* includes a belief in that very proposition). And if, at time t , S judges that “I am not in *this* total mental state’ and so comes to believe “S is not in M^* at t ” then the proposition will be false (since S will then be in M^* at t).

At this point it is worth spelling out the exact nature of the anti-expertise instability that occurs here in some detail. It is particularly important to clearly distinguish the demonstrative sentence or belief type: “I am not in *this* total mental state’ from the proposition that would be expressed or believed on some particular occasion of use (or some particular occasion of being tokened as a belief).

Let's call the subject's initial total mental state, when merely considering or contemplating the possible judgement "I am not in *this* total mental state", M^i . Now M^* can be defined as a possible total mental state that is exactly like M^i *except* that it includes an additional belief that: 'I am not in *this* total mental state'. Notice that the demonstrative element of this possible judgement, were it formed starting from M^i , would refer to the new total mental state M^* , which is exactly like M^i except that it would include this extra belief. (The demonstrative 'this' would *not* be referring back to the initial mental state M^i .)¹¹ So the total mental state M^* *includes* a belief whose content could be expressed non-indexically as "S is not in M^* at t".

If, at time t, S does *not* form the demonstrative belief "I am not in *this* mental state" and so remains in the initial state M^i , then the proposition that this judgement would have expressed is true at t – for the subject is not in M^* at t, she is still in M^i . Moreover, whilst still in M^i , S could be considering or contemplating forming the demonstrative judgement: 'I am not in *this* total mental state' and so can imaginatively grasp how the demonstrative element of that possible judgement would pick out a new, different total mental state that would include that very judgement. I.e. it would pick out M^* . In other words, when S is in the initial state M^i , S is already in a position to entertain and understand the proposition "S is not in M^* at t" by imagining forming the demonstrative judgement: 'I am not in *this* total mental state'. So there is no problem with S *entertaining* and *grasping* the proposition in question before she judges it.

But if S, *given that her initial state is M^i* , does go ahead at some time t and form a belief that: "I am not in *this* total mental state", then S will be in state M^* at t and so this belief is false (since the demonstrative thought effectively claims: S is *not* in M^* at t). So then we have the following kind of anti-expertise instability: given that S starts off in M^i and does not yet believe that "S is not in M^* at t", this proposition will be true at time t if S does nothing and remains in M^i at time t. But, *again given that S starts off in initial state M^i* , if at some time t she forms the demonstrative judgement "I am not in *this* total mental state", this judgement will have the content "S is not M^* at t", and this proposition will be false. Moreover, this is all something that S can grasp in advance simply by reflecting on her situation.

¹¹ If, we understand the demonstrative 'this' in the thought 'I am not in *this* mental state' to be referring to one's actual current total mental state, M^i , then the anti-expertise dilemma evaporates. For when I am in the initial state M^i I don't believe that: [I am not in M^i], which is just as it should be, since it is *false* that [I am not in M^i] when I am in M^i . Conversely if I were to go ahead and believe [I am not in M^i] then that belief would be *true*, for I would then be in a different total mental state, one which is neither M^i nor M^* – for recall, M^* is just like M^i except for the addition of the belief [I am not in M^*], whereas this new total mental state has added the different new belief [I am not in M^i].

Notice that this does not mean that it is just impossible for S to form a true belief in the specific proposition: S is not in M* at t. If S were in some *other* initial total mental state, Mⁱⁱ rather than Mⁱ, then it might well be possible for S to form a true belief in the proposition “S is not in M* at t’”. For if S starts off in Mⁱⁱ, rather than Mⁱ, forming a belief in that proposition would no longer put S into the total mental state M* – which was defined as exactly like Mⁱ except for one extra belief in the proposition “S is not in M* at t’”. If S starts in Mⁱⁱ, and then forms a belief in the proposition “S is not in M* at t’”, S would now be in some other new total mental state rather than M*, and so S would simply believe a further truth.

