

## Epistemicism about Vagueness and Meta-Linguistic Safety

Stephen Kearns  
Cornell University

Ofra Magidor  
University of Oxford (Balliol College)

### §1 Introduction

In his book ‘Vagueness’, Timothy Williamson defends the following claims.<sup>1</sup> First, sentences containing vague expressions satisfy (other things being equal) the relevant instances of the law of excluded middle and the law of bivalence. Even if, for example, Jack is a borderline case of tallness, Jack is tall or Jack is not tall and ‘Jack is tall’ is true or false. To speak more metaphorically: vague expressions and correspondingly the semantic values denoted by them have “sharp boundaries” or “sharp cut-off points”. Second, we do not and cannot know the sharp cut-off points corresponding to vague expressions.<sup>2</sup> If, for example, Jack is a borderline case of tallness then we do not and cannot know whether or not Jack is tall. Third, the explanation of why we do not and cannot know the sharp cut-off points corresponding to vague expressions is a safety-based explanation. The explanation goes roughly like this: in close-by possible worlds, ones in which the use facts for a certain vague expression differ only slightly from their actual use facts, the meaning and consequently the cut-off point of the expression is slightly different (or, to use the terminology of Hawthorne (2006), vague expressions are *semantically plastic*). This entails that our beliefs about the cut-off points corresponding to vague expressions are not safe and thus do not constitute knowledge.

It is important to note that the latter claim plays a central role in Williamson’s defense of the epistemicist position, for it provides a response to two standard objections to epistemicism. First objection: if vague expressions have sharp cut-off points, how is it that we are not able to say what these cut-off points are? Response: the safety based explanation shows why the cut-off points are unknowable, so it is of no surprise that we cannot say what they are.<sup>3</sup> Second objection: if vagueness consists of ignorance, then what makes vagueness a distinctive phenomenon, distinguished from other kinds of ignorance? Response: vagueness involves a special kind of ignorance, one that is

generated because of particular kinds of margin of error principles, namely ones that involve the semantic plasticity of vague expressions.<sup>4</sup> Williamson's specific safety-based explanation for our ignorance is thus supposed to characterize what is distinctive about this ignorance.

In this paper we challenge Williamson's safety based explanation for why we cannot know the cut-off points associated with vague expressions. We will assume throughout (most of) the paper that Williamson is correct in saying that vague expressions have sharp cut-off points, but we argue that Williamson's explanation for why we do not and cannot know these cut-off points is unsatisfactory. In §2, we present Williamson's position in more detail. In particular, we define the notion of meta-linguistic belief safety (MBS), which Williamson seems to implicitly assume is a necessary condition on knowledge. In §3, we argue that even if MBS were a necessary condition on knowledge, this would not entail that we cannot know the cut-off points associated with vague expressions. In §4, we present our main argument against Williamson's explanation. We argue that MBS is not a necessary condition on knowledge by presenting a series of cases where one's belief violates MBS but nonetheless constitutes knowledge. In §5, we discuss an objection to our view, one according to which if it were allowed that we know non-meta-linguistic cut-off point claims ('*m* is the cut-off point for tallness') that violate MBS, then it should be allowed that we know meta-linguistic cut-off point claims ('*m* is the cut-off point for 'tall') which violate a more straightforward safety condition on knowledge. We argue that this objection is not compelling. Finally, in §6 we briefly discuss what are the possible directions that a theory of vagueness can take if our objection to Williamson's theory is taken on board.

## **§2 Epistemicism and meta-linguistic safety**

According to Williamson's view vague expressions, just as non-vague ones, have 'sharp' semantic-values. What makes it the case that a certain expression receives a certain semantic value? We can divide the facts that go into determining which expressions receive which semantic values into two components. First, there is the totality of use facts of an expression: we follow Williamson in assuming that these can be reduced to a complex

collection of sharp, non-intentional facts (e.g. purely physical facts).<sup>5</sup> Second, there is a supervenience function - that is, the manner in which the meaning facts are determined by the use facts.

Given this observation, one might be tempted into one of two ways for explaining our ignorance of sharp cut-off points: the first is to argue that the ignorance is due (merely) to our ignorance of the exact use facts of a vague expression and the second is to argue that the ignorance is due (merely) to our ignorance of the supervenience function of meaning on use. As we understand Williamson, neither of these routes provides the correct explanation of our ignorance on his view.

One cannot explain our ignorance merely by appealing to ignorance of the use facts, simply because it seems that even if we knew the complete use facts for vague expressions, we still would not be able to figure out the sharp cut-off points associated with them. That is to say, we do not seem to have any idea how might go about discovering true conditionals of the form: "If 'tall' is used with the use facts  $u$ , then a person is tall if and only if they are over  $m$  meters in height".<sup>6</sup>

Less trivially, one cannot explain our ignorance merely by appealing to ignorance of the supervenience function. It is true that according to Williamson the supervenience function might be extremely complicated and unknowable *a priori*. But that in itself would be insufficient to show that the supervenience function is all together unknowable. After all, there are many facts about the world are both complicated and not knowable *a priori*, and yet we seem to be able to discover them. Complex principles of physics, for example, are discovered by a combination of direct observations and subtle theorizing. Moreover, it is plausible to think that empirical research can yield knowledge of certain *a posteriori* supervenience functions, such as the supervenience of mental states on physical states. Thus the fact that the supervenience function is complicated can at best explain why we do not currently know it, not why we could not come to know it subject to more research.

But Williamson claims that while we might come to know, for example, the supervenience function of the mental states on the physical states, we could not come to know the supervenience function of the meaning of vague expressions on their use. He argues that the two cases are disanalogous in the following way. He points out that in our efforts to discover the supervenience of the mental on the physical our investigations could perhaps appeal to both physical and mental facts to which we have *direct access*, that is to say, facts that we can know by routes other than just ‘calculating them’ from the relevant supervenience function. For example, we might know that a person is in a certain physical state – say brain state *b* - using some laboratory equipment. And at the same time we might know that the person is in a certain mental state – say being in pain - simply because they told us they are in pain. But Williamson maintains that in contrast, we have no direct access to facts about the cut-off points of vague predicates. As he puts it: “‘Everyone with physical measurements *m* is thin’ cannot be known *a posteriori* in a parallel way, for no route to independent knowledge of someone with physical measurements *m* that he is thin corresponds to asking someone whether he is in pain”.<sup>7, 8</sup> The upshot is, then, that even if one wishes to claim that we are principally ignorant of the relevant supervenience function, this ignorance is derivative on our ignorance of the sharp cut-off points of vague expressions and not vice versa.

Williamson’s account thus consists neither of arguing for ignorance of the use facts, nor for arguing for ignorance of the supervenience function. Instead, he provides a more direct explanation of our ignorance of sharp cut-off points. Here’s roughly how the account goes. We start with the idea that safety is a necessary condition for knowledge.<sup>9</sup> A simple definition of what it takes for a belief to be safe runs as follows:

**Simple Safety:** An agent *X*’s belief that *p* is *simply safe* if and only if there is no close-by possible world *w* in which (a) in *w*, *X* believes that *p* (b) in *w*, it is false that *p*.<sup>10</sup>

On the face of it, Simple Safety looks like a plausible necessary condition for knowledge: if one has a true belief that *p* but could have easily falsely believed that *p*, then it looks like the fact that one’s actual belief is true is no more than a lucky coincidence, and not

knowledge. But in so far as safety conditions are supposed to capture our intuition that luckily true belief cannot count as knowledge, Simple Safety is insufficient. For beliefs in necessary truths always satisfy Simple Safety, since there are no close-by possible worlds in which a necessary proposition  $q$  will be false. Yet intuitively, it seems that some beliefs in necessary propositions should count as merely luckily true beliefs, and hence in some sense unsafe. Indeed, this problem is relevant to the case in question: supposing that  $m$  is in fact the cut-off height in meters for ‘tall’, the proposition expressed by ‘A person is tall if and only if they are more than  $m$  meters in height’ is a necessarily true proposition.<sup>11</sup>

To avoid this problem, Williamson maintains that it is a different notion of safety that is violated in the case in question. He argues as follows: “Someone who asserts ‘Everyone with physical measurements  $m$  is thin’ is asserting a necessary truth, but he is still lucky to be speaking the truth... Although he could not have asserted the proposition he actually asserted without speaking truly, he could very easily have asserted a different and necessarily false proposition with the same words”.<sup>12</sup>

What is the exact safety condition that underlies this argument? In this passage Williamson seems to apply a notion of safety to assertions rather than beliefs:

**Meta-linguistic Assertion Safety:** An agent  $X$ ’s assertion of a sentence  $s$  is *meta-linguistically assertion safe* (MAS) if and only if there is no close-by possible world  $w$  such that (a) in  $w$ ,  $X$  asserts the sentence  $s$  (b) in  $w$ , the proposition expressed by this assertion of  $s$  is false.

Using this definition, we can explain how assertions of necessary truths can count as unsafe: a sentence  $s$  which actually expresses a necessary truth can, in a close-by possible world, express a different proposition, one that is not necessarily true and in fact is false in that world. Indeed, this is exactly what, according to Williamson, happens in the case of specifications of cut-off points such as ‘A person is tall if and only if they are more than  $m$  meters in height’. According to Williamson, slight differences in use facts will

entail slight differences in meaning facts.<sup>13</sup> It follows that if  $m$  is the actual cut-off point for ‘tall’, there are close-by possible worlds in which the use facts differ only slightly from the actual world, but in which the cut-off point for ‘tall’ will be slightly more or less than  $m$ . In such worlds, the proposition expressed by the sentence ‘A person is tall if and only if they are more than  $m$  meters in height’ is a necessary falsehood. An actual assertion of this sentence will thus violate MAS.

