Kelp, C. and Pritchard, D. (2009) Two deflationary approaches to Fitch-style reasoning. In: Salerno, J. (ed.) New Essays on the Knowability Paradox. Oxford University Press: Oxford, pp. 324-338. ISBN 9780199285495.

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.gla.ac.uk/140932/

Deposited on: 15 May 2017

TWO DEFLATIONARY APPROACHES TO FITCH-STYLE REASONING

Christoph Kelp and Duncan Pritchard

University of Stirling

ABSTRACT. This paper considers two *deflationary* responses to the Fitch argument on behalf of the semantic anti-realist—that is, two responses which aim to evade the conclusion of that argument by, on a principled basis, weakening one of the principles essentially employed. The first deflationary approach that is considered—which proceeds by weakening the factivity principle for knowledge—is shown to be ultimately unpromising, but a second approach—which proceeds by weakening the knowability principle that is at the heart of semantic anti-realism—is shown to have considerable *prima facie* appeal. It is then argued that some key objections that one might raise for this approach are on closer inspection ineffective.

0. INTRODUCTION

Frederic Fitch (1963) famously argued that the thesis that all truths are knowable (henceforth, the *knowability principle*), in conjunction with a handful of apparently highly plausible logical and epistemic principles, entails the obviously absurd claim that all truths are known. This argument has become known as the *paradox of knowability*. Of course, it is only a *paradox* if one finds the knowability principle highly plausible in the first place, since a basic pre-requisite of an argument qualifying as a paradox is surely that it involves a highly contentious—indeed, *unacceptable*—conclusion which validly follows from highly plausible premises. Moreover, there is good independent reason to think that such a principle is not so plausible. It seems that the principle knowability is naturally understood as applying to cognizers like us; that is, to subjects with finite cognitive capacities and a finite lifespan. At the same time, it is plausible that there are some propositions that are too large to be grasped by such cognizers—for instance, some disjunctions with infinitely many disjuncts. If these finite cognizers cannot grasp such propositions, however, then they cannot know them either. In consequence, for cognizers like us, some propositions must remain unknown. Provided that the principle of knowability is naturally understood as applying to cognizers like us, then it is not plausible that it is true. If so, however, the 'paradox' of knowability, then, isn't strictly speaking a paradox at all.[1,2]

Nevertheless, there are (as we will see in a moment) substantive philosophical grounds in favour of the knowability principle, and thus even if Fitch's argument does not point to a

paradox as such it may still be thought to be a potential *reductio* of those philosophical views which feel the theoretical need to incorporate this principle. Accordingly, if there are such theoretical views then its defenders had better have something compelling to say in response to Fitch's argument.

In this paper, we will look at one—perhaps the only—theoretical view to which, on the face of it, the knowability principle is of central importance. We will then consider two *deflationary* responses to Fitch's argument on behalf of defenders of this view. What we mean by a 'deflationary' response to the argument is a proposal which proceeds by weakening, on a principled basis, one of the principles essentially employed by that argument. The motivation for this strategy is this: *ceteris paribus*, if one can accommodate the considerations which prompt adoption of a certain principle by advancing a version of that principle which is (perhaps only slightly) logically weaker, then one ought to do so. If one can further show that the Fitch argument is blocked once the weaker 'deflated' version of the principle is adopted, then one will have succeeded in offering a deflationary response to the argument.[3]

The first deflationary response that we will consider proceeds by weakening the factivity principle for knowledge. We will argue that this strategy does not stand up to closer inspection. Nevertheless, we claim that there are good grounds for holding that the second deflationary response that we consider—which rejects the principle of knowability in favour of a weaker principle—is effective at resolving the problem posed by Fitch's argument.

1. SEMANTIC ANTI-REALISM

The view to which the knowability principle is, on the face of it, of central importance, is often labelled 'semantic anti-realism'. Semantic anti-realism is the rejection of realist theories of meaning (i.e., semantic realism). Indeed, semantic anti-realists often explicitly motivate their position by pointing to defects in realist theories of meaning. In this section, we will outline the problems which, according to the semantic anti-realist, beset realist theories of meaning and show how accepting the knowability principle can potentially avoid these problems.

To begin with, let us look at the credentials of realist theories of meaning. Realist theories of meaning are commonly construed as having the following two features:

(1) The meaning of a statement is identified with its truth conditions.
(2) There are evidence-transcendent truths.[4]

From these features of realist theories of meaning it follows that some statements have evidence-transcendent truth-conditions as their meanings. A general constraint on a theory of meaning is that it should at least be compatible with a theory of understanding—that is, for a theory of meaning to be satisfactory it must be compatible with an account of what a competent speakers' linguistic understanding consists in.[5] Semantic anti-realists suspect that realist theories of meaning will be unsatisfactory on just score because they are incompatible with a satisfactory account of our understanding of statements with evidence-transcendent truth-conditions.

