

# TRUTH, MARKS OF TRUTH, AND CONDITIONALS\*

Ian Rumfitt

In an essay published in 2002, David Wiggins summarized his key contentions about truth as follows: ‘Truth is indefinable and irreducible, is not grasped by anyone through a definition, but has certain marks by identifying which one may elucidate the notion, without altogether demystifying it. Not least among these are a mark that makes truth an inherently normative notion and another by which any true belief or statement to the effect that *p* will under the right circumstances command a certain convergence among persons well enough placed and well enough qualified to judge, the convergence in question being one that is *itself* to be explained by reference to the truth that *p*’ (Wiggins 2002, 316).

The present paper assesses these claims. While Wiggins’s account gets many things right, I believe it needs refinements and qualifications. Even where it does not need those, an exposition which emphasizes different points may be valuable. There is certainly value in attending to Wiggins’s 2002 paper. Although conceived and presented as a definitive statement of his view of truth, that essay has not so far attracted the discussion it deserves.

---

\* I am delighted to contribute to this special issue about the work of David Wiggins, and wish only that the present essay was a better return for nearly forty years of instruction and twenty of friendship.

## 1. On whether truth is definable

Wiggins claims that truth is undefinable, and I agree that nothing has yet been found which could properly be called a definition of the notion. But the claim's precise sense needs to be spelled out if truth is not to be more mystifying than it needs to be.

Wiggins's claim is quite different from the conclusion of Tarski's purported proof that truth is undefinable. Tarski's conclusion was that, within a consistent theory rich enough to prove the Diagonal Lemma, there can be no predicate which applies to all and only the true sentences in the language of the theory (Tarski 1935, 249-51). That is, Tarski identified a limitation on delineating the *extension* of 'true', as it applies to the sentences of certain formalized languages. Wiggins's concern, by contrast, is with the predicate's *sense*. It is the sense of a word which someone grasps or fails to grasp, and Wiggins's claim is that no one can grasp the sense of 'true' by way of a definition. Someone who does grasp that sense will be able to apply the word to sentences in various languages, both natural and formal. Elucidating that sense, then, is a very different task from that of delineating the extension of the true sentences of arithmetic, or of some other theory which meets the conditions of Tarski's undefinability theorem.<sup>1/</sup>

Why is Wiggins sure that the sense of 'true' admits of no 'non-circular analysis...or final dismantlement into supposedly ultimate constituent notions' (2002, 317)? His main argument is simply that we should not expect to find definitions or analyses of notions which have long occupied central places in human thought:

---

<sup>1</sup> Wiggins is well aware that he and Tarski are frying different fish: 'What does Tarski mean...by "definition"? He intends by it neither conceptual analysis as we attempt this in philosophy nor lexicographical treatment of the sense..., but the systematic determination of the extension of a predicate. His immediate purpose is rather metamathematical than philosophical' (Wiggins 2002, 320).

If a whole cluster of fundamental terms came into being more or less simultaneously, if onwards from the dawn of thought the uses of these terms evolved simultaneously in some sort of mutual reciprocity, why should we expect it to be possible to unscramble the meaning of any one of them from that of all the others? (Wiggins 2002, 317)

He also hints at an inductive argument from the failure of philosophers to find satisfactory definitions of other basic notions:

In philosophy as it is, how many cases are there of successful, substantive definitions that invite charges neither of circularity nor yet of wrongness? (2002, 316)<sup>2</sup>/

While the question is rhetorical, one might cite Frege's definition of cardinal number (*Anzahl*) as a case which is neither obviously circular nor obviously wrong. In any event, these arguments are plainly inconclusive. Are there better ones?

Frege gave an argument for the indefinability of truth, the best statement of which is found in the unpublished *Logik* of 1897:

Now it would be futile to employ a definition in order to make it clearer what is to be understood by 'true'. If, for example, we wished to say 'a representation is true if it agrees with reality' nothing would have been

---

<sup>2</sup> I once heard George Boolos joke that the answer to this question is precisely one: Socrates's definition of *mud* as earth mixed with water at *Theaetetus* 147c. But that was a joke: Boolos fully appreciated the merits of Frege's definition of number.

achieved, since in order to apply this definition we should have to decide whether the representation in a given case did agree with reality, in other words whether it was true that the representation agreed with reality. Thus we should have to presuppose the very thing that is being defined. The same would hold of any definition of the form ‘A is true if and only if it has such-and-such properties or stands in such-and-such a relation to such-and-such a thing’ (Frege 1969, 139-40 = Frege 1979, 128-9).

Why, though, does any application of a definition of truth presuppose the notion being defined? The key to Frege’s argument is his conception of *judgement*. Suppose the candidate definition is ‘A is true if and only if A is *F*’. In order to apply this definition to conclude that A is true, we would first have to judge that A is *F*. But what is it to make a judgement? The explanation Frege gave in the ‘Logik’ of 1897 is the same as that found in all his writings: to judge is ‘inwardly to recognize a thought as true’ (Frege 1969, 150 = Frege 1979, 139). This is why it is ‘futile to employ a definition in order to make it clearer what is to be understood by “true”’. Before you can apply a definition, you must be capable of making a judgement; but to judge is to recognize a thought as true; so anyone capable of making a judgement must already have grasped the notion of truth.<sup>3</sup>

This argument, too, is inconclusive, though. It rests on the premiss that to judge is to recognize the truth of a thought, but it is not obvious that we have to

---

<sup>3</sup> Unlike Wiggins (see *n.1*), Davidson took Frege’s argument to be directed to the same conclusion as Tarski’s purported proof: ‘Truth is, as G.E. Moore, Bertrand Russell, and Gottlob Frege maintained, and Alfred Tarski proved, an undefinable concept’ (Davidson 1996, 265). ‘Concept’ is ambiguous, but it is clear from the passage quoted that Frege is denying that there is any non-circular specification of the *sense* of ‘true’, whereas Tarski thought he had proved that no predicate *in* a language *L* delineates the *extension* of ‘true’ as restricted to the sentences of *L*.

invoke truth when elucidating judgement. One might instead treat judging as a mental activity constituted by certain rules or norms—perhaps ‘Judge that  $P$  only when you have strong evidence that  $P$ ’, or ‘Judge that  $P$  only when you know that  $P$ ’. This treatment would stand alongside attempts to elucidate assertion by describing the constitutive rules or norms which distinguish it from other speech acts (see e.g. Williamson 1996). For Frege, indeed, an assertion is the outward expression of a judgement (see again Frege 1969, 150 = Frege 1979, 139).