Whatever one thinks about the importance of making assertions that conform to MAS, the notion is insufficient for our purposes. What we would like is a notion of safety that applies to beliefs rather than assertions. After all, our aim is to explain our lack of knowledge and knowledge is a species of belief, not of assertion.

One might hope to bridge the gap between assertion and belief by adopting some view about the metaphysics of belief which ties the two together. For example, one might argue that having a belief amounts to a disposition to assent to a certain sentence; or that having a belief amounts to having a certain sentence written into one’s “belief box”; or that having a belief amounts to (in some sense) accepting a certain sentence. One could then attempt to say that a certain belief constituted in the relevant way by a sentence  $s$  is true if and only if were the agent to assert  $s$ , that assertion would have been safe.

We are pessimistic about the prospect of bridging the gap in this way. For a start, we are unsympathetic to the kind of view regarding beliefs of the sort presented above: a belief, we think, is constituted by a wide range of linguistic and non-linguistic dispositions. In addition, we are skeptical about the kind of suggested counterfactual analysis for tying belief to assertion (consider for example a case where the closest possible world in which one asserts  $s$  is one in which  $s$  has a radically different content than its actual content).

But whether or not one accepts any of these views regarding the metaphysics of beliefs, we still face familiar problems regarding the semantics of *belief ascriptions*. It is clear that the current discussion requires not only talking of agents’ having beliefs in general, but of them having a particular belief *that p* (e.g. a belief that a person is tall if and only if

they are more than  $m$  meters in height). And even given a view that ties an agents' belief directly to one particular sentence  $s$ , it is far from clear how one relates the sentence  $s$  to a sentence 'p' appearing in an adequate propositional attitude ascription of the form 'X believes that p'. It is clear that we do not want to be so strict as to require that  $s$  be *identical* to 'p': after all we can correctly characterize a mono-lingual Swahili speaker as believing that the sun is shining, even if they have no disposition to assent (or accept, or so forth) the English sentence 'The sun is shining'. On the other hand, we do not want to be so liberal as to require merely that  $s$  express the same coarse-grained proposition as 'p' does: after all, an ascription of the form 'Louis Lane believes that Clark Kent can fly' can be inadequate, even if Louis Lane has a disposition to assent to a sentence  $s$  with the same coarse-grained content.<sup>14</sup> Note that the latter point is particularly important for Williamson's purposes. This is so because if  $m$  meters is the sharp cut-off height for tallness then on a coarse grained enough notion of propositions the sentence 'tall means more than  $m$  meters in height' expresses the same proposition as the sentence 'tall means tall'. But as we shall see, Williamson wants to allow that one can know the latter without knowing the former.

Fortunately, we think we can avoid some of these hairy issues while still offering a charitable interpretation of Williamson. We will help ourselves only to the following assumptions. Whatever the metaphysics of belief is, beliefs have propositional content. A necessary condition on the truth of an ascription of the form 'X believes that p' is that X has a belief with the same propositional content as the proposition that 'p' expresses. And if propositions are fairly coarse-grained (in particular if they are something like Russellean or Stalnakerian propositions), sameness of content may not be a sufficient condition for the truth of the ascription.<sup>15</sup> We will also help ourselves to the following locution: *It is adequate to describe X's belief using 'p'*. Roughly speaking, what we mean by this locution is that the sentence (or a relevant utterance of the sentence) 'X believes that p' is true. In fact, relative to the actual world, X has a belief that can adequately be described using 'p' if and only if X believes that p. Relative to another possible world  $w$ , X can have a belief that can adequately be described using 'p' only if X believes that q, and where 'p' in  $w$  expresses the proposition that q. What complicates matters (and

makes our above characterization only rough) is that in other possible worlds (even close ones) the phrase ‘X believes that...’ might have a different meaning than its actual meaning. What we want is that if in a world  $w$  the phrase ‘\*\*\*’ means ‘X believes that..’ then in  $w$  it is adequate to describe X’s belief using ‘p’ if and only if in  $w$  the sentence (or a relevant utterance of) ‘\*\*\* p’ is true. More quibbles can be raised about how exactly to formulate this idea, but our intention, we hope, is clear.

With this background in place we can finally bridge the gap between assertion and belief, and suggest the following safety condition for belief on Williamson’s behalf:

**Meta-linguistic Belief Safety:** An agent X’s belief that p is *meta-linguistically belief safe* (MBS), if and only if there is no close-by possible world  $w$  such that (i) in  $w$ , X has a belief that can adequately be described in  $w$  using ‘p’. (ii) in  $w$ , that belief is false.<sup>16</sup>

Assuming that a belief’s being meta-linguistically safe is a necessary condition on knowledge we can see how Williamson’s account of our ignorance of cut-off points is supposed to work. Suppose Jane believes that a person is tall if and only if they are more than  $m$  meters in height. Let us also suppose that Jane’s belief is true. Jane’s belief will nevertheless fail to count as knowledge. Why? There is a close-by possible world  $w$  in which the use facts for ‘tall’ are slightly different, and hence the sharp cut-off point for ‘tall’ is slightly different than  $m$ . Now presumably, in  $w$ , Jane’s brain state, behavioral patterns, and dispositions are very similar (or perhaps even identical) to her actual ones. In particular, she still has a disposition to assent to ‘A person is tall if and only if they are more than  $m$  meters in height’ (assuming she has a disposition to assent to this sentence in the actual world), she still has this sentence “written in her belief box” (assuming she has it written in the actual world), and so forth. Given this, it would be adequate to describe Jane’s belief in  $w$  using ‘A person is tall if and only if they are more than  $m$  meters in height’, though of course by using this description in  $w$  we will be ascribing to Jane a belief with a different content than her actual belief. Now since in  $w$ ,  $m$  is *not* the correct cut-off point for ‘tall’, the content of Jane’s belief in  $w$  is false. So Jane’s actual



belief violates MBS. Since Jane's actual belief violates a necessary condition for knowledge, she fails to know.

A particularly compelling feature of this explanation is that it seems compatible with the fact that (in some sense) we do know the meaning of 'tall'.<sup>17</sup> For example, we do know that 'tall' means tall. Such knowledge does not violate MBS, because in close-by possible worlds we will have similar (or perhaps even identical) dispositions, brain states, etc. to our actual ones. In particular, we still have a disposition to assent to sentences such as "'tall' means tall'. Plausibly, this entails that in every close-by world  $w$  we will have a belief that can adequately be described in  $w$  using "'tall' means tall'. For sure, the content of the belief that is thus described in  $w$  is different than the content of our actual belief concerning the meaning of 'tall'. But the point is that the content of our belief in  $w$  is true in  $w$ , rendering our actual belief safe.

In spite of these compelling features we will now proceed to argue that appealing to MBS is an unsatisfactory way to explain our ignorance of the cut-off points of vague terms. First, we will argue that even if MBS were a necessary condition on knowledge it is not clear that Williamson's explanation would go through and second, we will argue that MBS is not a necessary condition on knowledge.

### **§3 Why MBS does not help**

Suppose that MBS is indeed a necessary condition on knowledge. Is it clear that it can be applied in order to show that we do not and cannot know the cut-off points associated with vague expressions? Suppose X comes to form a true belief that can be correctly described by saying 'X believes that every person is tall if and only if they are more than  $m$  meters in height'. According to Williamson this belief cannot constitute knowledge, because it violates MBS. But why must X's belief violate MBS? As we have seen, Williamson's explanation relies on the claim that in some very close-by world  $w$  in which the meaning of 'tall' differs only slightly from its actual meaning, X has such similar dispositions and so forth to her actual ones, so that X has a (false) belief that can

adequately be described in  $w$  using ‘A person is tall if and only if they are more than  $m$  meters in height’.

We shall not take issue with the claim that *if* in a close-by possible world  $w$ , X has similar dispositions, and so forth to her actual—in particular a disposition to assent to the sentence ‘A person is tall if and only if they are more than  $m$  meters in height’—then in  $w$  X has a belief that can adequately be described in  $w$  using this sentence. Rather, we wish to question the thought that there must be a close-by possible world in which X has such dispositions.

We start by asking ourselves how X came to form her actual true belief that a person is tall if and only if they are more than  $m$  meters in height. If X just *guessed* the cut-off point for tallness, then plausibly in some close-by possible worlds she would make the analogous (but false) guess. But one does not need MBS to tell us that guessing is not a good method for obtaining knowledge. Clearly, if X was to form a belief that is even a potential candidate for knowledge, we can assume that X arrived at this belief using some careful and sensitive deliberation or some other kind of reliable method.<sup>18</sup> But if X formed her belief via some reliable method, it is plausible to think that in every close-by possible world in which the cut-off point for ‘tall’ is  $m^*$ , X will apply the same reliable method and will arrive at the correct conclusion that she would put by saying ‘A person is tall if and only if they are more than  $m^*$  meters in height’ and not by saying ‘A person is tall if and only if they are more than  $m$  meters in height’.<sup>19</sup> Thus without telling us more about the method used by X, we have no reason to think that her belief violates MBS. Call this the *Belief-formation Challenge*.<sup>20</sup>

What can the Williamsonian say in response to the Belief-formation Challenge? One response might be to argue that it is hard to imagine what kind of method X could possibly use to yield a reliable prediction (rather than just a luckily true guess) of the cut-off point for tallness. But this response seems highly unsatisfactory because it reduces Williamson’s explanation of our ignorance to merely shrugging our shoulders and saying that one cannot know the cut-off points associated with vague expressions because we

cannot imagine how one could ever come to know such a thing. In particular, interpreting Williamson's explanation as a mere shrug-of-the-shoulders argument would seriously undermine the role his explanation plays in responding to those who think that our inability to imagine how one might come to know the cut-off points is good evidence that such cut-off points do not exist.