Semantic anti-realists base their suspicion on a challenge to realist theories of meaning which arises from what they consider to be an important Wittgensteinian insight into the nature of understanding—*viz.*, that understanding a concept consists in a set of practical abilities rather than in a state of mind. Certainly, if one is to be credited with a given practical ability then one must be able to manifest that ability in one's behaviour. For instance, a child will be credited with the ability to swim only if she is able to manifest swimming behaviour in suitable circumstances. Hence, if the Wittgensteinian insight is to be taken seriously—that is, if understanding is to be conceived of as a set of practical abilities—then understanding must be manifestable in behaviour too. Presumably, the kind of behaviour in which understanding must be manifestable is linguistic behaviour (i.e., language use). According to the semantic anti-realist, however, what would count—at least minimally—as a manifestation by a speaker of her understanding of a statement in use is that the speaker is able to *evaluate* her own and other peoples' use of the statement and, if circumstances render it appropriate, to *adjust* her use of it accordingly.[6]

Given that we understand what counts as manifestation of understanding in use in this way, however, it is hard to see how understanding of statements with evidence-transcendent truth-conditions could be manifested in use. After all, the truth-conditions of such statements are evidence-transcendent. As a result, there aren't any circumstances that would provide the basis for an evaluation of one's own or other people's use of such statements. And, similarly, there aren't any circumstances in the light of which one would adjust one's use of such statements. If there aren't any such circumstances, then understanding of statements with evidence-transcendent truth-conditions cannot be manifested. And if understanding of such statements cannot be manifested, then it does not consist in a set of practical abilities after all—contrary to what the Wittgensteinian insight suggests. Accordingly, the challenge that semantic anti-realists pose to their realist opponents is to provide an account of understanding

of statements with evidence-transcendent truth-conditions that is both faithful to the two core realist theses and respects the Wittgensteinian insight. Their suspicion is that this cannot be done.[7]

A related challenge that semantic anti-realists pose to semantic realists focuses on the acquisition of our understanding of statements with evidence-transcendent truth-conditions. Since if we accept a truth-conditional theory of meaning we acquire our understanding of a type of statement by bringing to bear evidence on the truth-values of instances of it, semantic anti-realists argue that it is hard to see how we could so much as acquire an understanding of statements with evidence-transcendent truth-conditions. Accordingly, semantic anti-realists challenge their opponents to provide an account of how we acquire our understanding of statements with evidence-transcendent truth-conditions.[8]

These two challenges were first advanced by Michael Dummett (1978) and have become known as the manifestation and the acquisition challenge, respectively.[9] Although there are further anti-realist arguments, these challenges—and the manifestation challenge in particular—appear to be the most common reason offered by semantic anti-realists as to why they find realist theories of meaning problematic.[10] Accordingly, semantic anti-realists have proceeded to deny at least one of the two core theses of realist theories of meaning.

Initially, semantic anti-realists were tempted to deny the realist's first core claim—i.e., the commitment to a truth-conditional theory of meaning—and replace it with a theory that identifies the meanings of statements with their assertibility conditions. The rationale for this is obvious, since by tying the meaning of a statement to its assertibility conditions (which are held *not* to be evidence-transcendent) rather than its truth-conditions, the anti-realist avoids the problems posed for a theory of meaning by allowing evidence-transcendent truths.

More recently, however, this option appears to have become less appealing to semantic anti-realists. Instead, they have tended to reject the realists' second core claim—i.e., that there are evidence-transcendent truths.[11] It ought to be clear that accepting the possibility of evidence-transcendent truths entails accepting the existence of unknowable truths, at least if one accepts the further (highly plausible) claim that in order to know a proposition one must have evidence in favour of it. Accordingly, if one holds, with the knowability principle, that there cannot be any unknowable truths, then it follows that one must reject the idea that there are evidence-transcendent truths as well. Given the foregoing, there is clearly a large theoretical pay-off in rejecting this key realist claim, since it avoids the worries just noted regarding our understanding of such truths. If there aren't any such truths, then the fact that it

is doubtful whether an understanding of them can be manifested—or acquired for that matter—won't be a problem for the semantic anti-realist.

Given that our primary interest is the Fitch argument, it is this second strand of semantic anti-realist thought—which, like the Fitch argument, has the knowability principle at its heart—that is our concern here. Henceforth, when we talk of 'semantic anti-realism' we will have this specific variety of semantic anti-realism in mind.

2. FITCH'S ARGUMENT

Fitch's argument clearly poses a fundamental challenge to semantic anti-realism. Indeed, given that it is an undeniable truth that we are not omniscient, unless the semantic anti-realist can find some way to block this argument then she is faced with a *reductio* of her position. In order to be able to discuss some options the semantic anti-realist may have to block Fitch's argument, it will be a good idea to look at how the argument proceeds in a bit more detail. First, we will formalise the knowability principle in the following way:[12]

$$(KP) \qquad (\forall P)(P \rightarrow \Diamond(\exists s,t)(Ks,t\, P))$$

Now we assume, for *reductio*, that one is not omniscient—i.e., that there is some truth (we'll call it 'P_1') which is unknown:

$$(1) \qquad P_1\, \&\, \neg(\exists s_1,t_1)(Ks_1,t_1\, P_1)$$

Given (KP), however, one can straightforwardly derive (2):

$$(2) \qquad \Diamond(\exists s_2,t_2)(Ks_2,t_2\, (P_1\, \&\, \neg(\exists s_1,t_1)(Ks_1,t_1\, P_1)))$$