And so it might seem that the threat of instability here could be evaded by *first changing* one’s total mental state from Mⁱ to some other total mental state Mⁱⁱ (e.g. by forming some other unrelated new belief or desire or experience etc.) and *then* forming a belief in the proposition “S is not in M* at t’”. However this would not be a way of evading the sort of instability in question – it would merely *shift* the instability to a new proposition. For as soon as one changes from initial total state Mⁱ to some other total state Mⁱⁱ, exactly the same kind of instability will just reappear concerning the proposition “S is not in M** at t’”, where M** is exactly like Mⁱⁱ except for the addition of the belief “S is not in M** at t’”¹². If one starts in the different initial state Mⁱⁱ and then forms the demonstrative judgement “I am not in *this* mental state”, this demonstrative judgement would no longer be forming a belief in the proposition “S is not in M* at t’” but rather in the proposition “S is not in M** at t’”. And so forming a belief in this latter proposition would no longer put S into the total mental state M*, but rather into the total mental state, M**.

At the risk of repetition, it may be useful to briefly summarise the discussion in this section: There is a certain kind of demonstrative judgement – I am not in *this* total mental state – which is such that when one merely considers it as a possibility, one can grasp by reflection alone that the proposition that one would thereby believe, were one to actually form the judgement, will be true if one does *not* actually form that judgement. But one can also grasp,

¹² Notice also that this sort of instability would not be evaded, but merely shifted, by *simultaneously* believing both “I am not in this total mental state” and some other unrelated new true proposition, n. Admittedly, forming a belief in the conjunction: ‘n & S is not in M* at t’ would then put S *not* into M* (which is exactly like Mⁱ except for the addition of the belief “S is not in M* at t’”), but into some other new total mental state. And so S’s belief that: ‘S is not in M* at t’ would here be true. But, again, this just shifts the instability to a different proposition. For consider, if starting in Mⁱ, one formed (all at once!) the judgement: “n & I am not in *this* mental state”, the demonstrative *this* would now be picking out a total mental state that is exactly like Mⁱ except for the addition of this conjunctive belief: ‘n & I am not in *this* mental state’. Call this latter new total mental state M***. Given that n is also a true proposition, then by remaining in the initial total mental state Mⁱ at time t, the proposition “n & S is not in M*** at t’ is true but not believed by S. But if, starting from Mⁱ, S were to go ahead at time t and actually believe that: n & S is not in M***, then this proposition would be false. So the anti-expertise instability remains.

by reflection alone, that were one to actually form that kind of demonstrative judgement the proposition believed would then be false. The specific proposition in question could perhaps be truly believed if one were to start from a different initial total mental state – i.e. if one first changed one’s total mental state in some other respect. But then, in this new total mental state, the possible demonstrative judgement: I am not in *this* mental state, would now express a different proposition, for which exactly the same instability applies – one can grasp that the proposition will be true so long as one does not actually form the demonstrative belief, but that if one were to go ahead and believe it then the proposition would be false.

This instability can also be described in non-indexical terms: a subject, S, is bound to start out, before time t, in some total mental state – call it M^i . If, at time t, the subject remains in this initial mental state, M^i , the proposition: S is not in M^* at t, is true (where M^* is exactly like M^i except for the addition of a belief in this very proposition). But if, starting from the state M^i , S believes this proposition at time t then it is false, as S is then in M^* at t. If S were to instead change her total mental state in some other way before time t, so that she is in total mental state M^{ii} , it could then be possible for S to form a true belief that: S is not in M^* at t. However, now the same kind of instability reappears for the proposition: S is not in M^{**} at t (where M^{**} is exactly like M^{ii} except for the addition of a belief in this proposition). Etc¹³.