A much more promising response the Williamsonian might offer to our Belief-formation Challenge is the following (call this *the Over-sensitivity Response*): as fine-grained as our measurement instruments become there will always be some subtle physical differences which are indiscernible to us. Take a possible world  $w$  which differs from the actual world in the use facts of 'tall' so slightly, that the distinction between the use facts in  $w$  and the actual use facts is indiscernible to us. According to Williamson, the meaning of 'tall' in  $w$  will be slightly different than in the actual world. Assume that the cut-off point for 'tall' in the actual world is  $m$ , while in  $w$  it is  $m^*$ . Now suppose that X managed to come up with a sensitive enough method that while in the actual world she forms an ' $m$ -belief' about the cut-off point, in  $w$  she forms an ' $m^*$ -belief' about the cut-off point. Since X can clearly notice the difference between an  $m$ -belief and an  $m^*$ -belief (e.g. because they are exhibited in clearly different assertions), it seems as though she can discriminate between the actual world and  $w$ , and hence discriminate between the use facts in the two worlds, in contradiction to our initial assumption. Does this not show that it simply cannot be the case that X forms a belief in such a sensitive manner, and thus that X's belief must violate MBS?

There are two reasons we have to think that the Over-sensitivity Response to our Belief – formation challenge does not work. First, even if X manages to form an  $m$ -belief in the actual world, but an  $m^*$  belief in another possible world  $w$ , this does not entail that X can extract back the use facts from the meaning facts on which they supervene. (This is especially true if X does not know the supervenience function of meaning on use, but even if she did, that would not help if the function were not one-to-one). At best, X can work out that the use facts in  $w$  are different than the actual use facts (though not in which respect), and even that is dubitable given that X is not really in a position to

compare both worlds at once. In short, X's forming the correct belief in each world does not entail that X has acquired some magical capability to discern some set of indiscernible physical use-facts.

Our second reason to think that the Over-sensitivity Response will not work is that we do not accept the assumption that in worlds in which use facts differ indiscernibly from the actual use facts, the meaning of 'tall' must be different from its actual meaning. Of course, a general sorites-like argument will show that for any arbitrarily small difference, there must be two possible worlds the use facts of which differ only by that amount, but in which the meaning of 'tall' differs: if there weren't such pair of worlds, we could have constructed a long chain of worlds where in each adjacent pair of worlds the use facts are so similar that they differ only by the said amount, but where in the two worlds at the extremes of the chain the use facts were so substantially different that there will be an obvious difference in meaning of 'tall'. (In fact, this argument has nothing to do with vagueness: it is equally true of the meaning of any word in the language). But the problem is that there is no particular reason to think that the *actual* world is one of a pair of adjacent worlds in such a series in which the meaning shifts.

How could it be that while some worlds do have close-by worlds containing meaning shifts, many worlds do not? A helpful way to illustrate this is to consider by analogy shifts of meaning over time rather than between worlds. Consider a non-vague word that changes its meaning over time via a gradual shift in use (rather than, for example, by an explicit new stipulation). At least one plausible story of how this happens (though by no means the only story available) is that the meaning of a word at a time  $t$  is determined by the totality of use facts of that word at all times prior to  $t$ . Initially, the word has meaning A, but more and more use facts accumulate that push it towards having meaning B. Sorites-like arguments show that at some point some very tiny extra shift in use serves as the last straw, tipping the meaning of the word to meaning B. This means that there is a time  $t$  such that just before  $t$  the meaning of the word was A, and just after  $t$  the meaning is B. Still, at most times (those before  $t$  and those after  $t$ ) the meaning of the word is stable. Most cases of small shifts in use are not such that they entail a shift in meaning.

The analogous picture for the possible world case is now clearer: there are many clusters of possible worlds such that in them small differences in the use facts of a word are not sufficient to entail a difference in meaning. It is only ‘edge-worlds’ – those worlds in which the use facts are really close to the tipping the meanings - that have very close-by possible worlds in which the meaning of the word is different. But we have no reason to think that the actual world is an edge-world (at least no reason to think that it’s an edge-world with respect to every single vague expression).

In response to our second reason above for rejecting the Over-sensitivity Response, the Williamsonian might argue that while this story is plausible for non-vague words which have quite stable meanings, matters are different for vague words: one thing that makes vague expressions special is that their meaning is extremely sensitive to small shifts in use. This however is an unreasonable claim. Even if vague expressions have somewhat less stable meanings than other expressions, there are good reasons to think that their meanings must be at least somewhat stable through shifts of use. Take for example the case of propositional attitude reports. Jack utters the sentence ‘Jill is tall’ and Joe reports him a minute later by saying ‘Jack said that Jill is tall’. Intuitively, his report is perfectly correct. But on the plausible assumption that a necessary condition on the correctness of the report is that Jack and Joe’s utterances have the same content ‘tall’ could not shifted its meaning in spite of the new use facts added in the minute that lapsed between Jack’s initial utterance and Joe’s utterance in the report.<sup>21</sup> One can of course resist either the intuitive judgment that Joe’s report is correct, or the plausible assumption that sameness of content is necessary for the correctness of the report – but rejecting either is non-trivial price to pay for the view. And similar problems arise when we consider communications between subjects (Joe utters a sentence  $s$  with meaning  $m$ , but by the time Jill hears him and processes the sentence it has meaning  $m^*$ ), and memory (Joe stores in his memory a fact using the sentence  $s$  which means  $m$ , but when he wants to recall the fact later, the sentence  $s$  now means  $m^*$ ).<sup>22, 23</sup>

One way for the Williamsonian to try and accommodate these considerations in favor of stability would be to concede that vague expressions have relatively stable meanings over

time, but to argue that they do not have correspondingly stable meanings across close possible worlds.<sup>24</sup> This position might be motivated as follows: perhaps the meaning of a vague expression  $e$  at a time  $t$  is determined in a large part parasitically on its meaning at previous times. That is to say, unless the use facts at  $t$  make it radically implausible for  $e$  to retain its previous meaning, it will by default retain the meaning it previously had. This will ensure that most of the time the meaning of  $e$  will be stable over time, but on the face of it would have no consequence to the meaning of  $e$  in close-by possible worlds because the meaning of  $e$  is in no way parasitic on its meaning in other worlds. Our problem with this proposal is that even if it were adopted, it would not necessarily help for Williamson's purposes. The problem is that it is plausible to think that in many cases close-by possible worlds have a shared history. Compare for example the actual world, with a world  $w$ , which is just like the actual world until May 1<sup>st</sup>, 2008, but where on May 2<sup>nd</sup>, 2008 there were three new uses of 'tall' in  $w$  that did not occur in the actual world. On Williamson's account, these differences in use are supposed to entail that after May 2<sup>nd</sup> the meaning of 'tall' in  $w$  is slightly different than its actual meaning. But according to the above proposal this is not so: up to May 1<sup>st</sup>, the meaning of 'tall' in the actual world is identical to its meaning in  $w$  because the use facts in the two worlds are identical. And after May 1<sup>st</sup> the meaning in each of the two worlds is supposed to be parasitic on its old meaning (in particular, it is not going to be very sensitive to the new use facts in  $w$ ) so that the meaning in both worlds remains identical. Thus on this proposal, the only way to ensure that the meaning in a world  $w$  and in the actual world differ, is by having different histories since the dawn of time. No close-by possible world which shares some initial history with the actual world can make for a violation in safety, which seems implausible.

We thus conclude that even discernible differences in use facts are not always sufficient to entail a difference in meaning of vague expressions. To clarify: we are not arguing that worlds in which the use facts are different enough to entail a difference in meaning are too far to count as close-by worlds. We are merely arguing that it is not always the case that we cannot discriminate between the use facts of those close-by worlds and ours. It thus seems that the claim that any true belief about the cut-off point for 'tall' would violate MBS at best requires further defense.

#### **§4 Why MBS is not a necessary condition on knowledge**

In this section we argue that MBS is not a necessary condition on knowledge. We will do so by presenting six cases in which a subject has a belief that on the one hand violates MBS (i.e. where the belief is not meta-linguistically safe), but on the other hand constitutes knowledge.

##### **Case 1: (possibility of) fake barns**

Suppose X looks at a (real) barn and forms a belief that he expresses using the sentence ‘That is a barn’. Now suppose that there is a close-by possible world  $w$  in which X is in a corresponding situation, looking at a fake barn. For example, we can imagine that prior to placing the real barn the barn-placer tossed a coin in order to decide whether to place a real or fake barn in that location. In the actual world the coin toss determines that a real barn is to be placed, while in the close-by possible world  $w$ , the coin toss determines that a fake barn is to be placed. In  $w$ , X is fooled by the fake barn and thus forms the false belief that he would express in  $w$  using the sentence ‘That is a barn’.