An essential feature of Fitch's argument at this point is a sub-argument to the effect that (2) is false. This proceeds by first assuming, for *reductio*, that the statement within the scope of the possibility operator at line (2) is true:

$$(3) \qquad (\exists s_2,t_2)(Ks_2,t_2\, (P_1\, \&\, \neg(\exists s_1,t_1)(Ks_1,t_1\, P_1)))$$

Plausibly, knowledge distributes across conjunctions, such that if a conjunction is known, then so are both of the conjuncts:

$$(4) \qquad (\exists s_2,t_2)(Ks_2,t_2\, P_1)\, \&\, (\exists s_2,t_2)(Ks_2,t_2\, \neg(\exists s_1,t_1)(Ks_1,t_1\, P_1))$$

Most will also agree that knowledge is factive, such that if one knows a proposition, then that

proposition must be true. We can thus conclude (5):

> (5) $(\exists s_2, t_2)(K s_2, t_2 \, P_1) \, \& \, \neg(\exists s_1, t_1)(K s_1, t_1 \, P_1)$

This is, of course, a contradiction. Since the assumption of this sub-argument leads to contradiction, we can therefore infer the negation of this assumption:

> (6) $\neg(\exists s_2, t_2)(K s_2, t_2 \, (P_1 \, \& \, \neg(\exists s_1, t_1)(K s_1, t_1 \, P_1)))$

Moreover, since this result has been derived based on no assumptions, we can also conclude that it is a necessary truth:

> (7) $\neg(\exists s_2, t_2)(K s_2, t_2 \, (P_1 \, \& \, \neg(\exists s_1, t_1)(K s_1, t_1 \, P_1)))$

Using standard modal logic, however, we can infer (8) from (7):

> (8) $\neg \Diamond (\exists s_2, t_2)(K s_2, t_2 \, (P_1 \, \& \, \neg(\exists s_1, t_1)(K s_1, t_1 \, P_1)))$

Now (8) is obviously inconsistent with (2). It therefore follows that the original assumption—that we are non-omniscient—must be denied.

The knowability principle, at least when combined with some very basic epistemic and modal logic, is therefore inconsistent with non-omniscience such that if we retain this principle then we must, it seems, accept the absurd conclusion that that all truths are known.[13] Fitch's argument therefore poses a serious problem for semantic anti-realism. In the remainder of this paper we will explore two deflationary approaches that the semantic anti-realist could pursue in order to evade this argument.

3. A DEFLATIONARY APPROACH TO FITCH'S ARGUMENT I: WEAKENING THE FACTIVITY PRINCIPLE

The first deflationary proposal that we will be exploring considers the prospects of offering an anti-realist response to Fitch's argument which denies the factivity of knowledge. It should be quite obvious that once the factivity of knowledge is denied, the argument that leads to the paradox, at least in its present form, will no longer go through since the step from (4) to (5) will no longer be valid.

Of course, it is easy to *say* that one does not accept factivity and that, therefore, one isn't impressed by Fitch's argument. However, factivity seems to play an important—indeed, indispensable—role in any plausible theory of knowledge. In particular, it is one of the central

guiding intuitions regarding knowledge that one cannot know falsehoods. That is, the very idea that there could be a case in which an agent knows a proposition and yet that proposition is false, just seems plain incoherent. Now of course one might claim that even the most deeply entrenched intuitions could be called into question on theoretical grounds. Even if this is so, however, it remains that any theory which denied factivity and thereby held that it was possible to know falsehoods would face a pretty severe up-hill struggle when it came to gaining widespread acceptance.

Nevertheless, there may be some room for manoeuvre here. After all, as we will now see, there is a potential logical gap—at least by semantic anti-realist lights—between the claim that there cannot exist any cases in which an agent knows a falsehood and the factivity claim that knowledge entails the truth of the proposition known. If this is right, then the semantic anti-realist can exploit this logical gap in order to motivate a weakened version of the factivity principle which can nevertheless retain the core guiding intuition behind factivity that there cannot exist cases of false knowledge.

In order to see this, let us first state factivity more formally:

(FAC) $(\forall P)(\forall s)(\forall t)(\mathrm{K}s,t\, P \rightarrow P)$

Furthermore, let us state explicitly the intuition that is meant to drive adoption of (FAC)—*viz.*, that there are no cases of false knowledge:[14]

(*) $\neg(\exists P)(\exists s)(\exists t)(\mathrm{K}s,t\, P\, \&\, \neg P)$

Now from (*) we can derive (**):

(**) $(\forall P)(\forall s)(\forall t)\neg(\mathrm{K}s,t\, P\, \&\, \neg P)$

And from (**) we can derive (***):

(***) $(\forall P)(\forall s)(\forall t)(\mathrm{K}s,t\, P \rightarrow \neg\neg P)$

From (***) it might seem like a very small move indeed to get to (FAC), since all one needs to do is introduce the double negation equivalence rule (DNE) to eliminate the double negation in the embedded consequent. Crucially, however, intuitionistic logic does not contain (DNE), and yet it is precisely this logic that semantic anti-realists typically endorse. Accordingly, it follows that an anti-realist can accept the intuition guiding adoption of (FAC)—which we have expressed as (*)—without being compelled to endorse (FAC) itself. Instead, this guiding intuition merely entails the weaker claim which we have expressed as (***), but which it is open to the semantic anti-realist to argue is itself a respectable version of

factivity. We will call this weakened version of factivity, (FAC*):

(FAC*) $(\forall P)(\forall s)(\forall t)(\mathrm{K}s,t\ P \rightarrow \neg\neg P)$

This line of reasoning seems deflationary in just the right sort of way, since it shows that there is, at least by the lights of a particular theoretical outlook, a way of properly responding to the core intuition motivating (FAC) which results in a logically weaker principle. If this logically weaker principle can help the semantic anti-realist block Fitch's argument, then this would thus be an extremely attractive way of resolving the situation.