So far as I can see, we have been given no reason to rule out, in advance of considering particular candidates, the possibility of finding a definition of truth. While the paucity of good analyses of basic notions makes success unlikely, we should take each candidate definition as it comes and assess it on its merits.

In a draft book *On Truth*, on which he was still working when he died, F.P. Ramsey proposed a definition of truth for beliefs which, once transposed to declarative sentences, comes to this:

$S$  is true if and only if  $\exists P(S \text{ expresses the thought that } P \wedge P)$ .<sup>4/</sup>

In the definiens, ‘ $\exists P$ ’ is a higher-order quantifier into sentence position, a quantifier whose logical properties A.N. Prior began to map out in his later writings.<sup>5/</sup>

---

<sup>4</sup> Cf. Ramsey 1991, 9. This account goes much deeper than the better known discussion in Ramsey’s earlier paper ‘Facts and Propositions’ (Ramsey 1927), which has been the wellspring of ‘redundancy’ or ‘deflationary’ theories of truth. Wiggins’s criticisms of those theories on pp.319-24 of his 2002 paper seem to me to be decisive.

<sup>5</sup> See esp. Prior 1971. On Ramsey’s account, ‘ $\exists P(S \text{ expresses the thought that } P)$ ’ is a straight consequence of ‘ $S$  is true’, but there are reasons for preferring a variant under which ‘ $S$  expresses a thought’ is a *presupposition* of ‘ $S$  is true’ (see Rumfitt 2019). When we emend Ramsey’s formula so as to make this a presupposition, we end up with a formal-dress version of P.F. Strawson’s elucidation of truth: ‘one who makes a

Does Ramsey's formula *define* truth? I think not. It exhibits the sort of circularity which Frege thought every candidate definition would manifest. The culprit in the present case is the sentential variable '*P*'. For which sentences are admissible substituents for '*P*'? Because any substituent must be capable of directly following the '^' sign, one constraint is this: any legitimate substituent must conform to the rules of the propositional calculus. That calculus, though, is a system of rules designed to ensure that whenever the premisses are true, the conclusion is also true. It follows that Ramsey's formula cannot be used as a strict definition of truth. In order to understand it, we need to know which sentences are legitimate substituents for '*P*'. But in order to know that, we must already possess at least an implicit grasp of the notion of truth. Against this candidate, then, Frege's objection bites: the putative definition 'presupposes the very thing that is being defined'.<sup>6</sup>

## 2. The value of Ramsey's formula as an elucidation of truth

I agree with Wiggins, then, that we have at present nothing which qualifies as an adequate definition of truth. I also agree that this should not deter a philosopher from trying to elucidate truth by describing its connections with other fundamental notions. Where we may differ, though, is that I regard Ramsey's formula as the best available starting point in that work of elucidation.

---

statement...makes a true statement if and only if things are as, in making that statement, he states them to be' (Strawson 1971, 180).

<sup>6</sup> The same objection would apply if Strawson's variant of Ramsey's formula (see the previous footnote) were put forward as a definition of truth. Not that Strawson did put it forward as such. He presented it simply as 'something uncontroversial and fairly general about truth' (Strawson 1971, 180).

To see one of its merits, let us compare it with a more familiar philosophical gloss on truth: a sentence is true when it corresponds to the facts. The latter gloss suggests that, in considering whether (for example) moral or aesthetic sentences may be assessed as true, the key question will be whether the world contains facts of an appropriate kind. Like Wiggins, I regard this approach as at best question-begging and at worst liable to leave us bogged down in a metaphysical morass (see Wiggins 1991, 140). Ramsey's formula points us well away from the mire. Given that formula, two questions will be key to deciding whether a sentence may be assessed as true. First, when someone utters it, does she express a thought, or must we give some other account of the rhetic act which she performs? Second, if she does express a thought, is the thought thereby expressed subject to the discipline of the propositional calculus? These questions may not be easy to answer, but they seem to be the right ones to address in determining whether a sentence may be assessed as true.

In order to appreciate the depth of the second question, it helps to consider an issue which perhaps raises fewer hackles than the alethic status of moral or aesthetic claims—viz., whether (so-called) 'indicative conditionals' may be assessed as true. A popular claim about indicatives—known variously as 'Adams's Thesis', 'the Ramsey Test', or simply 'the Equation'—says that a rational thinker's degree of belief in 'If  $A$ ,  $B$ ' will be her (conditional) degree of belief in  $B$ , given  $A$ . Symbolically,  $b(A \rightarrow B) = b(B|A)$ . An argument due to David Lewis, though, exposes a tension between the Equation and the thesis that indicative conditionals have truth values.<sup>7</sup> For let  $X$  and  $C$  be any propositions. By the propositional calculus,  $X$  is logically equivalent to

---

<sup>7</sup> Lewis 1976. Lewis himself held that indicative conditionals express propositions: he took their truth conditions to be those of the corresponding material conditional. However, other writers (notably Dorothy Edgington) have used his 'bombshell' arguments as part of their case for denying that indicative conditionals have truth values; see Edgington 1995, 271-8.