X’s belief violates MBS in this case. X believes that that is a barn. Yet there is a close-by possible world  $w$  in which X has a belief that can adequately be described in  $w$  using ‘That is a barn’, and that belief is false  $w$ . One might try to contest the claim that  $w$  should count as close-by on the grounds that close inspection of the fake barn would allow one to recognize that the barn is fake, and thus  $w$  is not indiscernible from the actual world. But given our discussion in §3, this will not do: Williamson wants worlds in which small shifts in meaning to occur to count as close-by, yet as we have argued such worlds often differ from the actual world in discernible manners.

The more interesting question is whether X’s belief should count as knowledge. We chose to open with this case precisely because intuitions are likely to differ here – highlighting the initial plausibility of MBS as a condition on knowledge on the one hand, and the potential problem with it on the other. For what it’s worth, our own intuition is that X’s belief should (other things being equal) count as knowledge. In particular, note that this case is much weaker than the standard fake barn cases: in the standard case, one

is faced with a real barn, in a situation in which there are many fake barns in one's immediate vicinity. Intuitions vary already with respect to that initial case,<sup>25</sup> but in our case there need not be a single fake barn in one's vicinity. All that is required is that one *could* have (easily) been in the vicinity of a fake barn.<sup>26</sup> If our intuition is correct then this is a case where X has a belief that violates MBS but nevertheless constitutes knowledge. However, if one does not share our intuition on this case we hope that (at least some of) the cases we discuss below will constitute clearer examples.<sup>27</sup>

### **Case 2: Table case**

Suppose Z is carrying tables into a room. She looks at a very large table and believes (correctly) that that table will not fit through the door. However, Z could easily have looked at a medium sized table instead. In particular, there is a close-by possible world  $w$  in which Z looks at a medium size table which could just about fit through the door. Let us suppose that in  $w$ , when looking at the medium sized table, Z erroneously believes that it will not fit through the door.

This case constitutes a counterexample to the claim that MBS is a necessary condition on knowledge. Z believes that that table will not fit through the door. However, in the close-by world  $w$ , Z has a belief that can adequately be described in  $w$  using the sentence 'That table will not fit through the door'. And Z's belief in  $w$  is false, thus violating MBS. However, it is clear that Z's actual belief that that table (namely, the large one) will not fit through the door does constitute knowledge. The table is very large and obviously will not fit through the door. Z's skill at judging such matters, though not perfect (which is clear from the fact that she would erroneously judge the case of the medium size table wrongly), can nevertheless be excellent. And because the very large table is not a borderline case, Z's judgment is bound to be accurate. Z knows that that table will not fit through the door.

One might try to argue that Z's belief does not after all constitute a violation of MBS by appealing to the notion of methods. Perhaps, the thought goes, a crucial ingredient in what makes a world  $w$  count as close-by (at least for the purposes of MBS) is that in  $w$



one forms the relevant belief using the same method that one uses to form the belief in the actual world. Thus although the world  $w$  in which  $Z$  picks the medium-size table is close in the sense that it could have easily been actual, it is not one in which  $Z$  forms her belief using the same method as in the actual world and thus cannot be a world which counts towards a violation of MBS.

We find this appeal to the notion of methods unconvincing. First, at least on the face of it, it seems that  $Z$  *does* use the same method in  $w$  as in the actual world (roughly: looking at a table and estimating whether it will fit). Of course one could find some fine-grained description of the methods in question according to which  $Z$  uses a different method in the actual world and in  $w$  (for example: in the actual world  $Z$  uses the method ‘look at a *large* table and estimating whether it will fit’, while in  $w$  she uses the method ‘looking at a *medium* table and estimating whether it will fit’). But in so far as we have any intuitive grasp on what the correct way to describe the method in question is in this case, the latter suggestion seems highly unnatural. Second and relatedly, the problem of individuating methods is a serious one for this notion of safety, and as we briefly argue in §5 the most promising solution to this problem - namely individuating methods in terms of knowledge - is one to which Williamson cannot appeal in this context. Finally, note that if in order to address our suggested counterexamples to MBS one needs to appeal to a fairly fine-grained notion of methods, this puts even more pressure on the Belief-formation Challenge we posed in §3: after all, if the only worlds that count as close-by are ones in which one forms the belief concerning the cut-off points in *exactly* the same manner as in the actual world, then it is even harder for Williamson to prove that there must always be close-by worlds in which one’s method leads one astray in determining the cut-off points. We thus conclude that appealing to the notion of methods in this case (as well as in the other cases we discuss below) will not be sufficient to defend MBS.

Finally, one might worry that cases 1 and 2 are somehow unacceptable because of their use of the indexical expression ‘that’. It is not clear to us why such use of an indexical would be illegitimate in this context. In particular, we take it that in so far as one is inclined to have the opposing intuition than ours in case 1 (namely if one thinks that  $X$

does *not* know that that is a barn), one would view the appeal to MBS as applied to the indexical ‘that’ to be a compelling (at least *prima facie* compelling) explanation of X’s ignorance. But at any rate, our point does not depend on this issue: none of the following cases rely on the use of indexical expressions.

### **Case 3: Bank Case**

Suppose that Y believes that there is a money-bank in Shrewsbury. His belief is true, justified, non-Gettiered, and so forth: i.e. a *prima facie* good candidate for counting as knowledge. Suppose, though, that Y could have easily formed the belief that there is a riverbank in Shrewsbury. For example, suppose that at the same time that Y was forming his money-bank belief, he receives an equal amount of compelling evidence in favor of the riverbank belief, and that he is equally interested in the question of whether there is a riverbank in Shrewsbury as in the question of whether there is a money-bank in Shrewsbury. There is thus a close-by possible world  $w$  in which Y forms the belief that there is a riverbank in Shrewsbury. Suppose also, though, that in  $w$  that belief is false (in fact, we can assume that there is no riverbank in Shrewsbury both in  $w$  and in the actual world—the only difference between  $w$  and the actual world is which belief Y forms).

Y’s money-bank belief violates MBS in this case. Y believes that there is a bank in Shrewsbury. Yet there is a close-by possible world  $w$  in which Y has a belief (the riverbank belief) which can adequately be described in  $w$  using ‘There is a bank in Shrewsbury’, and this belief is false in  $w$ . Yet it seems highly implausible to say that just because Y could have easily formed the riverbank belief, his money-bank belief does not count as knowledge.

Again, arguing that  $w$  is not a close-by world because Y can discern the case where he forms a money-bank belief from the case where he forms a riverbank belief will not help. Not only (as we have argued) are there close-by worlds which are discernable, but also we can vary the case so that Y’s belief violates MBS in virtue of a world  $w$  that is *identical* to the actual world. Suppose that in the actual world Y forms *both* the riverbank and the money-bank belief. Y’s belief would still violate MBS: Y believes (and knows)

that there is a bank in Shrewsbury (namely, a money bank). But there is a close-by world  $w$  – namely the actual world – in which Y has a *false* belief that can adequately be described using ‘There is a bank in Shrewsbury’ (namely, Y’s riverbank belief). So Y’s belief violates MBS, where the only constraint on the similarity of worlds is that the actual world be close-by to itself.

One might nevertheless feel that there is something unacceptable about this case, perhaps having to do with the fact that the two uses of ‘bank’ do not even have similar meanings and that Y can clearly discern the two ‘bank’-beliefs. It is not clear to us why such use of a word with two non-similar meanings would be illegitimate in this context. But at any rate, our point does not depend on this issue: the next case is similar to this one but, involves a word which has two very similar meanings.

#### **Case 4: Modified bank case**

Let us suppose the following about the word ‘bank’: ‘Bank’ is ambiguous between a looser sense of the word according to which building societies count as banks and a stricter sense of the word according to which building societies do not count as banks.<sup>28</sup> Suppose that Y learnt the term ‘bank’ in its loose sense, but could have easily learnt the term in its strict sense. Y is not aware that ‘bank’ has two senses, and in fact Y does not know that there is a difference between building societies and banks (understood in the strict sense). Y uses the term ‘bank’ in its loose sense simply because (roughly) the causal chain by which Y picked up the term from his community goes back to the loose sense. But this casual chain could have easily gone back to the strict sense, and Y would not have noticed the difference.

Walking down the centre of town, Y sees a branch of Nationwide (a building society). Y sees that Nationwide offers many services offered by a bank (mortgages, savings accounts, ATMs, etc.) and comes to form a belief that he would put using the words ‘There is a bank in the centre of town’, and that we can correctly report (appealing to the loose sense of the word) by saying that Y believes that there is a bank in the centre town. Suppose also that there are no strict-sense-banks in the centre of town. There is a (close)

possible world  $w$  which is just like the actual world, except that the word ‘bank’ in Y’s mouth picks out strict-sense-banks rather than loose-sense banks. In  $w$ , just like in the actual world, Y walks down the centre of town, sees the branch of Nationwide, and forms a belief that he would put using the word ‘There is a bank in the centre of town’. (Recall that Y is not aware that there is such a thing as a building society and that the word ‘bank’ as he is learnt it does not strictly speaking apply to these). In  $w$ , Y has a belief that can adequately be described using ‘There is a bank in the centre of town’ (appealing to the strict sense of the word, which is one of its two senses in  $w$  as well). But this belief is false in  $w$  because Nationwide is a building society, not a strict-sense-bank. So Y’s belief violates MBS.