Unfortunately, however, closer inspection reveals that the present proposal is ultimately unsuccessful. True, on the face of it, (FAC*) blocks the move from line (4) to line (5) in that it only gives us (5*):

(5*) $(\exists s_2,t_2)(\mathrm{K}s_2,t_2\ P_1)\ \&\ \neg\neg\neg(\exists s_1,t_1)(\mathrm{K}s_1,t_1\ P_1)$

Crucially, however, this triple negation collapses into a single negation, even within an intuitionistic logic, and thus one will be able to derive line (5) of the paradox of knowability anyway, even without having to appeal to (FAC).

So one won't solve the paradox of knowability by rejecting the factivity of knowledge and replacing it by the ever so slightly weaker (FAC*). In order to get this line to work one would have to replace the factivity principle with something that is weaker even than (FAC*). The difficulty facing such a proposal, however, is that it will not be able to do full justice to our intuition that one cannot know falsehoods. In this way, it is highly doubtful whether the present deflationary strategy can ultimately be successful.[15]

4. A DEFLATIONARY APPROACH TO FITCH'S ARGUMENT II: WEAKENING THE KNOWABILITY PRINCIPLE

Although the proposal to deny factivity will not do the trick, there is a second proposal available that is in the same deflationary spirit and which is much more promising. The thought is that instead of rejecting one of the epistemic principles which are employed within Fitch's argument, one instead rejects the very principle that is the target of that argument—i.e., the knowability principle itself. In its stead is then put forward a slightly weaker principle which can nevertheless accommodate the guiding motivation behind the knowability principle.

Informally, the weakened principle that we have in mind is as follows: for all true propositions, it must be possible to justifiably believe them. More formally:

(JP) $\quad (\forall P)(P \rightarrow \Diamond(\exists s,t)(JBs,t\,P))$

In order to see why this principle of justified believability, as we will call it, suits the purposes of the semantic anti-realist it is important to first notice that it accommodates the semantic anti-realists' worries regarding realist theories of meaning.

Recall that the semantic anti-realist argued that realist theories of meaning will have a problem explaining how we can acquire and manifest an understanding of the meanings of statements with evidence-transcendent truth-conditions. Recall, furthermore, that we saw that accepting the knowability principle will avoid this problem. Given the plausible additional assumption that one knows a proposition only if one also has evidence for it, it follows that if all truths are knowable then it must also be possible to have evidence for them. If it must be possible to have evidence for all true propositions, however, then there can be no evidence-transcendent truths. In this way, the semantic anti-realist can resist the realist's second core claim that there are evidence-transcendent truths by accepting the knowability principle.

Notice, however, that a parallel argument will show that accepting the justified believability principle will do the job just as well. After all, it is also plausible that one justifiably believes a proposition only if one has evidence for it. That means, however, that if, for all truths, it must be possible to justifiably believe them, then it must also be possible to have evidence for them. But if it must be possible to have evidence for all true propositions, then there can be no evidence-transcendent truths. In this way, the semantic anti-realist can resist the realist's second core claim by accepting the justified believability principle. In short, this principle will do the job for the semantic anti-realist just as well as the knowability principle.

Notice that while replacing the knowability principle with the justified believability principle will allow the semantic anti-realist to avoid the conclusion that we are omniscient— after all, justified belief is not knowledge[16]—that does not mean that the semantic anti-realist is no longer susceptible to refutation by a Fitch-style argument. After all, a parallel argument for justified belief threatens to show that the justified believability principle entails that all statements are justifiably believed. And that, it would seem, is almost as bad for the semantic anti-realist as the original conclusion of Fitch's argument. So there is still work to be done.

One might think, however, that even if there is work to be done, it is not much work. After all, justified belief, as opposed to knowledge, is not factive. That is, one can justifiably

believe a falsehood. For instance, one might reliably and conscientiously form the belief that there is a barn over there—and thereby have a *justified* belief in this proposition—even though this belief is nonetheless false because, unbeknownst to you, what you are in fact looking at is a barn façade. Accordingly, since Fitch's argument relies on the factivity of knowledge, it follows that it will not go through if the knowledge operator is replaced by a justified belief operator. Hence it would seem as though all the semantic anti-realist has to do is to replace the knowability principle with the justified believability principle in order to avoid the conclusion of Fitch-style reasoning.