Although we can, from the outside as it were, work through this anti-expertise instability in non-indexical terms, the possibility of forming beliefs in these propositions via a demonstrative judgement of the form: “I am not in *this* total mental state” is crucial since it is what allows the subject to be able to deduce by reflection alone that this kind of anti-expertise instability applies to her. If a complex, total mental state is described or referred to in non-indexical terms, the subject may have no idea whether or not she is currently in it or whether she would be in it were she to form some belief, or what the time is expressed in non-indexical terms. This is not something she can work out just by reflecting on her situation. It is the demonstrative form of the belief “I am not in *this* total mental state” that allows S to grasp that the proposition in question will be true if the belief remains unformed, but also that actually forming such a demonstrative belief in this proposition is bound to be false.

¹³ I am extremely grateful to an anonymous referee for this journal whose helpful comments substantially improved the non-indexical formulations in this section.

4. Comparisons & Clarifications

In this section I briefly compare and contrast the anti-expertise condition presented above with various other puzzling or paradoxical judgements and sentences that have been discussed in the literature.

Let's begin with the following puzzling sentence (see e.g. Burge 1978, Tymoczko 1984, Caie 2012, though the idea arguably goes back to Buridan's *Sophismata*):

'I do not believe that this sentence is true'

The sentence above contains a demonstrative '*this*' which refers to that very sentence itself and might well be thought to create an anti-expertise situation where: $SB(p) \leftrightarrow \neg p$. However, there is a familiar sort of worry whether the sentence: 'This sentence is true' really expresses the sort of genuine propositional content that could be believed in the first place. *Which* proposition or possible state of affairs is one supposed to believe/not-believe to be true? It is far from obvious, as with 'The Liar', that such a 'Truth-teller' sentence has a (classical) truth-value¹⁴. And indeed, the focus of Caie's (2012) illuminating discussion is the question: which doxastic attitude should one take towards a claim that is neither determinately true nor determinately false? In contrast, the judgement that: I am currently in *this* total mental state, seems to be straightforwardly and determinately true, just as judging that I am currently *this* height (whilst demonstrating the top of one's head) is straightforwardly, though trivially, true. Conversely, judging that one is *not* in this current total mental state would seem to be straightforwardly and determinately false. An advantage of the present proposal then is that it avoids any worries about whether the sentence in question really expresses a genuine content that is fit to be believed or about whether the content is determinately true or false. So the anti-expertise dilemma I have presented here cannot be avoided by insisting that there is really nothing that is fit to be the content of a genuine true/false belief in the first place.

Kripke (1980) famously offered as an alleged example¹⁵ of a contingent truth that can be known *a priori*: "Stick S is one metre long at time t", as uttered by someone who is defining the unit 'one metre' by reference to stick S at time t. This sort of stipulative reference-fixing of what is to count as 'one metre' – i.e. *This* length shall henceforth be called 'one metre' –

¹⁴ See, e.g. Kripke 1975, Mortensen & Priest, 1981.

¹⁵ The example occurs on p.56 of his *Naming and Necessity* (1980). Kripke's other example of the contingent a priori was "If Neptune exists, Neptune is the cause of the perturbations in the orbit of Uranus" – see Kripke 1980, p75.

certainly bears a similarity to the sort of demonstrative judgements, such as: ‘I am *this* tall’, ‘I am located *here*’, ‘I am in *this* mental state’, which were discussed above. I think it’s fair to say that Kripke’s Paris Metre Rod example has been much more controversial than his celebrated examples of necessary a posteriori truths¹⁶. So it is worth clarifying that nothing I have said here requires accepting the existence of contingent a priori truths. What I have claimed is that one can know by reflection alone (without relying on any empirical evidence) that a certain kind of anti-expertise condition applies to oneself. This anti-expertise condition can be captured by a pair of conditional claims:

- IF, at some time *t*, one were to form the demonstrative judgement ‘I am not in *this* total mental state’ THEN the proposition that would be expressed by forming that judgement at *t* would be false.
- IF, at some time *t*, one does not form the demonstrative judgement “I am not in *this* total mental state’ THEN the proposition that would be expressed by forming that judgement at *t* would be true.