It seems clear, however, that Y’s belief can constitute knowledge. Y is very well justified in believing that there is a (loose-sense) bank in the centre of town. Y may know full well that any business that offers the services that Nationwide does is a (loose-sense) bank. Thus in spite of the fact that two uses of ‘bank’ are similar here, and that Y cannot discern the case where he has a ‘loose-sense-bank’ belief from a ‘strict-sense-bank’ belief, this case constitutes another counterexample to MBS’s being a necessary condition on knowledge.

One might worry that the appeal to ambiguity is somehow illegitimate. But again, our case does not rest on this issue either. The next two cases involve neither indexicality nor ambiguity.

#### **Case 5: H<sub>2</sub>O vs. XYZ**

Suppose that there are two kinds of water-like substances on earth: XYZ and H<sub>2</sub>O. Since to the layman the two seem exactly the same, it is assumed by everyone that there is just one water-like substance, and that this substance is called ‘water’. Now let us suppose that as it happens the totality of use facts (some of which have to do with the exact distribution of H<sub>2</sub>O and XYZ in the world) entail that the word ‘water’ picks out H<sub>2</sub>O. Suppose also, though, that had the distribution of H<sub>2</sub>O and XYZ been ever so slightly different, the word ‘water’ would have picked out XYZ. (As a toy model of this you can

imagine that the reference of the word is determined merely by which substance is more common on earth. In the actual world there is one molecule more of  $H_2O$  than of XYZ, but it could have very easily been the case that there were one molecule more of XYZ than of  $H_2O$ ).

Suppose that scientists now decide to investigate the nature of water. They pick various samples of what they think is water, and which as it happens *is* water, i.e. samples of  $H_2O$ . They investigate these samples in their laboratory and discover that they have the chemical constitution of  $H_2O$ . They therefore come to form a belief that they express using the words 'Water is  $H_2O$ '.

Now let us consider the same scientists in a very close-by possible world  $w$ , one in which due to a slightly different distribution of the molecules on earth, 'water' means XYZ. Let us suppose that the scientists in  $w$  perform exactly the same experiments as they did in the actual world using exactly the same samples of water-like substance as in the actual world. Since all the water-like samples they pick are samples of  $H_2O$ , and since their experiments correctly reveal that these are samples of  $H_2O$ , the scientists in  $w$  come to form a belief that they express using the words 'Water is  $H_2O$ '.

First, we argue, the scientists (actual) belief violates MBS. The scientists believe that water is  $H_2O$ . However, in the close-by world  $w$  they also have a belief that can adequately be described using 'Water is  $H_2O$ '. But since the content of their belief in  $w$  is that XYZ is  $H_2O$  (or some more fine-grained content that entails this), their belief is false in  $w$  making their actual belief that water is  $H_2O$  violate MBS. Now one might question the claim that the scientists' belief in  $w$  can adequately be described in  $w$  using 'Water is  $H_2O$ '. True, the argument goes, the scientists in  $w$  have a disposition to assent to the sentence 'Water is  $H_2O$ ' and to dissent to the sentence 'Water is not  $H_2O$ '. But they also have some other dispositions that tell against the fact that in  $w$  one might truly say 'They believe that water is  $H_2O$ '. For example, when the scientists examine a sample of XYZ in  $w$  they still classify it as XYZ and not as  $H_2O$ . But recall that if one were to report the scientists' belief in  $w$  using 'Water is  $H_2O$ ' one would be ascribing to them the belief that

XYZ is H<sub>2</sub>O (or some more fine-grained proposition that entails this) - an ascription that does not sit well with their disposition to refuse to classify XYZ samples as being H<sub>2</sub>O. In response to this argument we do not necessarily need to settle the tricky question of how to weigh up conflicting dispositions. All we need to do is to add a few details to our case. Suppose that (both in the actual world and in  $w$ ), after conducting the initial set of experiments, the scientists decide that their experiments are completely conclusive and that they do not need to test any more samples of watery-stuff. This means that whenever they encounter a sample of XYZ (or for that matter H<sub>2</sub>O) in the future, they simply assume it is water, and since they also assume that water is H<sub>2</sub>O, they (falsely) conclude that the sample in question must be H<sub>2</sub>O. In short, both in the actual world and in  $w$  they are disposed both to assent to sentences such as 'Water is H<sub>2</sub>O', and to treat particular samples of XYZ as if they were H<sub>2</sub>O. Thus we should have no hesitation in claiming that in  $w$  they have a belief that can correctly be described using 'Water is H<sub>2</sub>O'.

Next, we argue that the scientists' (actual) belief should count as knowledge. The scientists obtain their belief that water is H<sub>2</sub>O by what seem to be standard sound scientific method: they test large samples of water in the lab, and conclude that these samples are H<sub>2</sub>O. One could try to argue, at least if one assumes our toy model where there is an almost equal distribution of H<sub>2</sub>O and XYZ molecules on earth, that the scientists were just lucky to have tested H<sub>2</sub>O samples rather than XYZ samples. Had they tested XYZ samples they would have reached the false conclusion that water is XYZ, rendering their actual belief unsafe (according to some different formulation of safety from MBS, one yet to be specified). But whether or not we are troubled by this alleged violation of safety, we can specify the case so that the scientists were not merely lucky to test H<sub>2</sub>O rather than XYZ samples. For example, we could specify that XYZ samples are all in areas that are very far away from where the scientists live. (Recall the scientists assume that there is only one water-like substance, so they have no reason to make a special effort to seek far out samples of the water-like stuff). Or alternatively we could specify that XYZ is always located at the bottom of lakes, whereas the water-like samples are taken from the top of the lakes. In short, it is not hard to see how one can build the

case so that the scientists could not have easily picked out XYZ samples and were not merely lucky to have picked out H<sub>2</sub>O samples.

Another way to see why the scientists do know that water is H<sub>2</sub>O in this case is to consider the following scenario. Suppose that in the actual world there are two competing teams of scientists – the English and the French scientists. Both believe that all watery-stuff is of the same substance, and let us suppose that the use facts (including the distribution of H<sub>2</sub>O and XYZ) determine that both the English word ‘water’ and the French word ‘eau’ pick out H<sub>2</sub>O. The French and English teams conduct exactly the same experiments as we have described above, and reach a conclusion that the English scientists describe by saying ‘Water is H<sub>2</sub>O’ and the French scientists describe by saying ‘L’eau est H<sub>2</sub>O’: namely the scientists on both teams believe that water is H<sub>2</sub>O. Now suppose that there is a close-by possible world  $w$  in which a very slight shift in the distribution of the two substances entails that the English word ‘water’ picks out XYZ, but the French word ‘eau’ still picks out H<sub>2</sub>O. (For this we would need to slightly complicate our toy model for how the meanings of the words are related to the distribution of substances. For example, perhaps in  $w$  slight differences between the distribution of H<sub>2</sub>O and XYZ in England versus in France mean that the English the word means XYZ, while the French word means H<sub>2</sub>O). As before, in  $w$  both teams conduct the same experiments on the same samples of H<sub>2</sub>O.

Now one thing that should be uncontroversial about this scenario is the French scientists in the actual world know that water is H<sub>2</sub>O. Not only have they used sound scientific methods to reach a true conclusion, but they could not have easily gone wrong: in the close-by world  $w$  they use the same methods to reach an equally true conclusion which they express in  $w$  using the true sentence ‘L’eau est H<sub>2</sub>O’.<sup>29, 30</sup> So if one wanted to claim that the English scientists do not have knowledge in this scenario one must accept the claim that while the French scientists know that water is H<sub>2</sub>O the English ones do not. But this seems absurd: the two teams conduct exactly the same experiments and reach exactly the same conclusion. The only differences between them is subtle semantic facts about the stability of meaning of a certain word in their respective languages.

Investigating such subtle semantic facts should not be part of a scientific investigation into the nature of water, and failure to notice such semantic facts should not bar scientists from coming to know the chemical nature of water. We thus conclude that the English scientists do know that water is  $H_2O$ , in spite of violating MBS.

### **Case 6: Identical twins**

One might worry that the previous case somehow relies on some controversial claim about the semantics of natural kind terms. But our point does not depend on this issue either. In this final case, we will present a similar example to the previous one, but one which involves ordinary proper names.

Suppose that A and B are two identical twins. Many people refer to A by the name 'Joe', in such a way that determines that the name 'Joe' semantically denotes A. Suppose, though, that since A and B are identical twins, people often confuse them and accidentally refer to B by the name 'Joe'. Now if enough people had referred to B enough of the time using 'Joe', the name would have semantically denoted B, rather than A. So suppose that although as things stand 'Joe' refers to A, had slightly more people referred to B using 'Joe' the name would have denoted B. In short, 'Joe' denotes A, but could have easily denoted B.