5. PROBLEMS WITH THE SECOND DEFLATIONARY APPROACH TO FITCH-STYLE REASONING

There are, however, problems on the horizon for this line of reasoning. In particular, one might think that the conclusion just canvassed is either false or uninteresting. To take the first horn first, one might think that it is false because even granted that justified belief is not factive, the following reflection principle does, nonetheless, hold: if, at a certain time, one justifiably believes that one does not at that time justifiably believe a proposition, then one does not at that time justifiably believe that proposition. More formally, we can express this principle as follows:

(RP) $(\forall P)(\forall s)(\forall t)(JBs,t\,(\neg JBs,t\,P) \rightarrow \neg JBs,t\,P))$

The significance of this principle is that the relevant 'factivity' move in a Fitch-style argument employing the justified belief operator would be from (4') to (5'):

(4') $(\exists s_2,t_2)(JBs_2,t_2\,P_1) \,\&\, (\exists s_2,t_2)(JBs_2,t_2\,\neg(\exists s_1,t_1)(JBs_1,t_1\,P_1))$
(5') $(\exists s_2,t_2)(JBs_2,t_2\,P_1) \,\&\, \neg(\exists s_1,t_1)(JBs_1,t_1\,P_1)$

If the justified belief operator were factive, then that would straightforwardly licence this inference. It ought to be clear, however, that even if the justified belief is not factive, then this inference will go through just so long as (RP) holds. So one might object that the mere fact that justified belief is not factive does not get the semantic anti-realist off the hook, since it is still plausible that justified belief satisfies the reflection principle which, it would seem, suffices to generate the Fitch result.

On the other hand, one might object that the result is uninteresting because it has long been established that the semantic anti-realist can resist the conclusion of Fitch-style

argument by stating the epistemic constraint on truth in terms of justified believability. J. L. Mackie makes the point in the following passage:

> Suppose we read *K* [*the knowledge operator in Fitch's argument*] as 'It is justifiably believed at t_1 that …' This will distribute over &, but we might expect the argument now to fail at step 4 [*to 5 in the above statement of the argument*], since this *K* is not truth-entailing. But step 4 [*to 5*] still goes through. If it is justifiably believed that *p* at t_1 that *p* is not justifiably believed at t_1, then *p* is *not* justifiably believed at t_1. On the other hand, if we read *K* as 'It is justifiably believed at some time that…', then step 4 does not go through. It does not follow that if it is justifiably believed at any time that *p* is not justifiably believed at any time, then *p* is not justifiably believed at any time. It might be justifiable at t_1 to think that *p* is false and never has been and never will be justifiably believed and yet there might be some other time t_2 at which *p* was, or will be justifiably believed. So the argument does not enable us to reject the principle that what is true can be justifiably believed at some time. (Mackie 1980, 91-2)

In this passage, Mackie distinguishes between two reflection principles for justified belief, one which he deems plausible and one which he deems implausible. The plausible reflection principle has it that if it is justifiably believed at t_1 that it is not justifiably believed at t_1 that *p*, then it is not justifiably believed at t_1 that *p*. This is, of course, the reflection principle—(RP)—that we formulated above. In contrast, according to the implausible reflection principle, if someone at some time justifiably believes that no-one ever justifiably believes that *p*, then no-one ever justifiably believes that *p*. This principle can be formalised in the following way:

(RP*) $(\forall P)((\exists s,t)(\text{JB}s,t \neg(\exists s_1,t_1)(\text{JB}s_1,t_1\ P)) \rightarrow \neg(\exists s_1,t_1)(\text{JB}s_1,t_1\ P)))$

Mackie claims, correctly and for the right reasons, that (RP*) is false. He goes on to claim, again correctly, that the conclusion of Fitch's argument can be avoided if the epistemic constraint is construed in terms of justified believability at some time—i.e., what we have called the justified believability principle. Unfortunately, however, this last claim, while correct, is made for the wrong reasons. For, as we are about to show, Fitch's conclusion can be derived from (RP), which Mackie deems plausible, and the justified believability principle.

To begin with, we start with the relevant assumption for *reductio*—someone at some time justifiably believes that *p* and that no-one ever justifiably believes that *p*:

(3'') $(\exists s_2,t_2)(\text{JB}s_2,t_2\ (P_1\ \&\ \neg(\exists s_1,t_1)(\text{JB}s_1,t_1\ P_1)))$

If (3'') is true, then so is an instance of it. Or in other words, if someone at some time justifiably believes that *p* and that no-one ever justifiably believes that *p*, then there must be a particular epistemic subject who believes this conjunction at a particular time. Let the epistemic subject and time be s_3 and t_3, respectively. We then get:

(4'') $JBs_3,t_3 (P_1 \& \neg(\exists s_1,t_1)(JBs_1,t_1 P_1)))$

Since justified belief distributes across conjunctions, we get:

(5'') $JBs_3,t_3 P_1 \& JBs_3,t_3 \neg(\exists s_1,t_1)(JBs_1,t_1 P_1)$

Now if one justifiably believes that there is no-one at any time who justifiably believes that P_1, then one also justifiably believes that, currently, one does not justifiably believe P_1 oneself.[17] Accordingly, from (5'') we can derive:

(6'') $JBs_3,t_3 P_1 \& JBs_3,t_3 (\neg JBs_3,t_3 P_1)$

Given (RP), however, the second conjunct of (6'') entails that s_3 does not justifiably believe P_1 at t_3:

(7'') $JBs_3,t_3 P_1 \& \neg JBs_3,t_3 P_1$

From here the Fitch-style argument proceeds as rehearsed. So we can argue to its conclusion without having to appeal to the implausible reflection principle, (RP*). All that we need is (RP) which Mackie deems a plausible reflection principle. So Mackie's distinction between the two reflection principles will not help the semantic anti-realist. If the semantic anti-realist is to get any mileage out of rejecting the principle of knowability and replacing it by the weaker principle of justified believability, then she must have some other way of resisting the Fitch-style conclusion.