These conditional claims are knowable *a priori* but they are *necessary* truths, not contingent truths. Notice also that in order to grasp that this anti-expertise condition obtains the subject does not need to *actually make* any demonstrative judgement herself – such as “I am not in *this* mental state’. The subject need only *contemplate* the possible demonstrative judgement in order to grasp that what it *would* express will be false if the judgement is actually made, but will be true if the judgement remains unformed.

In Makinson’s (1965) ‘Preface Paradox’, the subject starts off by rationally believing each individual proposition in a large set of propositions, but then due to rational concerns about her own fallibility, the subject comes to believe that the conjunction of all the propositions in the set is false. Likewise, in Lottery cases (Kyburg 1961) a subject can apparently rationally believe of each individual ticket that it will lose and yet also rationally believe that not all of the tickets will be losers. There may appear to be some similarity to the case I have presented here insofar as the subject starts in some total doxastic state, believing each individual proposition in some large set of propositions, and then considers forming the judgement: I am not in *this* total doxastic state. However, there are also a number of significant dissimilarities. Firstly, denying that all of the propositions in some set of propositions are *true* (as in the Preface and Lottery cases) is different from denying that some set of propositions comprises all of the propositions that one believes (which is what the judgement “I am not in *this* total doxastic state” would do). Secondly, the extra proposition in the Preface and Lottery cases, which effectively says: “But not all of those propositions are correct!”, is not talking about

¹⁶ See, e.g. Donellan (1977), Casullo (1977), Bonjour (1998), Turri (2011) for criticisms which deny that Kripke’s example really is contingent a priori.

itself. We start with a set of propositions, B, all of which are initially believed, and then we have an extra proposition which says: at least one proposition in B is false. This extra proposition is not itself a member of B. No self-reference is involved. Whereas, what is crucial to the anti-expertise case I have presented here is that the demonstrative judgement: "I am not in *this* total mental state" is itself included in the total state mental state that it refers to. Thirdly, in both the preface and lottery cases, it seems eminently rational to actually go ahead and believe the extra proposition, which denies that all of the other propositions can all be true. (Indeed in the case of the lottery one can *know* that not all the tickets will lose.) Whereas, with the sort of anti-expertise condition I have presented here, it clearly would not be rational to actually go ahead and form the demonstrative judgement in question. For the subject can easily deduce that *were* she to actually go ahead and form the judgement "I am not in *this* total mental state" the proposition thus judged would then be false. It is only if it remains unbelievably by the subject that the proposition in question will be true; the very act of actually forming the belief will make it false. In the preface and lottery paradoxes, there is nothing like this dependence of a proposition's truth-value on whether the subject believes it. Conversely, what is paradoxical or problematic about the preface and lottery cases is that the subject seems to be rationally led into having a manifestly inconsistent set of beliefs. Whereas, in the present anti-expertise case having an inconsistent set of beliefs is not the issue; the issue is *instability* concerning a single proposition.

The following case, which does concern a single proposition (at least on one way of individuating propositions), is due to Mark Crimmins (1992):

'You have known me for years', explained Gonzales, 'But there is something you have not discovered. You know me under two guises, just as Lois Lane knows Superman. You do not realize that I am the person you know under another guise. On that way of thinking about me, you have quite different opinions of me. In fact, you think me an idiot.'

'Knowing your cleverness', I replied, 'I must with some embarrassment accept what you say. Since I do not know what guise you mean, I do not know which belief to revise. Until I find out, it seems, I falsely believe that you are an idiot!' (Crimmins, 1992, 191)

This case may also seem similar to the sort of anti-expertise I have discussed, since in Crimmins' case the subject, S, seems to be rationally required to believe: SBp & ¬p, and

Crimmins' case also involves an indexical form of judgement ('*you* are an idiot!')¹⁷. However, again, the truth-value of the proposition in Crimmins' case does not depend on whether the subject believes it. And whereas Crimmins' case relies on there being two different modes of presentation ('guises') for the same item, the sort of anti-expertise case I have presented does not involve both believing and disbelieving something under two different modes of presentation. Moreover, Crimmins' case relies on the subject having *empirical evidence* to support the proposition that: I falsely believe that you are an idiot. In contrast, as I have already emphasized, the anti-expertise condition I have been discussing can be known to obtain *a priori*.