$(X \wedge C) \vee (X \wedge \neg C)$ , where  $X \wedge C$  and  $X \wedge \neg C$  are logically incompatible. Moreover, these logical facts are obvious. Assuming, then, that a rational thinker's degrees of belief respect obvious logical facts, we shall have that  $b(X) = b(X \wedge C) + b(X \wedge \neg C)$  for any propositions  $X$  and  $C$ . Now where  $b(C) \neq 0$ ,  $b(X|C) = b(X \wedge C)/b(C)$  so, if  $b(C) \neq 0$  and  $b(\neg C) \neq 0$ ,  $b(X) = b(X|C).b(C) + b(X|\neg C).b(\neg C)$ . Accordingly, if  $A \rightarrow B$  is a *bona fide* proposition, we shall have that  $b(A \rightarrow B) = b(A \rightarrow B|C).b(C) + b(A \rightarrow B|\neg C).b(\neg C)$ , for any  $C$  for which  $b(C) \neq 0$  and  $b(\neg C) \neq 0$ . In particular, then,  $b(A \rightarrow B) = b(A \rightarrow B|B).b(B) + b(A \rightarrow B|\neg B).b(\neg B)$ , whenever  $b(B) \neq 0$  and  $b(\neg B) \neq 0$ . Given, finally, that rational degrees of beliefs are closed under conditionalization, the Equation implies that  $b(A \rightarrow B|B) = b(B|A \wedge B)$  and  $b(A \rightarrow B|\neg B) = b(B|A \wedge \neg B)$ . Since  $b(B|A \wedge B) = 1$  and  $b(B|A \wedge \neg B) = 0$ , this yields the conclusion that  $b(A \rightarrow B) = 1.b(B) + 0.b(\neg B) = b(B)$ . That conclusion, though, is patently absurd. There are many conditionals whose consequents are neither certainly true nor certainly false, but where a rational thinker's degree of belief in the conditional is not the same as her degree of belief in the consequent.

I do not think that Lewis's argument refutes the hypothesis that indicative conditionals have truth values. While the Equation and the assumption of closure under conditionalization are good rules of thumb in most circumstances, our uses of vernacular conditionals do not strictly conform to either principle, so the 'bombshell' fails to explode (see Rumfitt 2013). Lewis's argument, though, does illustrate the power of the requirement that a true sentence must express a thought which is subject to the discipline of the propositional calculus, for what drives it is precisely the logical equivalence of  $X$  with  $(X \wedge C) \vee (X \wedge \neg C)$ . That in turn shows how Ramsey's formula yields a real constraint on truth evaluability. We can imagine speakers who use indicative conditionals in strict conformity both with the Equation and with



closure under conditionalization. Lewis's argument would show that their conditionals were not to be assessed as true or false.

Ramsey's formula, then, yields non-trivial requirements on truth. How does it compare with Wiggins's own 'marks of truth'? Wiggins's approach to the topic is far less direct. He does not consider a candidate elucidation in the form '*S* is true if and only if  $\Phi(S)$ '. Rather, he seeks to locate marks of truth by asking 'what must the concept *true* be like if its restrictions, true-in- $L_1$ , true-in- $L_2$  (etc), are to play the role that they play in the account of declarative meaning that has issued in (iii)?' (2002, 329). Turning back to (iii), we find this:

(iii) where *s* is declarative, *s* signifies or conveys literally in *L* that *p* if and only if

(a) there is a truth-definition  $\Theta$  for *L* which implies that *s* is true just in case *p* and

(b) the output of  $\Theta$ , when applied to the utterances of speakers of *L*, advances substantially and unimprovably the effort to make total sense of the whole conduct, linguistic and non-linguistic, of speakers of *L* (Wiggins 2002, 327-8).

In other words, truth is constrained by the requirement that a truth theory for *L*, constructed along Tarskian lines, should be the kernel of a theory of meaning or interpretation for the speakers of *L*.

I am sympathetic to the underlying idea that the kernel of a theory of meaning for a language will be a systematic assignment of truth conditions to its declarative sentences. However, I cannot follow Wiggins in his Davidson-inspired way of

elaborating this idea. The breach comes over Tarski's 'Convention **T**' (Tarski 1935, 187f). Like Davidson, Wiggins envisages truth conditions being assigned by Tarskian *T*-sentences—i.e. material biconditionals in the form '*S* is true if and only if *p*'. If a truth theory did not imply a *T*-sentence for every sentence of the relevant language, he writes, it 'would certainly fall short of reflecting the intuitive meaning of "true" in the object language' (Wiggins 2002, 321). For anyone guided by Ramsey's formula, however, this is misconceived: the Liar Sentence—crucial to Tarski's purported indefinability proof—provides a flat-out counterexample to Convention **T**. The Liar Sentence,  $\lambda$ , viz. 'Sentence  $\lambda$  is not true' is plainly meaningful, so Convention **T** would appear to imply that the following *T*-sentence is true:

( $T\lambda$ ) Sentence  $\lambda$  is true if and only if sentence  $\lambda$  is not true.

Clearly, however, ( $T\lambda$ ) is not true. Tarski saves Convention **T** by denying that  $\lambda$  belongs to the language whose true sentences are being delineated. Hence his extraordinarily restrictive conclusion: a truth predicate *for* a language *L* cannot be a predicate *in* *L*. For a Ramseyan, though, this is quite the wrong moral to draw. The Liar is a perfectly meaningful sentence of English. Given Ramsey's formula, however, the reasoning of the Liar Paradox constitutes a proof that there is no thought which an utterance or inscription of it expresses. The reasoning amounts, in other words, to a proof that  $\neg\exists P(\lambda \text{ expresses the thought that } P)$  (see e.g. Williamson 1998, 12-15). We must acknowledge, then, that some utterances of meaningful English sentences fail to express thoughts and hence are 'undefined' in truth value. Now ( $T\lambda$ ) is also a meaningful sentence of English. Its right-hand side is undefined and its left-hand side is either false or undefined, so on no sensible treatment of the biconditional

will it come out true. Accordingly, the Tarskian requirement that a satisfactory theory of truth for  $L$  should imply a  $T$ -theorem for each meaningful declarative sentence of  $L$  must be rejected.

An elucidation of truth which stems from Ramsey's formula, then, will differ from Wiggins's. Because English and other natural languages contain truth predicates which apply (amongst other things) to their own sentences, I cannot accept an account which builds in Convention **T** and with it Tarski's rejection of semantically closed languages. For this reason, I prefer Ramsey's elucidation of truth to Wiggins's. As the case of indicative conditionals shows, Ramsey's account is far from nugatory. It imposes constraints on the application of 'true' which certain sorts of sentences may fail to meet.

### **3. Why convergence is not a mark of truth**

Wiggins's specific marks of truth are supposed to derive from the fundamental requirement that a Tarskian truth theory will be the kernel of a systematic theory of interpretation. In particular, this applies to his mark of *convergence*. Explaining that mark, he writes:

If  $x$  is true,  $x$  will under favourable circumstances command convergence [i.e. different speakers will converge on an agreement that  $x$  is indeed true], and the best explanation of the existence of this convergence will either require the actual truth of  $x$  or be inconsistent with the denial of  $x$  (Wiggins 1991, 147).