Fortunately for the semantic anti-realist, however, there is excellent reason to believe that (RP) does not hold. Consider the following case due to Saul Kripke: Pierre is a Frenchman who has lived most of his life in France. Having just returned from a trip to London, one of Pierre's best friends asserts "Londres est jolie." Since Pierre knows his friend to be a man of exceptional taste he believes what his friend asserted and hence comes to believe that London is pretty. Now suppose that, by some unfortunate circumstance, Pierre finds himself stuck in a particularly unattractive part of London. Pierre is forced take on a badly paid job that will just pay him enough to buy food and accommodation. At this time he learns English 'directly'—that is, by direct interactions with other English speakers rather than referring to, say, translation manuals. Pierre uses the term 'London' as his neighbours do and learns everything his neighbours know about it which, let us suppose for the sake of argument, is not very much. On the basis of his experiences in the city he comes to believe that London is not pretty. At the same time, Pierre is still sometimes thinking about his nice

life in France, and sometimes even of his friend who told him about the pretty city of London. In such moments Pierre thinks to himself: "Si seulement je serais en Londres …" Obviously, Pierre still believes that London is pretty and hence he has inconsistent beliefs. Moreover, his inconsistent beliefs are both justified. The testimony from a person with exceptional taste justifies his belief that London is pretty while his direct experiences justify his belief that London isn't pretty.

It is plausible that whilst having inconsistent beliefs that are both justified, Pierre may also believe, justifiably, that he does not believe that London is pretty. Perhaps some psychologist analyses him and tells him that the source of his recent unhappiness is simply that he no longer believes himself to be living in a pretty city. Pierre thus comes to believe, and justifiably so (since on the basis of the reliable testimony from the psychologist), that he does not believe that London is pretty. But if one justifiably believes that one does not believe a proposition, then one also justifiably believes that one does not justifiably believe that proposition. Accordingly, Pierre also justifiably believes that he does not justifiably believe that London is a pretty city.

Pierre's case thus indicates that one can simultaneously justifiably believe all of the following: (a) a proposition, *P*, (b) its negation, not-*P*, and (c) the proposition that one does not justifiably believe *P*. Given that this is so, however, it can easily be seen that the reflection principle (RP) must fail. For if (RP) held, it would follow that Pierre both does and does not justifiably believe that London is pretty. (RP) turns an inconsistency in Pierre's belief-system (in conjunction with a second-order belief), into an inconsistency in the world. So it must be false.[18] If (RP) is false, however, then the relevant Fitch-style argument no longer goes through. The semantic anti-realist is off the hook.

There is, however, a further difficulty for the semantic anti-realist who endorses the justified believability principle. It remains true that since there are some statements that are true but will never be justifiably believed, it must, by the justified believability principle, also be possible for someone at some time to justifiably believe an instance of this. Among other things that means that it must be possible for someone at some time to justifiably believe statements of the form "*P* but no-one ever justifiably believes *P*" and, similarly, "*P* but I don't justifiably believe *P*". And, as Dorothy Edgington (1985, 558) has pointed out, one might think that this is already bad enough for the semantic anti-realist. After all, it would seem that one just couldn't have any evidence for statements of either form. If so, then it would seem that one also cannot justifiably believe such statements.

Moreover, recall that the semantic anti-realist introduces the justified believability principle in order to ensure that meaning, construed truth-conditionally, can always be manifested in use. If there are truths of the form "*P* and no-one ever justifiably believes *P*" and "*P* and I don't justifiably believe *P*", then one must be able to manifest the meaning of those statements in understanding. Since justified believability is supposed to secure manifestability, it must be possible to justifiably believe statements of the form "*P* and no-one ever justifiably believes *P*" and "*P* and I don't justifiably believe *P*". But if it is impossible to have evidence that would support statements of this form, then one cannot justifiably believe such statements. So even if the semantic anti-realist can deny the reflection principle, (RP), Fitch's argument shows that things are already bad enough for the semantic anti-realist even before the problematic principle comes into play.