Suppose that Jill is a friend of Joe (i.e. A), and has never encountered his twin brother (i.e. B). In fact, she lives very far away from the twin brother B and is under no risk of ever encountering him. She follows the use of her community in calling Joe (i.e. A) by the name 'Joe'. She knows Joe very well and believes that he is a nice guy. Her belief is true and based on all the right kind of evidence (knowing Joe well, observing him interact in various situations, testimony of other people, and so forth). It seems quite clear that Jill's belief should count as knowledge. Yet assuming that B is not a nice guy, Jill's belief violates MBS: there is a close-by world  $w$  in which (unbeknownst to Jill) 'Joe' refers to B. In  $w$ , Jill would still have a belief that she would express using the sentence 'Joe is a nice guy'. As above, one might worry that other dispositions entail that in  $w$  Jill nevertheless does not believe that B is a nice a guy, and therefore that in  $w$  we cannot



adequately describe her belief using ‘Joe is a nice guy’. But as above, all we need to is to add more details to the case. For example, suppose that Jill also believes that B is a nice guy (perhaps simply because she is assuming that A’s twin brother must also be nice). Or suppose that if Jill ever encountered B, she would confuse him with A, she would assume on the basis of that that he is a nice guy. So with enough details added it should be unproblematic that in  $w$  Jill has a belief that can adequately be described in  $w$  using ‘Joe is a nice guy’. But, we assume, B (both in the actual world and in  $w$ ) is an evil-twin, and hence Jill’s belief in  $w$  is false, thus violating MBS. We conclude that MBS cannot be a correct necessary condition on knowledge.

### **§5 Knowledge of meta-linguistic cut-off points**

Given our rejection of MBS as a necessary condition on knowledge one might raise the following objection to our position.

Let us call beliefs such as ‘A person is tall if and only if they are  $m$  meters in height’ beliefs about *non-meta-linguistic cut-off points*, and beliefs such as ‘The cut-off point for ‘tall’ is  $m$  meters’ beliefs about *meta-linguistic cut-off points*. Suppose that X came to know a non-meta-linguistic cut-off point claim, for example that a person is tall if and only if they are more than  $m$  meters in height. It seems that she could then reasonably infer from this the corresponding meta-linguistic-cut-off claim, i.e. that the cut-off point for ‘tall’ is  $m$  meters. Now it seems like the meta-linguistic cut-off belief that she forms on the basis of this reasonable inference from a known premise should count as knowledge.

But now comes the objection: suppose Williamson is correct in thinking that there is a close-by possible world  $w$  in which the cut-off point for ‘tall’ is  $m^*$  rather than  $m$ , and that in  $w$  X still has a belief that she would express in  $w$  by saying ‘A person is tall if and only if they are  $m$  meters in height’. It seems that in  $w$ , X would make a similar inference and reach the meta-linguistic conclusion that she would express in  $w$  by saying ‘The cut-off point for ‘tall’ is  $m$ ’. But this conclusion is false in  $w$ . The upshot is that X’s (actual) meta-linguistic cut-off belief violates a much more straightforward condition on

knowledge, namely Simple Safety. After all, the content of the meta-linguistic claim ‘ $m$  is the cut-off point for ‘tall’ does not vary between the actual world and  $w$ , and thus it is the same belief that  $X$  has in the two worlds. But that belief is true in the actual world and false in  $w$ , so Simple Safety is violated which seems to entail that  $X$  does not know the meta-linguistic claim after all, contrary to what we have supposed above. Moreover, one might view this line of thought as an argument in favor of taking MBS as a necessary condition on knowledge after all: it is because meta-linguistic cut-off beliefs violate Simple Safety that the corresponding non-meta-linguistic cut-off beliefs violate MBS. Consequently, the argument goes, violations of MBS should not count as knowledge because they entail corresponding violations of Simple Safety.<sup>31</sup>

We have several things to say in response to this objection. First, the argument proceeds by supposing (by way of contradiction) that one can know the non-meta-linguistic cut-off point claims. But we have not argued anywhere in this paper that we have such knowledge: we merely argued that if indeed we cannot have such knowledge, then violations of MBS cannot be the correct explanation for this ignorance.

Second, as we have argued in §3, it is not clear that our non-meta-linguistic cut-off claims violate MBS. For corresponding reasons, it is not clear that the meta-linguistic cut-off claims violate Simple Safety. As we have argued, if one has a true belief that the (non-meta-linguistic) cut-off point for tallness is  $m$ , and if that belief was formed via an appropriately sensitive method, then in the close-by world  $w$ , one would form the belief that the (non-meta-linguistic) cut-off point for tallness is  $m^*$ . But then using an inference such as above, one would in the actual world form the true belief that the (meta-linguistic) cut-off point for ‘tall’ is  $m$ , and in  $w$  form the equally true belief that the (meta-linguistic) cut-off point for ‘tall’ is  $m^*$ , and so one would *not* violate Simple Safety after all.

Finally, although it is a more plausible condition than MBS, we do not think that Simple Safety is an adequate necessary condition on knowledge either, and for familiar reasons. Suppose for example that Jill looks at her hands and forms the very reasonable belief that

she has hands. Suppose also that as she was forming this belief a very heavy stone was falling right next to her and just missed her. Plausibly, there is a very close-by possible world  $w$  in which the stone does not miss her. Suppose that in  $w$ , the stone hits Jill and has the following gruesome effect: first, it chops off her hands, and second, it knocks her on the head in a way that makes her hallucinate that she has hands. Her actual belief violates Simple Safety in this case: there is a close-by possible world in which Jill believes that she has hands but that belief is false. And yet it seems very plausible that her actual belief should nevertheless count as knowledge.

The equally familiar response to cases like this is to change the formulation of Simple Safety so as to include some talk of methods: the reason the close-by possible world  $w$  above is not sufficient to undermine Jill's actual knowledge is that in that in  $w$  Jill forms her belief using a very different method than the actual one (hallucination rather than perception). One might thus propose to replace the appeal to Simple Safety in the above argument with an appeal to the following safety condition:<sup>32</sup>

**Methods Safety:** An agent's belief in the proposition that  $p$  is *methods safe* if and only if there is no close-by possible world  $w$  in which (a) she believes that  $p$  (b) she formed that belief using a relevantly similar method to the actual method she used to form the belief (c) It is false that  $p$ .

The problems with Method Safety are well-known: how does one adequately individuate methods? Note in particular that this issue comes up in this context even if one accepted that our non-meta-linguistic cut-off beliefs always violate MBS (i.e. if one forms an  $m$ -belief in the actual world, then one would still form  $m$ -beliefs, and not  $m^*$ -beliefs in close-by worlds). For even if one accepted this assumption, one would still face the question of whether inferring the meta-linguistic cut-off claim in the actual world is done using the same method as in other close-by worlds. In one case, we have an inference from a known belief and in the other, from an unknown, false belief. Aren't these methods relevantly different, showing that our non-meta-linguistic belief does not after all violate Methods Safety?<sup>33</sup>

Of course, one promising solution to the problem of individuation of methods is to take knowledge as a basic concept and interpret Methods Safety in whatever manner is needed to make it an adequate condition on knowledge. We have no objection to this move in general, but applying it will not help in this case, because it would undermine Williamson's claim that there is a safety based *explanation* of our ignorance of the cut-off points of vague predicates. Indeed, if knowledge is considered basic and safety is defined in terms of it, it is our knowledge (or ignorance) of a fact that explains the safety (or lack thereof) of our belief in that fact, and not *vice versa*. We thus conclude that the objection from meta-linguistic cut-off points does not undermine our rejection of MBS as an explanation of our non-meta-linguistic ignorance.

### **§6 Directions for a theory of Vagueness**

In this paper we have argued that Williamson's explanation for why we do not and cannot know the cut-off points of vague expressions is unconvincing. If our conclusion is correct, what are we to conclude about the project of giving an adequate account of vagueness? As far as we can see, there are four main routes one could take.

The first route is to reject the claim that vague expressions have sharp cut-off points. We find this to be the least promising route, not only because we find most of Williamson's arguments in favor of sharp cut-off points compelling, but also because, as Williamson points out, puzzles analogous to the one about the apparent unknowability of cut-off points seem to arise even if the view that there are sharp cut-off points is rejected.<sup>34</sup> For example, if one is a supervaluationist about vagueness, then one claims that the meaning of the word 'tall' is determined by a range of admissible precisifications. But which precisifications exactly are the admissible ones? We do not seem to know and it is hard to imagine how we would come to know this. If one is a fuzzy-logician then one claims that if Bill is a border-line case of 'tall' then 'Bill is tall' is true to a degree greater than zero and smaller than one. But which degree exactly? We do not seem to know and it is hard to imagine how we would come to know this fact. In short, the puzzle about why we seem to be unable to specify in precise terms what meanings of vague predicates are seems to arise even if one thinks those meanings are not "sharp".<sup>35</sup>

The second route is to reject the claim that the cut-off points of vague predicates are unknowable. Perhaps these cut-off points are merely *unknown* but not unknowable. If sharp cut-off points are knowable, then how might we come to know them? It is not even clear that we owe an answer to this question. Perhaps it is not unknowable whether Goldbach's conjecture is true, and yet we have no idea how we might come to know whether it is true. Moreover, we can offer some rough ideas about how we might come to know the sharp cut-off points associated with vague expressions. Here is one such crude story. Physical facts are discovered roughly as follows: first, one observes various directly observable facts. Then one theorizes and comes up with some general principles that are compatible with and explain these facts, though are probably not *entailed* by these facts. Plausibly, there are various competing principles that are compatible with and explain the facts, but one can choose between them using standard considerations such as simplicity, elegance, and so forth. Finally, we can use these general principles to come to know new particular facts which we are unable to observe. Perhaps, then, we can come to know the cut-off points of vague predicates by an analogous procedure. First, we collect a lot of data about vague predicates. The data consists both of use facts and of those meaning facts which we do know (For example, we do know that the sharp cut-off point for 'tall person' is larger than one meter and smaller than two and a half meters). We then come up with a general theory: let us suppose that this theory is a claim about the supervenience function of meaning on use. The supervenience function we postulate must be compatible with all the use/meaning facts we do know but would not be entailed by these facts. Rather we would pick it over competing hypothesis by standard considerations such as simplicity, elegance, and so forth. Finally, once we infer the supervenience function we can use it to come to know new particular facts such as whether 1.9 meters is above or below the cut-off point for 'tall' – facts that we are not able to directly observe about our language.