I cannot myself see how the derivation of this mark from the fundamental requirement is supposed to go. Davidson's idea that interpretation presupposes widespread agreement seems to be in the mix, but the details elude me. In any case, the purported derivation will really only interest those who accept that truth is the property whose extensions (as restricted to particular languages) are delineated by Tarskian theories. For all that, I shall venture some comments about this putative mark. Convergence, I contend, is reasonably expected only for some truths. While there is a universal mark of truth in the vicinity, it is not the one Wiggins identifies.

Wiggins holds that competent judges converge on some, but not all, moral claims. They will agree about sentences in which a 'thick' moral term is directly predicated of an action or state-of-affairs. Thus, discussing Gilbert Harman's case of hoodlums torturing a cat, Wiggins contends that (1) people will converge on the judgement 'What the hoodlums did was callous and cruel'; and (2) the best explanation of this convergence is that what they did was indeed callous and cruel (Wiggins 1991, 156 *f*). He does not contend that 'there must be the prospect of a convergence among *all rational intelligences* whatever, or convergence among *all rational creatures of human fabric and constitution*...The expectation can be legitimately confined to people of a certain culture who have what it takes to understand a certain sort of judgment' (1991, 160; emphasis in the original).

I accept that true sentences of this simple form—'pure, relatively specific evaluations', as Wiggins calls them (*ibid.*)—will command convergence. The terms 'callous' and 'cruel' have meanings, and it is hard to see how could they could have acquired meanings unless there were widespread agreement that they apply to certain cases. Cat-torturing is one of these paradigms. This convergence may be specific to our culture and it may involve the moral sentiments which acculturation instils. Like

Wiggins, though, I do not see why that cultural specificity should disqualify pure evaluations from being true. ‘Pillar boxes are red’ is true, even though we only expect creatures with our visual systems to converge upon it.

Wiggins, though, also applies his mark of convergence contrapositively, to attack the position he calls ‘strong cognitivism’, which embraces ‘in its scope of applications not only pure valuations but everything else that falls within the provinces of the prudential, aesthetic, and moral’ (Wiggins 1991, 161). Consider the sentence ‘Captain Vere did the right thing when he signed Billy Budd’s death warrant’. Wiggins’s discussion of Budd’s story is subtle and intricate, and it introduces issues which are orthogonal to truth. However, the upshot of his ‘underdeterminationist position’ (*op. cit.*, 178) seems to be that this sentence about Vere is neither true nor false. It is not true, because competent authorities even from our own culture do not converge upon its acceptance. It is not false, because those same authorities also do not converge upon acceptance of its negation.

I agree with Wiggins that competent judges will not converge in this case. Indeed, in the operatic version of the story, Vere himself fails to reach a settled view: the Prologue and Epilogue portray him as having been tormented for decades by the question of whether he did the right thing in having Budd executed. Moreover, I do not dispute Wiggins’s evaluation of the sentence. Since I do not accept the unrestricted Principle of Bivalence, I allow that there are sentences which are of a sort to be assessed as true or false, but which in fact are neither. The sentence about Vere may well be one of these. What I want to challenge is Wiggins’s *argument* for identifying here a truth-value gap, or some similar species of underdetermination. The principle of that argument (‘no convergence, *ergo* not true’) is the contrapositive of his mark of convergence, but it seems to me to be unsound. For there are all sorts

of things which might account for a lack of convergence even when the relevant sentence is true.

To illustrate this, let me turn from morals to musical aesthetics. The relevant judges here will be musical people, but who are they? What marks them out, I would suggest, is that they get the ‘pure, relatively specific [musical] evaluations’ right. When a musical phrase is elegant, they agree that it is elegant. (Even better, of course, if they produce some elegant phrases of their own.) When a composition is banal, they agree that it is banal. There is no problem so far for Wiggins. Such judges will converge on accepting ‘The overture to Wagner’s *Rienzi* is banal’, and that sentence is plainly true.

It is when we turn to what might be called *baggier* musical judgements, however, that a problem for Wiggins comes into view. One might suppose that musical people would also converge on ‘Mozart was a fine composer’, but in fact this is not so. Glenn Gould was a notorious counterexample. He was unquestionably musical, but he had a low opinion of Mozart’s mature output. Some expressions of this view were, no doubt, designed to shock or to wind people up: ‘Mozart died too late rather than too soon’; ‘The [second] G minor symphony consists of eight remarkable measures...surrounded by a half-hour of banality’ (Gould 1976, 32, 35). However, Gould’s lectures and articles on Mozart are too closely argued to be dismissed as having been written merely *pour épater les bien-pensants*. One striking thing about them is that some of the features he deems to be compositional flaws are cited approvingly by those who appreciate Mozart. The use of arpeggios and scales to fill out successive four-bar phrases—which some of us hear as contributing to the music’s accessibility and its sense of proportion—Gould heard as being so trite as to sound almost prefabricated, as though Mozart’s compositions were no more than

supremely competent exercises in the musical equivalent of painting by numbers. That is far from being my own response, but one might hesitate to dismiss Gould's judgement as simply *mistaken* when it was both longstanding<sup>8</sup>/ and based on a close study of the scores. Few, certainly, would have had the nerve to challenge him with the words Wiggins sometimes uses to gloss his mark of convergence, and tell him to his face that 'there is nothing else to think than that Mozart was a fine composer' (cf. e.g. Wiggins 1991, 164). For Gould clearly had thought hard about the matter, but come to a very different view.