Having highlighted the contrasts between the anti-expertise condition I have presented and various other well-known cases in the literature, I will finish this section by emphasizing one point of similarity between this anti-expertise condition and Gödel's 1st incompleteness theorem¹⁸. We saw above that for any given initial total mental state, M^i , there will be a proposition which will be true if, starting from M^i , one does not form a belief in it, but which will be false if, starting from M^i , one does form a belief in it. That does not mean that this proposition cannot ever be truly believed, for it might be truly believable starting from some other initial state. But then for this different initial total state there will then be some further proposition for which just the same sort of instability obtains. There is, I suggest, a parallel here with Gödel's first incompleteness theorem, which famously showed that for any consistent formal system that can express arithmetic, L , there is a true sentence which cannot be proved in that system. Speaking rather loosely and informally, the 'Gödel sentence' in question effectively 'says': "I am not provable in L ". Slightly less loosely: this sentence makes some arithmetical claim that will be true iff there is no proof of this sentence in L . This does not mean that this sentence is unprovable in *any* consistent formal system that can express arithmetic. If we *change* the formal system in question from L to L^* – e.g. by simply adding the original Gödel sentence to the axioms of L , or by moving to the meta-language in which the original result about L was demonstrated – then the sentence may be provable in L^* . But then there will be some further, new Gödel sentence for L^* which effectively says "I am not provable in L^* ".

So then, to spell it out explicitly, the parallel is this: Gödel's celebrated result showed that for any specific formal system there will be a specific sentence that is either true but not provable

¹⁷ In fact I don't think that the use of the indexicals 'you' or 'I' is essential for Crimmins' case – the same issues would arise for 'S falsely believes that Gonzalez, master of disguise, is an idiot'.

¹⁸ I am very grateful to Peter Brössel for helpful conversations about the comparison with Gödel's theorem.

in that system or provable in that system but false. The sentence may be both true and provable in some other formal system but then the same issue will arise with some other sentence for this other formal system. Likewise, we have seen that for any specific initial total mental state there is a proposition that will be true if a belief in the proposition is *not* formed from that total initial mental state, but which will become false if a belief in the proposition *is* formed from that total initial mental state. This specific proposition may be truly believed if it were formed starting from some other initial total state. But then the same situation will arise with some other proposition for this other initial total state. Now, needless to say, there are *all sorts* of important differences between the Gödel sentence for some formal system and the demonstrative judgement “I am not in *this* total mental state”. No small part of Gödel’s achievement was to come up with the system of (what we now call) ‘Gödel Numbering’, allowing him to construct his proof for the language of arithmetic in which there are not any self-referential devices like: ‘this sentence...’ or ‘I am not provable...’. Indeed, it might seem dangerously lacking in humility to even mention the sort of anti-expertise condition I have presented here in the same breath as Gödel’s great proof! So let me be clear that the comparison is only intended to illuminate how the kind of anti-expertise condition that is created by the thought ‘I am not in *this* total mental state’ will always keep reappearing as one shifts from one total mental state to another.

5. Conclusion

The moral of the story for present purposes is that what I have called the ‘Sorensen-Jeffrey’ line of response to (alleged) cases of instability does not always work. For just by considering or contemplating the possible judgement “I am not in this current total mental state” one is able to grasp that a certain kind of anti-expertise instability obtains. Sorensen’s suggestion that we can rationally discount any ‘amount of empirical evidence’ that we are in an anti-expertise situation, simply does not apply here as empirical evidence plays no role. Nor is it at all obvious how Jeffrey’s advice – that you would ‘do well to reassess your beliefs’ – would apply here either, since the only basis we need for grasping that this kind of anti-expertise obtains is pure *a priori* deduction.