It would seem, however, that there are ways for the semantic anti-realist to respond to this difficulty. Let us begin with statements of the form "*P* and I don't justifiably believe that *P*". In order to argue that statements of this form can be justifiably believed, the semantic anti-realist can simply point to Pierre's case again and claim that Pierre might well come to believe that London is pretty (by believing that the proposition expressed by "Londres est jolie" is true) and that he does not believe that London is pretty (by believing that the proposition expressed by "I don't believe that London is pretty" is also true). Since both of his beliefs are justified and since we typically acquire a justified belief in a conjunction by conjoining the justification we have for the beliefs in the conjuncts,[19] it would follow that the semantic anti-realist can comfortably allow that Pierre justifiably believes that London is pretty and that he does not justifiably believe that London is pretty.[20]

Things are a bit more difficult when it comes to statements of the form "*P* and no-one ever justifiably believes that *P*". However, the situation is not hopeless for the semantic anti-realist. Consider, for instance, a case in which I am stranded on a lonely island. The only thing I have with me is a book about psychology which is written in English. I read about a brain lesion, named 'X', and learn that the only symptom of X, which occurs in 99% of the cases, is a continuous strong belief on the part of the patient that she has X. Through introspection, I find that I don't believe that I have X. I therefore come to believe that I don't believe that I have X. My belief, since based on reliable introspective capacities, is justified. I justifiably believe that I don't believe that I have X. Since if one justifiably believes that one does not believe a proposition, *P*, then one also justifiably believes that one does not justifiably believe *P*, I justifiably believe that I don't justifiably believe that I have X. Moreover, since I am on a lonely island, without drinkable water, and since I have every reason to believe that no-one

will come to my rescue and that I will die fairly soon, I also justifiably believe that no-one will ever justifiably believe that I have X. To finish the story off, suppose that just before I left for my holiday, I called my doctor to get the results for some brain tests that they had done on me. My doctor told me that everything was fine except that I have a brain condition called 'Y', which is completely harmless, and that I could go on the trip without any problem. On the basis of the testimony from my doctor I come to believe, justifiably, that I have Y. Now, since I am German and have talked to my German doctor our conversation was naturally in German. What I don't know is that the German expression 'Y' and the English expression 'X' are names for the same brain lesion and that I am among the lucky 1% of patients who don't suffer from the symptoms. I am now in a condition in which I justifiably believe that I have X (on the basis of testimony from my doctor: 'Sie haben Y') and I also justifiably believe that no-one ever justifiably believes that I have X (on the basis of introspection, what I have read about X in the psychology book and my unfortunate predicament of being on a lonely island about to die). If I were to conjoin my two beliefs, I would justifiably believe that I have X and that no-one ever justifiably believes that I have X.

There is thus a way for the semantic anti-realist to respond to Edgington's worry. The semantic anti-realist may point out that, contrary to what Edgington claimed, one can justifiability believe statements of the form "*P* and I don't justifiably believe *P*" as well as "*P* and no-one ever justifiably believes *P*". The semantic anti-realist is, again, off the hook. In general, there are good grounds for holding that the deflationary strategy of replacing the knowability principle with the justified believing principle may well offer the semantic anti-realist a way of avoiding Fitch-style reasoning.[21]

REFERENCES

Brogaard, B., & Salerno, J. (2004). 'Fitch's Paradox of Knowability', *Stanford Encyclopædia of Philosophy*, (ed.) E. N. Zalta, http://www.seop.leeds.ac.uk/entries/fitch-paradox/.
Dummett, M. (1978). *Truth and Other Enigmas*, London: Duckworth.
——(1993). *The Seas of Language*, Oxford: Clarendon Press.
Edgington, D. (1985). 'The Paradox of Knowability', *Mind* 94, 557-68.
Fitch, F. (1963). 'A Logical Analysis of Some Value Concepts', *The Journal of Symbolic Logic* 28, 135-42.
Greenough, P. (2002). *Knowledge, Lies and Vagueness: A Minimalist Treatment*, Ph.D. dissertation, University of St. Andrews, St. Andrews, Scotland.
Kvanvig, J. (1995). 'The Knowability Paradox and the Prospects for Anti-Realism', *Noûs* 29, 481-500.
——(2006). *The Knowability Paradox*, Oxford: Oxford University Press.
Mackie, J. L. (1980). 'Truth and Knowability', *Analysis* 40, 90-2.

Pritchard, D. H. (2004). 'Epistemic Deflationism', *The Southern Journal of Philosophy* 42, 103-34.

Putnam, H. (1981). *Reason, Truth and History*, Cambridge, UK: Cambridge University Press.

Tennant, N. (1997). *The Taming of the True*, Oxford: Clarendon Press.

Williamson, T. (1988). 'Knowability and Constructivism', *Philosophical Quarterly* 38, 422-32.

—— (1992). 'On Intuitionistic Modal Epistemic Logic', *Journal of Philosophical Logic* 21, 63-89.

—— (2000). *Knowledge and Its Limits*, Oxford: OUP.

Wright, C. (1993). *Realism, Meaning and Truth*, Oxford: Blackwell.

—— (1994). *Truth and Objectivity*, Cambridge, Massachusetts: Harvard University Press.

—— (2000). 'Truth as Sort of Epistemic: Putnam's Peregrinations', *Journal of Philosophy* 97, 335-64.

NOTES

[1] The thesis that the 'paradox' of knowability is not really a paradox has also been defended by Williamson (2000, ch. 12).

[2] Of course, one might find it independently puzzling that it is even possible to derive the conclusion that all truths are known from the premise that all truths are knowable. Fitch's argument would then be philosophically interesting even if one did not find the knowability principle plausible. Still, what is making the argument—taken in isolation—philosophically interesting is *not* that it poses a paradox.