That said, I think we can explain why Gould thought as he did about Mozart. The key to the explanation is a throwaway remark in an interview. Trying to justify his having ignored most of the *sforzando* markings in his recordings of Mozart's piano sonatas, Gould commented that 'they represent an element of theatricality to which my puritan soul strenuously objects' (1976, 36). Now Mozart's mature compositions—even his purely instrumental compositions—are full of theatrical tropes. More particularly, they are imbued with the gestures of *opera buffa*. This is especially true of the twelve piano concertos (K.449 to K.503) which Mozart wrote in the years (1784 to 1786) when he was first developing the musical language he would need for a large-scale comic opera, and then actually composing *Le Nozze di Figaro*. Now if you are the sort of puritan who cannot abide *opera buffa*, you will react against musical works which, although purely instrumental, contain tropes from comic opera, and Gould duly deemed Mozart's mature concertos to be 'unfixable' (*op. cit.*, 32) and refused to play them. Furthermore, we can explain in related terms Gould's aversion to other Mozartian pieces. He derided the sonata-allegro form—

---

<sup>8</sup> '[As a student] I simply couldn't understand how my teachers, and other presumably sane adults of my acquaintance, could count these pieces among the great musical treasures of Western man' (Gould 1976, 33).

ubiquitous in the works of Mozart and his contemporaries—as an uninteresting simplification of the more complex structures of the Renaissance and Baroque masters. As many of the numbers in *Figaro* show, though, sonata-allegro form *is* well suited to give musical expression to the sort of comic imbroglio in which an incident or situation creates a tension which is at least partly resolved. One begins to understand why the very language of the high classical style—a style which reached its apogee in *Figaro*—would trouble a puritan soul.

How does this bear on Wiggins? I think it undermines his thesis that convergence is a general mark of truth. The truth or falsity of ‘Mozart was a fine composer’ depends on the quality of his scores, and those scores are surely such as to make the sentence true. What Gould’s case shows, though, is that irrelevant factors can intrude and prevent even a highly musical person from apprehending that truth. If a given group of musical people contains enough puritans, there will be no hope of its converging on that truth. If, indeed, it contains only puritans, it may even converge on its negation. But these vicissitudes of convergence are plainly irrelevant to the *truth* of the sentence in question, which depends only on the quality of Mozart’s compositions.

The underlying problem here is general. There are any number of bizarre and irrelevant factors which may come in ‘from left field’, as Americans say, and derail any prospect of convergence. It is hard to see, though, why the intrusion of such factors should be enough to refute the hypothesis that a given sentence is true, even when it ‘falls within the provinces of the prudential, aesthetic, or moral’. Wiggins’s mark of convergence, however, implies that this is sufficient to refute any claim to truth. I conclude that that mark ought to be rejected.



#### 4. A Fregean mark of truth

One of the presuppositions of convergence, by contrast, is a far more plausible candidate to be a universal mark of truth.

To see what this presupposition is, it helps to return to Frege's unfinished *Logik* of 1897. For Frege, the primary bearers of truth are what he calls *thoughts* (*Gedanken*), about which he says:

A thought does not belong specially to the person who thinks it, as a mental representation [*Vorstellung*] does to the person who has it, but stands to everyone who grasps it in the same way and as the same thought. Otherwise two people would never attach the same thought to the same sentence, but each would have his own thought; and if, for example, one man put  $2 \cdot 2 = 4$  forward as true whilst another denied it, there would be no contradiction, because what was asserted by one would be different from what was rejected by the other. It would be quite impossible for the assertions of different people to contradict one another, for a contradiction occurs only when it is exactly the same thought that one person is asserting to be true and another to be false. A dispute about the truth of something would thus be futile. There would simply be no common ground to fight on; each thought would be enclosed in its own private world [*Innenwelt*] and a contradiction between the thoughts of different people would be like a war between ourselves and the inhabitants of Mars (Frege 1969, 145 = Frege 1979, 133).

In a few cases, Frege thinks, there *is* only the appearance of contradiction, and disputes about the truth of something are merely apparent. In a paper published more than twenty years later, he gave the example of two men arguing about the properties of a rainbow. ‘People who had used in conversation the expression “This rainbow” could be brought to see ‘that they had not been designating anything by these words, since what each of them had had was a phenomenon of which he himself was the owner’ (Frege 1919, 146). I am not sure if this is the best account of rainbows. One might do better to revert to the famous analogy with which Frege glossed his theory of sense and reference, and liken a rainbow to ‘the real image projected by the object glass in the interior of [a] telescope’—which ‘is still objective, inasmuch as it can be used by several observers’ (Frege 1892, 30). This questionable example, however, does not imperil Frege’s general point, which is that arguments about a sentence’s truth presuppose that the parties are attaching the same sense to it. That holds good whether the argument ends in agreement or disagreement. Accordingly, attaching a common sense to a sentence is a presupposition of convergence.

We can distil from these observations a genuine mark of truth: if a sentence is to be true—if, indeed, it is to be a candidate for truth—it must be possible for different speakers of the relevant language to attach a common sense to it. Moreover, it must be possible for such speakers to know that they are attaching a common sense to it. The possibility in question must be a real or practical one: it is not enough that it should obtain in some remote possible world.

There is, however, a popular position in current philosophy of language which implicitly rejects this Fregean mark of truth. The position concerns indicative conditionals and posits that, while these have truth values, the propositions they

express depend on speakers' epistemic state in such a way as to make it practically impossible for a hearer to know exactly what is said by uttering one.

The position in question is an attempt to solve a problem generated by the 'Equation' of §2: a rational thinker's degree of belief in 'If  $A$ ,  $B$ ' is the same as her (conditional) degree of belief in  $B$ , given  $A$ . Weak assumptions about conditional belief entail that no rational thinker can simultaneously have a high degree of belief in  $B$  given  $A$ , and a high degree of belief in not  $B$  given  $A$ . Assuming the Equation, it follows that no rational thinker can simultaneously have high degrees of belief in the pair of indicative conditionals 'If  $A$ ,  $B$ ' and 'If  $A$ , not  $B$ '.

The problem arises because there are circumstances where it does seem rational simultaneously to accept such a pair. It was Allan Gibbard (1981) who first drew attention to this with his tale of 'Sly Pete', but the asymmetries of that story have distracted some commentators so I shall focus on a more recent example due to Timothy Williamson (2020, 89 *f*). We are at a nuclear power plant where a detector beside the nuclear core is connected to various warning lights. When the detector is working and the core is overheating, each light is red. When the detector is working and the core is not overheating, each light is green. When the detector is not working, each light is red or green at random, independently of the others. A competent and trustworthy engineer, East, sees only the easternmost light, which is red, so he sends the following message to the plant's controller:

- (1) If the detector is working, the core is overheating.