However it is important to note that the spirit of this ‘Sorensen-Jeffrey’ line of thought appeals to the (supposed) *holism* of our ‘web of beliefs’ and to the idea that it can sometimes be rationally permitted to make drastic adjustments to our background beliefs and assumptions in order to hold onto some very important belief elsewhere in the web – e.g.

Sorensen at one point speaks of ‘a Duhemian response’ to anti-expertise cases¹⁹. And if one takes this Quinean-Duhemian sort of view, one might hold that ‘no statement is immune to revision’²⁰ (Quine, 1961, 43), *including statements that express supposedly a priori, necessary truths*, such as logical or mathematical laws. E.g. Quine (1961) suggested that one might reject the law of excluded middle in order to ‘simplify quantum mechanics’. And so, given such a thoroughly holistic picture of how beliefs can be revised, it must be conceded that one could, in principle, try to resist accepting that the anti-expertise condition presented here applies to oneself by revising or rejecting one of the (apparently) a priori and necessary truths involved or one of the logical laws or rules of inference that I relied on in presenting the case. But it is hard to see how such a move could be plausibly motivated. Sorensen, recall, claimed that we are ‘justified in making costly revisions to our background assumptions’ in order to ‘avoid positions which cannot be consistently accepted’. When these revisions merely involve questioning whether past observations or memories are reliable, it seems plausible enough that the cost, even if it is high, may be less than the cost of accepting that one is caught in a rational dilemma. But if the revisions would require us to jettison (apparently) necessary truths or basic rules of deductive inference then it becomes much less plausible that the cost should be any less than the cost of accepting that one is subject to anti-expertise. After all, there is no obvious reason to think that our views about the nature of such a dilemma – and whether it really should be avoided at all costs – are any more central or well-entrenched in our web of beliefs than (what we take to be) other necessary *a priori* truths or rules of deductive inference²¹. Moreover, if Sorensen were still to insist that a rational subject ought not accept that they are subject to the kind of anti-expertise I have presented here, he would be recommending exactly what he criticised Conee for recommending: viz., refraining from drawing a deductive consequence that one is perfectly able to grasp. Sorensen took himself to be striking a *different* bargain to Conee. But in this case rejecting the idea that one is subject to a form of anti-expertise would be to fail to draw a manifestly valid deduction, which is just what Conee originally advised.

So then, assuming that the foregoing discussion has been on the right track, we have here a kind of anti-expertise instability that is not based on any empirical set-up – such as a science-fiction brain-scanning machine, or an evil demon or Oracle, etc. – and to which we are all

¹⁹ I am very grateful to an anonymous referee for this journal for pressing me to discuss this point.

²⁰ It might be thought that the statement: ‘no statement is immune to revision’ is in danger of leading to paradox if it applies to itself. See, e.g. Katz (1988), Elstein (2007), Ebbs (2016) and Baumann (2017) for discussion.

²¹ Notice also, though it is an *ad hominem*, that Sorensen himself is well-known for championing the epistemicist approach to vagueness on the basis that it allows us to retain classical logic, and for rejecting any rival approaches which would require us to revise or deviate from classical logic.

constantly subject. It is thus a promising candidate to be a genuine anti-expertise dilemma, one that cannot be avoided in the manner Sorensen and Jeffrey recommended²².