We realize this suggestion is quite radical, and we are not claiming that it can realistically be carried out (or indeed that if it could be carried out, it would in any way be an interesting project to do so). We are merely arguing that it is not utterly unimaginable that

*some* methods could yield us knowledge of cut-off points associated with vague expressions. That one cannot figure out these cut-off points by a quick introspection on one's language is hardly an argument that they are unknowable, just as many known facts of physics were not learned by a mere quick introspection of our immediate environment.

Suppose that cut-off points are knowable but unknown. Does this undermine the epistemicist view of vagueness? If cut-off points are, as we suggested above, knowable but extremely difficult to know then the epistemicist still has a response to the first objection raised against the view, namely that as much as we try, we cannot seem figure out what the cut-off points of vague predicates are. The response is that one has not tried hard enough. What about the second objection, the one concerning the distinctiveness of vagueness? Here matters are trickier. The mere knowability of the cut-off points is not a problem here (after all we can argue that once we come to know the cut-off points of a vague expression it ceases to be vague, and thus we no longer need to explain what is distinctive about it). But the question is whether we can characterize our ignorance of cut-off points of vague expressions (when we are ignorant of them) in an interesting way. It will not do to say that vague terms are terms for which we are ignorant of their meanings: in some sense we are not ignorant of the meanings of vague terms (we know that 'tall' means tall) and in another sense we are ignorant of the meaning of almost any non-vague expression (we do not know whether the meaning of 'prime' is such as to include 2,345,457 or whether the meaning of 'elm' is such as to include the tree in the back yard). The advantage of Williamson's explanation was that it was more specific than this: it explained the exact source of our ignorance as a result of slight shifts in meaning. Once this explanation has been rejected, it is not clear that we are left with a good way to characterize what is distinctive about vagueness. We leave it open whether or not this consequence should be particularly worrying for epistemicism.

A third route which explicitly endorses the claim that there is no special way to characterize the ignorance involved in vagueness, is one that states that sharp-cut off points are unknowable and there is simply no interesting explanation of this fact. According to the third route the idea that one requires an explanation for the

unknowability of a certain set of facts presupposes that the default is that facts are knowable, and that some special explanation is needed in those cases where facts are not knowable. According to the proponents of the third route this presupposition is false. Indeed, if anything, the cases that require explanation are those in which we *are* able to obtain knowledge. Thus no explanation is needed for why a certain set of facts in general, and facts concerning the sharp cut-off points of vague expressions in particular, are unknowable.<sup>36</sup>

The fourth and final route is to maintain that sharp cut-off points are unknowable, to accept the challenge of providing an explanation for this unknowability, but to respond to this challenge by providing a different explanation for the unknowability than Williamson's. One direction one could pursue is to try to find more careful formulations of safety conditions that would achieve this goal. We are pessimistic about the prospects of this direction but perhaps we could be proven wrong. A different strategy is found in Sorensen 2001. Sorensen argues that propositions about sharp cut-off points for vague properties have determinate truth *values* but lack *truthmakers* (Sorensen 2001, p. 176). This being so, such truths are 'epistemic islands' (p. 175). That is, because there are no grounds for the truth of propositions about the cut-off points of vague properties, there is no way of coming to know these propositions. A related idea is suggested by recent work (which one of the current authors is engaged in) on the notion of arbitrary reference.<sup>37</sup> The claim defended in this work is that it is possible to refer to a particular object arbitrarily or randomly. When one does so, the referent of the term does not (completely) supervene on the use of the term, and hence we do not and cannot know (in any informative way) what the referent is. For example, a stipulation such as 'Let *n* be an arbitrary number' results in '*n*' referring to a particular number (perhaps 245), but since there are possible worlds which (up to the point of the stipulation) the use facts are *identical* to the actual use facts, but in which the stipulation results in '*n*' referring to another particular number (perhaps 234), we do not and cannot know to which number '*n*' refers to, even upon the most careful examination of the use facts. Similarly, it may be that the meaning of 'tall' is picked out arbitrarily or randomly from a range of acceptable meanings, but because its resultant meaning does not supervene on use, we do not and

cannot know its exact meaning. Of course, there are many questions to be answered here (amongst them: why believe that it is possible to refer in this way? how is the range of acceptable meanings determined? How does this help with the problem of characterizing what is unique about vagueness?). This is not the place to address these questions or indeed to defend the ‘arbitrary reference’ view of vagueness. We merely mention this view as yet another direction for explaining why sharp cut-off points are unknowable.

We conclude, then, that even if one accepts Williamson’s main claim that vague expressions have sharp cut-off points, the project of specifying a complete and adequate theory of vagueness requires serious further work.<sup>38</sup>

## References

Breckenridge, W., and Magidor, O.: ‘Arbitrary reference’ (unpublished manuscript).

Gendler, T. and Hawthorne, J. (2005): ‘The real guide to fake barns: a catalogue of gifts to your epistemic enemies’, *Philosophical Studies* 124, pp. 331-352.

Gomez-Torrente, M. (2002): ‘Vagueness and margin for error principles’, *Philosophy and Phenomenological Research* 64, pp. 107-125.

Hawthorne, J. (2006): ‘Epistemicism and Semantic Plasticity’, in his *Metaphysical Essays*, OUP.

Salmon, N. (1986): *Frege’s Puzzle*, MIT Press.

Soames, S. (1989): ‘Direct reference, propositional attitudes and semantic content’, *Philosophical topics* 15, pp. 44-87.

Sorensen, R. (1988) *Blindspots*, OUP

Sorensen, R. (2001) *Vagueness and Contradiction*, OUP



Williamson, T., (1994): *Vagueness*, Routledge.

Williamson T. (1997a): 'Précis of vagueness', *Philosophy and Phenomenological Research* 57, pp. 921-928.

Williamson T., (1997b): 'Reply to commentators: [Horgan, Gomez-Torrente, Tye]', *Philosophical Issues* 8, pp. 255-256.

Williamson, T., (1997c): 'Reply to commentators', *Philosophy and Phenomenological Research*, 57, pp. 945-953.

Williamson T., (1999): 'Schiffer on the Epistemic theory of Vagueness', *Philosophical Perspectives* 13, pp. 505-517.

Williamson T. (2000), *Knowledge and its Limits*, OUP.

## Notes

---

<sup>1</sup> For a concise summary of these claims see Hawthorne (2006), pp. 185-6.

<sup>2</sup> Special scenarios such as those involving the testimony of omniscient beings are being excluded here (see for example Williamson (1997a), p. 926).

<sup>3</sup> “Once one has seen this point [the explanation of our ignorance], one can hardly regard our inability to find them [the sharp boundaries] as evidence that they do not exist. But if one has not seen the point, one might naturally suppose that if they exist then we should be able to find them, and so regard our inability as evidence that they do not exist” (Williamson (1994), p. 234).

<sup>4</sup> “What distinguishes vagueness as a source of inexactness is that the margin of error principles to which it gives rise advert to small difference in meaning, not to small differences in the objects under discussion.” (Williamson (1994), p. 231).

<sup>5</sup> See for example, Williamson (1997a), p. 927.

<sup>6</sup> Cf. Hawthorne (2006), p. 200.

<sup>7</sup> Williamson (1994), p. 204.

<sup>8</sup> It is not clear to us that Williamson is entirely correct on this point: after all we do have direct access to at least *some* facts about the cut-off points of predicates such ‘thin’ (we do know, for example, of a man of an average height that weighs 40kg that he is thin and of a man of an average height that weighs 200kg that he is not thin). We will return to this point in §6.

<sup>9</sup> In *Vagueness*, Williamson does not use the term ‘safety’, but he clearly appeals to the concept. He explicitly relates the notion of ‘margins for error’ which is discussed in *Vagueness* with the notion of safety in Williamson (2000), ch. 5.

<sup>10</sup> A subtle question here is whether saying that *q* is false in a close-by world should be equated with saying that *q* could have easily been false. Williamson, at least, seems to shift fairly freely between saying that a belief is safe if it is true in all similar cases (see e.g. Williamson (1994), p. 226), and saying that it is safe if it couldn’t have easily been false (see e.g. Williamson (1994), p. 230). But it is not clear that the two should be equated: in a billion ticket lottery, we could not easily have drawn ticket no. 345, but presumably the world in which we draw ticket no. 345 is just as close as any world in which we draw any (non-actual) other ticket. (Thanks to John Hawthorne for bringing this point to our attention). It thus seems that there is at least a sense according to which ‘*q* could not have easily been false’ means roughly that there aren’t *many* close worlds in which *q* is false, rather than that there aren’t *any* close worlds in which *q* is false. For the rest of the paper we will follow Williamson in assuming that if there is *one* close-by world in which *p*, then *p* could have easily been the case, though as far as we can see our arguments will work just as well if one wished to talk of *many* close-by worlds instead.