Another competent and trustworthy engineer, West, sees only the westernmost light, which is green, so he sends this message to the controller:

- (2) If the detector is working, the core is not overheating.

Since the controller trusts both the engineers, it seems she can rationally accept both messages. Indeed, having accepted them, it seems that she can rationally infer

- (3) The detector is not working.

As we have seen, though, the Equation implies that the controller cannot rationally accept both (1) and (2).

Now one way out of the problem would be to accept that in the case described we have a counterexample to the Equation. Williamson thinks that in the end we shall have to accept this, and I agree. Some of the Equation's determined defenders, though, hold that, because the two speakers know different things, the thought or proposition which East expresses by (1) is not incompatible with the thought West expresses by (2). This solution is bought at the cost of violating the Fregean mark of truth: since a hearer will usually know little about a speaker's epistemic state, she will also not know which proposition a speaker who affirms a conditional is putting forward as true. The proposition thereby affirmed will precisely not 'stand to everyone' who hears the conditional and understands the relevant language 'as the same thought'.

Which way ought we to go? Should we reject the Equation, or defend it by rejecting the Fregean mark of truth? As Williamson (*op. cit.*) pertinaciously argues, there is really no contest here: it is an illusion that postulating this extreme form of context-sensitivity solves the problem. For we still have to explain how the controller

can rationally infer (3) from what the engineers tell her. The only available manoeuvre here is to claim that the material conditionals

(1') If the detector is working  $\supset$  the core is overheating

and

(2') If the detector is working  $\supset$  the core is not overheating

express the publicly available residue of the engineers' distinct private thoughts; in accepting (1') and (2'), the controller is accepting premisses which logically entail (3). As Williamson shows, though, this explanation breaks down over only slightly more elaborate cases (2020, 93*f*).

Even if we solve the immediate difficulty by rejecting the Equation, though, residual problems remain. If conditionals have truth values, we have to say under what conditions they are true; the truth conditions we assign must validate the inference from (1) and (2) to (3). If *per contra* we deny that indicative conditionals have truth values, we still need to spell out what is communicated by saying 'If *A*, *B*', and the present case presents a difficulty for the obvious account:

In felicitous cases, I utter an indicative conditional, and thereby insure that the audience comes to accept that I have a certain conditional belief, belief in *B* given *A*. The audience does so because it trusts my sincerity and command of language. The audience then infers from my believing *B* given *A* that I have some good grounds for so believing, and takes that as a reason for believing *B*

given *A*. Thus is my conditional belief communicated to them (Gibbard 1981, 230).

This model of communication, however, breaks down in the present case. The controller will take East's utterance as a reason for herself believing *B* given *A*. *Pari passu*, she will take West's utterance as a reason for herself believing not *B* given *A*. The problem is that she cannot coherently have both conditional beliefs. If she finds herself trying to hold both, she has to go back and revisit the assumptions (that both East and West are competent observers, that both are trustworthy, that both can speak English...) which have led her to this pass. Gibbard's account of communication using indicative conditionals fails, then, in this case. It does not account for what needs to be explained, which is that the controller rationally accepts both (1) and (2) and then rationally infers (3).

## **5. Two ways forward**

The question of whether indicative conditionals have truth values is a large and complex issue, on which many different considerations bear. Rather than prejudge it, I shall conclude this paper by suggesting two ways forward: one for those who deny that conditionals have truth values, the other for those who accept that they do.

Even leaving aside its failure to deal with the present case, Gibbard's account of communication is already suspect. As he says (1981, 230 *n.15*), it extends to conditionals Grice's theory, whereby communication—or, at least, the sort of communication characteristic of declarative speech—is essentially a matter of a

speaker's trying to instil *beliefs* in an audience (Grice 1957). That theory, though, fails to cover all instances of successful communication even when conditionals are not under consideration. The central notion in declarative communication is 'Speaker *S* tells hearer *H* that *P*', but it is wrong to analyse that in terms of *S*'s attempting to get *H* to believe something. There are myriad cases where a speaker tells someone something without caring in the slightest whether the hearer believes him or not. There are also cases where a speaker tells someone something even though he knows full well that he will not be believed. The latter will be failed attempts to *persuade* or *convince* the audience, but they may be successful instances of *telling*. Any account of '*S* told *H* that *P*' needs to respect the conceptual distance between that and '*S* persuaded *H* that *P*' or '*S* convinced *H* that *P*'.

But if telling does not essentially involve instilling a belief, what is characteristic of it? The account I prefer has roots in Thomas Reid and C.S. Peirce but has been most fully developed in our day by Richard Moran (2018). For *S* to tell *H* that *P* is for *S* to utter something by which he offers *H* his *assurance* that it is the case that *P*. *S* can make this offer without caring whether it is taken up; he can make it, indeed, in circumstances where he knows it will be spurned. As Moran shows in detail, this analysis of telling helps to explain how being told things is a fertile source of knowledge.

Supposing that something along these lines is a satisfactory account of acts of telling using non-conditional declaratives, how might it be extended to cover conditionals? Alongside the notion of an outright assurance, we have the general idea of a conditional assurance. Suppose a friend wishes to stand for Parliament but is worried that his family will be impoverished if he loses his deposit. If I think it would be a good idea for him to stand, I might offer him a conditional assurance: I might

assure him that, in the event of his losing his deposit, I will reimburse him. Such an offer may affect the behaviour of an agent who accepts it even if the relevant condition never obtains. Suppose my friend does not end up losing his deposit. It may still be the case that he would not have stood for Parliament without my conditional assurance of reimbursement in the event that he did.

How might this general notion of a conditional assurance be applied to elucidate the act of using a conditional to tell someone something? When *S* uses ‘If *A*, *B*’ to tell *H* something, we might say, *S* offers *H* his conditional assurance that *B* is true, in the event that *A* is true. This account presupposes that the antecedent and consequent of a conditional are candidates for truth. It does not, however, require that the whole conditional is such a candidate.<sup>9</sup>

One merit of this analysis is that it accounts for the controller’s reaction to the two messages she receives. East has offered her his conditional assurance that the core is overheating, in the event that the detector is working. West has offered her *his* conditional assurance that the core is not overheating, in the event that the detector is working. The controller may rationally accept both offers. In the event that the detector is working, East’s assurance will imply that ‘The core is overheating’ is true, and West’s that ‘The core is not overheating’ is true. In that case, the controller’s trust in at least one of her engineers will turn out to be misplaced. However, the two conditional assurances will come into conflict *only if* the detector is working. That is why the controller infers, and is entitled to infer, that the detector is not working.