References

- Baumann, P. (2017) “Is Everything Revisable?” *Ergo: An Open Access Journal of Philosophy* 4: 349-357.
- Bommarito, N. (2010) ‘Rationally Self-Ascribed Anti-Expertise’, *Philosophical Studies* 151(3): 413-419.
- Bonjour, L. (1998) *In Defense of Pure Reason*. New York: Cambridge University Press.
- Burge, T. (1978) ‘Buridan and epistemic paradox’, *Philosophical Studies*, 34: 21–35.
- Caie, M. (2013) ‘Belief and Indeterminacy’, *Philosophical Review*, 122(4): 527-575.
- Casullo, A. (1977) “Kripke on the A Priori and the Necessary”, *Analysis* 37: 152 – 159.
- Christensen, D. (2007) ‘Epistemic Self-Respect’, *Proceedings of the Aristotelian Society* 107 (1pt3): 319-337.
- Christensen, D. (2010) ‘Higher-Order Evidence’ *Philosophy and Phenomenological Research* 81(1): 185-215.
- Conee, E. (1982) ‘Utilitarianism and Rationality’ *Analysis* 42(1): 55-59.
- Crimmins, M. (1992) ‘I Falsely Believe that p’, *Analysis* 52 (3):191 (1992)
- Donnellan, K. (1977) “The Contingent A Priori and Rigid Designators”, *Midwest Studies in Philosophy* 2: 12-27.
- Ebbs, G. (2016) ‘Reading Quine’s Claim that no Statement Is Immune to Revision’, in Frederique Janssen-Lauret and Gary Kemp (Eds.), *Quine and His Place in History*, Palgrave Press.
- Egan & Elga (2005) ‘I Can’t Believe I’m Stupid’ *Philosophical Perspectives* 19(1): 77–93
- Elstein, D. (2007) “A New Revisability Paradox”, *Pacific Philosophical Quarterly*, 88(3), 308–318
- Gibbard & Harper (1978) ‘Counterfactuals and Two Kinds of Expected Utility’, in A. Hooker, J. J. Leach & E. F. McClennen (eds.), *Foundations and Applications of Decision Theory*. D. Reidel.
- Jeffrey, R. (1983) *The Logic of Decision* (2nd Edition), University of Chicago Press.
- Katz, J. (1988) *Realistic Rationalism*. Cambridge, MA: MIT Press.
- Kripke, S. (1975) ‘Outline of a theory of truth’ *Journal of Philosophy*, 72: 690–712.
- Kripke, S. (1980) *Naming and Necessity*, Cambridge, MA: Harvard University Press.
- Koons, R. (1990) ‘Doxastic paradoxes without self-reference’, *Australasian Journal of Philosophy* 68(2): 168 – 177.
- Kyburg, H. (1961) *Probability and the Logic of Rational Belief*, Middletown: Wesleyan University Press.
- Makinson, D. (1965) “The Paradox of the Preface”, *Analysis*, 25: 205–207.
- Mortensen, C. & Priest, G. (1981) ‘The Truth Teller Paradox’, *Logique et Analyse*, 95-96: 381-388.
- Quine, W.V. O. (1961) “Two Dogmas of Empiricism”, in *From a Logical Point of View* (2nd ed.) Harvard University Press.
- Richter, R. (1990) ‘Ideal Rationality and Hand Waving’, *Australasian Journal of Philosophy* 68 (2): 147 – 156

²² Earlier versions of this paper were presented at the conference ‘Hard Cases and Rational Choice’ at the University of Bern, at the workshop ‘Rationality: Epistemic and Practical Perspectives’ and at the colloquium on Logic and Epistemology, both at the Ruhr University Bochum. I am very grateful to the audiences on all those occasions for helpful questions and feedback. Many thanks in particular to Peter Brössel, Ruth Chang, Insa Lawler, Jim Pryor, Kevin Reuter, Christian Straßer and Filippo Vindrola for their comments, objections and advice. Finally, I am especially grateful to the anonymous referees for this journal and for another journal, whose reports very substantially improved this paper.

- Sorensen, R. (1987) 'Anti-Expertise, Instability, and Rational Choice', *Australasian Journal of Philosophy* 65 (3): 301 – 315
- Sorensen, R. (1988) *Blindspots*, Oxford University Press
- Titelbaum, M. (2015) "Rationality's Fixed Point (Or: In Defense of Right Reason)" *Oxford Studies in Epistemology* 5: 253–294
- Turri, J. (2011) 'Contingent A Priori Knowledge', *Philosophy and Phenomenological Research* 83 (2):327-344.
- Tymoczko, T. (1984) 'An Unsolved Puzzle about Knowledge', *The Philosophical Quarterly*, 34 (137): 437-458.
- Wittgenstein, L. (1953) *Philosophical Investigations*, G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.