<sup>11</sup> For simplicity of exposition we will use the simplifying assumption that the sharp-boundary for ‘tall’ can be reduced to a cut-off height. We will also assume that ‘tall’ is restricted by context to the relevant comparison class – in all cases discussed in the paper this will be ‘tall for a person’.

---

<sup>12</sup> Williamson (1994), p. 230. As Williamson notes in n. 12, words here are not taken to be individuated by their meanings.

<sup>13</sup> Although sorites considerations show that for *any* expression *some* slight differences in use entail difference in meaning, Williamson thinks this phenomenon is particularly prevalent in the case of vague expressions because for them there are no natural properties or objects in the vicinity to stabilize the meanings. (see Williamson (1994), p. 231). We discuss this issue further in §3 of this paper.

<sup>14</sup> One might, of course, opt for a view of propositional attitude ascriptions according to which ascriptions of this kind are literally true, though inadequate for *pragmatic reasons* (as is defended for example by Salmon (1986) and Soames (1989)). In order to avoid unnecessary complications, we will talk henceforth as though such ascriptions are false. If one holds that they are true, though, one can translate all claims of the form ‘X believes that p’ in the paper in into claims of the form: ‘X believes that p, in a manner that makes it pragmatically adequate to describe her belief by saying ‘X believes that p’.

<sup>15</sup> But the discussion in n. 14 is relevant here too.

<sup>16</sup> We ignore here the complications arising with cases where X has a true belief of the form ‘That is green’ in the actual world, but lacks any belief in a close by world, because of, e.g., hallucination. No cases of this sort will be discussed in this paper.

<sup>17</sup> See e.g. Williamson (1997b), p. 263.

<sup>18</sup> See §6 below for the kind of method that might be relevant here.

<sup>19</sup> Note that we assume here that the method used to form the belief can play a role in determining the closeness of worlds. Alternatively, one can appeal to a notion of closeness that does not take methods into account, but restrict the safety condition so as to only require true beliefs in close-by worlds in which one uses the same method. It is also worth pointing out that we are not assuming here that reliable methods must be absolutely reliable, i.e. never yield false beliefs as applied in any possible circumstances. All we require is that there are at least some cases of the kind described where one’s application of the reliable method in the particular case in question will yield the correct results in all close-by worlds. The possibility of forming beliefs in this manner should be accepted by any proponent of a non-sceptical safety-theoretic account of knowledge: if many of our true beliefs are safe, then it must be possible to use belief-forming methods that yield correct results in all close-by worlds.

<sup>20</sup> A similar problem (though phrased in different terms) is raised in Gomez-Torrente (2002), pp. 112-113.

<sup>21</sup> We are assuming here, again, that the meaning of a word at time *t* is determined by its use at all times up to *t*, but not by future use facts.

<sup>22</sup> In particular note that since Williamson wants our knowledge to be exhibited in safe *assertions* further problems arise if vague terms do not have relatively stable meanings. For example, if our assertion of ‘‘tall’ means tall’ is to be true, then we should allow for enough stability of meaning so that ‘tall’ does not shift meaning between its first occurrence and its second occurrence in assertions of this sentence.

<sup>23</sup> It should be noted that in various places Williamson does briefly address this problem. In Williamson (1997c), p. 952, he suggests that an utterance doesn’t exactly have one determinate meaning at a time: what

---

one's utterance "means" or "says", depends on what counts as one speech community, and what counts as one's speech community may depend on context in which the words 'means' or 'says' are applied. So, for example, two agents can simultaneously say 'Jack said that Jill is tall', one speaking truly and one falsely, because their respective contexts of utterance mean that the word 'said' picks out a different relation as uttered by each. In a similar vein, Williamson (1999), p. 513 suggests that in the context propositional attitude reports the embedded utterance might receive a meaning that is parasitic on the meaning of the speaker reported, rather than its ordinary meaning. These proposals seem to us much too radical to accept as a response to this problem and have been convincingly argued against in Hawthorne (2006) (see especially §12-13).

<sup>24</sup> Thanks to John Hawthorne for this suggestion.

<sup>25</sup> See for example Gendler and Hawthorne (2005).

<sup>26</sup> Interestingly, Williamson seems to share our intuition about this when he says "our ability to recognize our friends and relations is not undermined by the mere possibility of look-alikes, although it might be undermined by their actual presence in the neighbourhood" (Williamson (1994), p. 236).

<sup>27</sup> One worry that some might raise with respect to this particular example is the following. Perhaps what counts as close-by worlds for the purposes of safety is a time-sensitive matter, so that one's judgment about this case depends on when the coin was tossed. If the coin was tossed fairly recently, the worry goes, then the belief indeed violates MBS but it is also fairly plausible that X lacks knowledge (because, so to speak, X was under a realistic risk of seeing a fake barn). On the other hand, if the coin was tossed a long while ago, then it is plausible that X does have knowledge, but then it is perhaps also plausible that X's belief does not violate MBS (because now the world where the coin lands differently and a fake barn is placed no longer counts as close). Either way, the objection goes, the example does not show that MBS is not a necessary condition on knowledge.

Leaving aside the question of the general merits of this time-sensitive notion of closeness, we think this train of thought is not one to which Williamson can appeal. For suppose that we only count worlds that diverge from the actual worlds recently as ones that might cause trouble for safety. Williamson would then need to argue that any world which is responsible for the violation of MBS is one which until recently had a shared history with the actual world (including the history concerning the meaning of the expression in question), but in which some small and recent difference in use facts entail a recent shift in meaning for that expression. But this puts pressure again on the idea of stability of meaning over time: the world in question would be a close world in which the meaning of the word is highly unstable over time.

<sup>28</sup> For the reader unfamiliar with the term 'building society' we note that the term refers to a kind of British financial institution which functions very much like a bank (building societies are similar to American credit unions). In fact, our supposed description of how the word 'bank' functions might well be an accurate description of British English.

---

<sup>29</sup> Of course, as we have set things out there could be another close-by world  $w^*$  where the word ‘eau’ does shift its meaning to XYZ in an analogous way to ‘water’. We will assume that there are natural facts that make some re-distributions of the H<sub>2</sub>O and XYZ molecules not easily achievable, ruling out this possibility.

<sup>30</sup> One could, of course, try to argue that the French scientists’ belief violates MBS, as it is described in English: In English we can adequately describe the French scientists (actual) belief by saying that they believe that water is H<sub>2</sub>O and one might try to argue that even in  $w$  one can adequately describe their belief in  $w$  using ‘water is H<sub>2</sub>O’, because that ascription in  $w$  would be ascribing to them the belief that XYZ is H<sub>2</sub>O and they have a disposition to misclassify XYZ samples as H<sub>2</sub>O. But even if the French scientists’ belief does violate MBS, rather than proving that they do not know this would merely constitute yet another counter-example to MBS being a necessary condition on knowledge.

<sup>31</sup> It is worth noting that Williamson never offers this motivation for MBS. We suspect that the reason for this is that Williamson wants to use MBS to characterize the unique kind of ignorance associated with vagueness, and since our ignorance regarding meta-linguistic cut-off points also seems to be part of the phenomenon of vagueness, he wants to appeal directly to MBS in the meta-linguistic cases as well. Hawthorne (2006) argues against the suggestion that ignorance of meta-linguistic cut-off points can be explained by direct appeal to the standard semantic-plasticity explanation, and instead offers an explanation of the meta-linguistic ignorance by appealing to something like our explanation involving Simple Safety above.

<sup>32</sup> For maximum clarity we now explicitly add a methods constraint, rather than assuming that it is already addressed by the relevant notion of closeness.

<sup>33</sup> By analogy you might consider the following case. Suppose Jill in the twin case above (§4, case 6) knows that the guy in front of her is Joe, and she also knows that Joe is called ‘Joe’. She therefore infers the guy in front of her is called ‘Joe’. But in the close-by possible world  $w$  in which ‘Joe’ refers to the other twin, she still assents to the sentence ‘The guy in front of me is Joe’ (which is false in this case), and ‘Joe is called ‘Joe’’ (which is still true in this case), and infers that the guy in front of her is called ‘Joe’. Her conclusion in  $w$  is false, and thus her actual belief violates Simple Safety, and, depending on how we interpret it, possibly also Methods Safety. Yet this does not seem to entail that in the actual case Jill does not know that the guy in front of her is called ‘Joe’.

<sup>34</sup> See for example Williamson (1994), §7.3.

<sup>35</sup> We are not ruling out the possibility, though, that some views on vagueness which do not accept sharp boundaries can avoid this problem. (Views which use a vague or indeterminate meta-language might well have more success here).

<sup>36</sup> Something like this position is advocated in Sorenson (1988).

<sup>37</sup> Breckenridge and Magidor (MS).

<sup>38</sup> Thanks to Wylie Breckenridge, John Hawthorne, Mike Martin, Richard Price, and Scott Sturgeon, as well as to audiences in the ‘Logic and Metaphysics Forum’, London, and in the ‘Vagueness and language use’ conference, Paris for helpful discussions.

---