We have, then, a promising account of acts of telling using indicative conditionals which does not assume that conditionals have truth values. But what if

---

<sup>9</sup> What about cases where the consequent is itself a conditional? We might analyse ‘If *A*, then if *B*, *C*’ as ‘If *A* and *B*, *C*’.



they do? Since (1') and (2') jointly entail (3), a theory on which 'If  $A$ ,  $B$ ' has the truth conditions of the corresponding material conditional accounts for this case very straightforwardly, and Williamson's book is an extended defence of just such a theory. Interesting and ingenious as it is, though, the defence is an uphill battle. Is there another assignment of truth conditions to indicative conditionals which gives us what we want?

I think there is, and we find it by applying a method—that of paying close attention to the conditional construction's grammar—which David Wiggins would find congenial.<sup>10</sup> As Lewis remarked some years ago, we often find a conditional clause within the scope of an adverb or adverbial phrase, as in 'Usually/always/mostly/on Sundays, if Mary is here, she is angry' (see Lewis 1975). As he also pointed out, the role of the 'if'-clause in these sentences is to *restrict* a quantifier. It is a delicate issue what sort of thing these adverbs are quantifying over, but if we suppose (for the sake of definiteness) that they are quantifying over situations, then the sentence 'Usually, if Mary is here, she is angry' has the sense of 'Most situations in which Mary is here are situations in which she is angry'. Now 'most' is an irreducibly binary quantifier, which takes a plural term (the 'restrictor') and a verb phrase (the 'matrix') to form a sentence. The role of the 'if'-clause in the sentence, then, is to indicate the appropriate restrictor: the ' $A$ ' of 'If  $A$ ,  $B$ ' supplies 'situations in which  $A$ ', or ' $A$ -situations' for short. A similar analysis works when 'usually' is replaced by 'necessarily', 'probably', 'it is likely that', 'it must be'.

---

<sup>10</sup> Wiggins 1986 is a marvelous advertisement for the benefits which can accrue in philosophy from scrupulousness over grammatical points. When Wiggins and I co-taught the B.Phil. proseminar in the Michaelmas Term of 1998, he insisted that we should start each class by reading a couple of paragraphs from Michael Dummett's *Grammar and Style: For Examination Candidates and Others* (Dummett 1993). While some of the students found this eccentric, one hopes that by now they see the point.

What, though, about a ‘bare’ conditional ‘If  $A$ ,  $B$ ’, where the ‘if’-clause does not appear to lie within the scope of any other operator? Loosely following Angelika Kratzer (1981) we might assign to the bare ‘If  $A$ ,  $B$ ’ the truth conditions of ‘ $A$ -situations are  $B$ -situations’, where the situations in question encompass those in nearby possible worlds, as well as those which actually obtain. This semantic postulate explains a number of features of bare conditionals, of which I mention two.

The first is the fact that in many languages, we can reverse the roles of antecedent and consequent in a bare conditional by prefacing it with a word meaning ‘only’ (and in some cases adjusting the word order). If we start with ‘If John has worked hard, he will get a First’, and preface this with ‘only’, we get ‘Only if John has worked hard will he get a First’, which has the same truth conditions as ‘If John will get First, he will have worked hard’. The postulate explains this, for we find the same reversal with ‘ $A$ s are  $B$ s’. If we start with ‘Those who have worked hard will get Firsts’ and preface this with ‘only’, we get ‘Only those who have worked hard will get Firsts’, which has the same truth conditions as ‘Those who will get Firsts will have worked hard’. We find the same effect in French (*‘seulement si’*), in German (*‘nur wenn’*), and in many other languages. The postulate shows how the meaning of ‘only’ combines with the truth conditions of conditionals to produce it.

The second merit concerns the logic of bare conditionals. Given the postulate, the logic of ‘if...then’ will be an off-shoot of quantificational logic; this accounts for familiar logical principles for conditionals while explaining why those principles have apparent exceptions. ‘ $A$ -situations are  $B$ -situations’ combines with ‘The actual situation is an  $A$ -situation’ to entail ‘The actual situation is a  $B$ -situation’; this accounts for the validity of *Modus Ponens*: ‘If  $A$ ,  $B$ ;  $A$ ; ergo  $B$ ’. Similarly, ‘ $A$ -situations are  $B$ -situations’ combines with ‘ $B$ -situations are  $C$ -situations’ to entail ‘ $A$ -

situations are *C*-situations'; this accounts for the validity of Hypothetical Syllogism:

'If *A*, *B*; if *B*, *C*; ergo if *A*, *C*'.

This is all well and good, but in applying these principles to actual arguments, we need to take care that the relevant domain of possible situations does not shift.

Apparent counterexamples to the principles will arise when they do. Thus John Burgess gives the following putative counterexample to Hypothetical Syllogism (Burgess 2009, 80-81):

If Clinton wins the primary, Obama will come in second

If Obama dies before the primary, Clinton will win it

So: if Obama dies before the primary, he will come in second

On the present analysis, the anomaly is easily explained. The argument's first premiss is uttered in a context in which it is assumed that Obama will live to take part in the primary. That is, situations in which he dies before the primary are not counted as being 'nearby'. The second premiss, though, presents his early death as a serious possibility, so as the argument proceeds the 'nearby' situations expand to include some in which Obama dies before the primary. When the relevant domain of quantification shifts as the argument progresses, we cannot expect Hypothetical Syllogism to remain applicable. However, it remains a sound logical principle.

How, though, might this account of conditionals be applied to Williamson's problem case? In her own discussion of 'Sly Pete', Kratzer develops Gibbard's story so as to ensure that the conditional messages sent by Pete's henchmen are incompatible (Kratzer 2012, 102-4). That discussion, then, does not cover the sort of case which concerns us, in which two bare conditionals 'If *A*, *B*' and 'If *A*, not *B*' are

both true. Applying the semantic postulate directly to East's conditional, however, the truth conditions of

- (1) If the detector is working, the core is overheating,

will be given by

- (1'') Situations in which the detector is working are situations in which the core is overheating.