[3] This deflationary strategy—applied to epistemological issues—is explored and defended at greater length in Pritchard (2004). See also Greenough (2002).

[4] Cf. Wright (1993, 250). We take it that for present purposes this is an adequate representation of Wright's statement of realism: "Realism about a given discourse, for the purposes of the Manifestation Challenge, is simply the combination of views (a) that the proper account of our understanding of its statements is evidence-transcendent truth-conditional, and (b) that the world on occasion exploits, so to speak, this understanding—does on occasion deliver undetectable truth-conferrers to such statements." (Ibid.)

[5] Cf. Wright (1993, 47).

[6] Cf. Wright (1993, 247).

[7] Cf. Wright (1993, 247-8).

[8] Cf. Wright (1993, 87).

[9] See also Dummett (1993).

[10] For two further anti-realist arguments, see Wright (1993), who outlines a challenge that proceeds from the normativity of meaning, and Putnam (1981), who adduces the so-called 'model-theoretic' argument. Most contemporary anti-realists appear to accept the manifestation challenge. For examples, see Dummett (1978; 1993), Wright (1993) and Tennant (1997).

[11] For a good contrast of the two anti-realist approaches, compare the early and later essays collected in Wright (1993).

[12] We will take quantification over propositions (P, P_1 etc.,), subjects (s, s_1 etc.,), and times (t, t_1 etc.,) for granted. For a statement of the argument that does not rely on substitutional quantification, see Kvanvig (1995).

[13] See Williamson (1988; 1992) for an argument to the effect that the conclusion just canvassed—that all truths are known—will not follow within intuitionistic logic from Fitch's reasoning, even though it does follow that non-omniscience is false. For an excellent overview of the debate regarding Fitch's puzzle, see Brogaard & Salerno (2004).

[14] Note that to avoid unnecessary complications, we have here expressed the intuition in a slightly weaker form—i.e., that there are no cases of false knowledge, rather than that it is *impossible* for there to be cases of false knowledge. Nothing in what follows turns on this.

[15] Interestingly, in some recent (and unpublished) work Finn Spicer and Allan Hazlett have independently argued that there are good grounds for rejecting (FAC) outright. In particular, they argue for the plausibility of the claim that knowledge is best understood as reliable true belief. Thus, since one can have a reliably-formed false belief, it follows that (FAC) must go. If such an argument could be made compelling, then it would hold out the prospect that the semantic anti-realist could exploit this proposal in order to block the Fitch argument on non-factivity grounds. Notice, however, that such a suggestion is not within a deflationary spirit. That is, unlike the deflationary proposal explored here, the move is not towards offering a logically weaker formulation of

factivity which can nevertheless do justice (by the broader lights of that theory at any rate) to the intuitions which drive acceptance of factivity. Instead, this line of argument involves the straightforward rejection of those intuitions. This feature of the proposal makes this sort of manoeuvre dialectically problematic, and certainly ensures that it is not relevant for our purposes here.

[16] At least on standard views of justification at any rate. If one held that justification was factive, then there would be scope to contend that there is no logical gap between justified belief and knowledge. Such a theory of justification would be highly revisionary, however, and so we can legitimately set this possibility to one side here.

[17] Some may object that the step from (5'') to (6'') will not go through because justified belief is not closed under known entailment (never mind under entailment *simpliciter*). However, the argument does not depend on such closure principles. It is plausible that if (RP) is valid—that is if one justifiably believes at t_1 that one does not justifiably believe that p at t_1, then one does not justifiably believe that p at t_1—then so is the following reflection principle: if one justifiably believes at a given time that no-one ever justifiably believes that p, then at that time one does not justifiably believe that p. Stated formally:

$$(\text{RP**}) \quad (\forall P)(\forall s)(\forall t)(\text{JB}s,t \; \neg(\exists s_1,t_1)(\text{JB}s_1,t_1 \; P) \rightarrow \neg \text{JB}s,t \; P)$$

(RP**) will validate the inference to (7'') without relying on any further closure principles.

[18] If one wants to run the argument by appeal to (RP**) instead of (RP)—see the above footnote—one will need a slightly different case to make the present point. We provide such a case below.

[19] Cf. Kvanvig (2006, 21).

[20] Of course, another consequence is that Pierre can come to justifiably believe a contradiction. This may initially seem an unwelcome consequence of the view. However, it is not clear why one could not justifiably believe a contradiction (at least so long as it is not an obvious one). A clever logician could easily tell me that what, in effect, is a complicated contradiction is true and I could come to believe this contradiction on that basis. Since the logician is an otherwise reliable informant on such issues, and since the testimony of reliable informants furnishes us with justified beliefs, my belief in the contradiction is surely justified.

[21] We are grateful to Brit Brogaard, Tony Brueckner, Laurence Goldstein, Patrick Greenough, Allan Hazlett, Stephen Maitzen, Aidan McGlynn, Alan Millar, Paul O'Grady, David Papineau, Sven Rosenkranz, Peter Sullivan and to an anonymous referee from Oxford University Press. Special thanks also go to the editor of this volume, Joe Salerno, for all his (considerable) help.