Similarly, the truth conditions of West's

- (2) If the detector is working, the core is not overheating,

will be given by

- (2'') Situations in which the detector is working are situations in which the core is not overheating.

(1'') and (2'') are fully compatible, and together they entail

- (3'') There are no situations in which the detector is working.

(Compare the sound argument: 'Unicorns are pink; unicorns are not pink; *ergo* there are no unicorns'.) Since the actual situation is by definition 'nearby', (3) implies

(3) The detector is not working,

exactly as required. Among theories which take conditionals to have truth values, then, we have found two which can account for the validity of the inference from (1) and (2) to (3). The Kratzer-inspired semantic postulate offers an alternative to taking the truth conditions of ‘If  $A$ ,  $B$ ’ to be those of ‘ $A \supset B$ ’.

## 6. Conclusion

Wiggins is right to seek to elucidate the notion of truth rather than define it. However, an elucidation which stems from Ramsey’s formula fares better than Wiggins’s own approach, based as it is on Tarskian truth theories. Moreover, his mark of convergence should be replaced by the Fregean requirement that different speakers must be able to attach a common sense to a sentence which is up for assessment as true. Once these changes are made, we can make progress on some of the hard cases.<sup>11/</sup>

---

<sup>11</sup> For their comments on a draft, I am much indebted to Stephen Everson, Richard Holton, Christopher Peacocke, and Timothy Williamson.

## REFERENCES

Burgess, J.P. 2009. *Philosophical Logic*. Princeton, NJ: Princeton University Press.

Davidson, D.H. 1996. 'The folly of trying to define truth'. *The Journal of Philosophy* **93**: 263-78.

Dummett, M.A.E. 1993. *Grammar and Style: For Examination Candidates and Others*. London: Duckworth.

Edgington, D.M.D. 1995. 'On conditionals'. *Mind* **104**: 235-329.

Frege, G. 1982. 'Über Sinn und Bedeutung'. *Zeitschrift für Philosophie und philosophische Kritik* **100**: 25-50.

———. 1919. 'Die Verneinung'. *Beiträge zur Philosophie des deutschen Idealismus* **I**: 143-57.

———. 1969. *Nachgelassene Schriften*, ed. H. Hermes et al. Hamburg: Felix Meiner.

———. 1979. *Posthumous Writings*, translated by P. Long and R. White. Oxford: Blackwell.

Gibbard, A.F. 1981. 'Two recent theories of conditionals'. In W. Harper et al., eds., *Ifs* (Dordrecht: Reidel), pp.211-47.

Gould, G.H. 1976. 'On Mozart and related matters: Glenn Gould in conversation with Bruno Montsaingon'. *The Piano Quarterly* **24**. Page references are to the reprinting in Gould, ed. T. Page, *The Glenn Gould Reader* (New York: Vintage Book, 1984), pp.32-43.

Grice, H.P. 1957. 'Meaning'. *The Philosophical Review* **66**: 377-88.

Kratzer, A. 1981. 'The notional category of modality'. In H.-J. Eikmeyer & H. Reiser, eds., *Words, Worlds, and Contexts* (Berlin: Walter de Gruyter), pp.38-74.

———. 2012. 'Conditionals'. In Kratzer, *Modals and Conditionals* (Oxford: Oxford University Press), pp.86-108. (An expanded version of a paper originally published in 1986.)

Lewis, D.K. 1975. 'Adverbs of quantification'. In E. Keenan, ed., *Semantics of Natural Language* (Cambridge: Cambridge University Press), pp.3-15.

———. 1976. 'Probabilities of conditionals and conditional probabilities'. *The Philosophical Review* **85**: 297-315.

Moran, R. 2018. *The Exchange of Words: Speech, Testimony, and Intersubjectivity*. Oxford: Oxford University Press.

Prior, A.N. 1971. *Objects of Thought*, ed. P.T. Geach & A.J.P. Kenny. Oxford: Clarendon Press.

- Ramsey, F.P. 1927. 'Facts and propositions'. *Proceedings of the Aristotelian Society, Supplementary Volume 7*: 153-70.
- . 1991. *On Truth*. Dordrecht: Kluwer.
- Rumfitt, I. 2013. 'Old Adams buried'. *Analytic Philosophy* **54**: 157-88.
- . 2019. 'Truth'. *Oxford Studies in the Philosophy of Language* **1**: 148-77.
- Stalnaker, R.C. 1984. *Inquiry*. Cambridge, Mass.: MIT Press.
- Strawson, P.F. 1971. 'Meaning and truth'. In his *Logico-Linguistic Papers*. (London: Methuen), pp.170-89.
- Tarski, A. 1935. 'Der Wahrheitsbegriff in den formalisierten Sprachen'. *Studia Philosophica* **I**: 261-405. Page references are to the translation by J.H. Woodger, 'The concept of truth in formalized languages', in Tarski, eds. Woodger & J. Corcoran, *Logic, Semantics, Metamathematics* (Indianapolis: Hackett, 1983), pp.152-278.
- Wiggins, D.R.P. 1986. 'Verbs and adverbs, and some other modes of grammatical combination'. *Proceedings of the Aristotelian Society* **86**: 273-304.
- . 1991. 'Truth, and truth as predicated of moral judgements'. In his *Needs, Values, Truth: Essays in the Philosophy of Value*, 2<sup>nd</sup> ed. (Oxford: Blackwell), pp.139-84. (An edited version of a paper originally published in 1987.)
- . 2002. 'An indefinibilist cum normative view of truth and the marks of truth'. In R. Schantz, ed., *Current Issues in Theoretical Philosophy, Volume 1: What is Truth?* (Berlin: Walter De Gruyter), pp.316-32.



Williamson, T. 1996. 'Knowing and asserting'. *The Philosophical Review* **105**: 489-523.

———. 1998. 'Indefinite extensibility'. *Grazer Philosophische Studien* **55**: 1-24.

———. 2020. *Suppose and Tell: The Semantics and Heuristics of Conditionals*.

Oxford: Oxford University Press.