

Second Thoughts about Criteria

Author(s): Crispin Wright

Source: *Synthese*, Vol. 58, No. 3, Essays on Wittgenstein's Later Philosophy (Mar., 1984), pp. 383-405

Published by: Springer

Stable URL: <http://www.jstor.org/stable/20115974>

Accessed: 21-11-2017 14:16 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



Springer is collaborating with JSTOR to digitize, preserve and extend access to *Synthese*

CRISPIN WRIGHT

SECOND THOUGHTS ABOUT CRITERIA

This paper continues themes with which I was concerned in a lecture given at the invitation of the Royal Institute of Philosophy in 1978.¹ The main aims of that lecture were to sharpen the formulation of what I took to be the notion of a criterion introduced in the works of the later Wittgenstein, and to display its connection with anti-realist ideas on meaning. I am much less confident than formerly about the Wittgensteinian pedigree of the notion I was trying to sharpen, though the formulations of the lecture may still be of value in explicating the concept of a criterion that the philosophical community largely believes, rightly or wrongly, to have originated in Wittgenstein's work. And I still hold that those philosophers of language who believe that assertion-conditions, rather than truth-conditions, should provide the basis for the philosophical theory of meaning are committed to explaining and utilising some such concept. But it seems to me very much less clear now whether the concept can play the vigorous anti-sceptical role in the theory of knowledge that its advocates have confidently expected of it; or whether, indeed, as ordinarily interpreted, it is so much as coherent.

Orthodoxy in the interpretation of Wittgenstein attributes to criteria five cardinal features:²

that recognition of satisfaction of criteria for *P* can confer sceptic-proof knowledge that *P*;

that *P*'s criteria determine *necessarily* good evidence for *P*, and thereby fix its content; that the criteria for *P* will typically be multiple;

that satisfaction of a criterion for *P* will always be a 'public' matter;

and that to know of the satisfaction of criteria for *P* is always consistent with having, or discovering, further information whose effect is that the claim that *P* is not justified after all.

I shall canvass reasons for thinking that it is seriously unclear whether, with or without a realist background, the fifth feature - *Defeasibility* - can be made to harmonise either with the first - *the Knowledge Feature* - or with the second - *the Meaning Feature*. But criteria would not be interestingly different from (public) truth conditions if Defeasibility was

Synthese 58 (1984) 383–405. 0039–7857/84/0583–0383 \$02.30

© 1984 by D. Reidel Publishing Company

waived; and jettison of the Meaning Feature would threaten the Knowledge Feature as well. While to forego the Knowledge Feature would be to deprive the notion of what most advocates have seen as its principal point.

1. THE KNOWLEDGE FEATURE, DEFEASIBILITY, AND SCEPTICISM

It will be useful to consider what response a proponent of criteria ought to make to the following simple objection.³ 'If I claim to know *P*, I will be understood to be claiming that my belief that *P* is guaranteed correct; so I must have a *conclusive* basis for that belief if my claim is to be true. But knowledge that criteria for *P* are satisfied is to be consistent' – by Defeasibility – 'with my obtaining further information as a result of which I no longer have a basis for the belief that *P* at all. Information that is genuinely conclusive for a certain belief, however, cannot lose that status as a result of being *added to*; so knowledge that criteria for *P* are satisfied cannot ever amount to knowledge that *P* – the two states have different essential characteristics.'

The natural response is that this misunderstands the sense in which recognition of the satisfaction of criteria can confer knowledge. To be sure, knowledge that 'criteria for *P* are satisfied' will not, even in the most favourable circumstances, *constitute* knowledge that '*P* is true'. The point is rather that while recognition of the satisfaction of criteria for *P* will, other things being equal, *entitle one to claim* knowledge that *P*, this claim, like the assertion of *P* itself, may have to be withdrawn if certain sorts of development take place. It is, in other words, analytic of the concept of knowledge that whenever one is entitled to make a particular assertion on the basis of satisfaction of criteria for it, one is *eo ipso* entitled to make the corresponding knowledge-claim on the same basis. In this way the concepts of knowledge and of criterion are interlinked, and there can be no coherent sceptical doubt about our entitlement to advance knowledge-claims when appropriate criteria are satisfied. Our entitlement is simply a consequence of the content of those claims; and the would-be sceptic is merely "objecting to a convention".⁴

No one ought to want to complain that this response leaves it unclear how the claim to know that *P* can differ in content from the claim that *P*. Their behaviour when embedded in more complex contexts is

obviously quite different; for example, the circumstances that justify my denial that I know *P* need not, obviously, justify a denial of *P*. In any case, there is no suggestion that *every* type of ground for asserting *P* is to be a ground for claiming to know *P*, but only that the criteria for *P* are criteria for the knowledge-claim as well. But what the response does leave unclear is in what precise way the content of the two claims is supposed to differ; that is, in effect, what *knowledge* is being taken to be. If recognition that criteria for *P* are satisfied can justify a knowledge-claim that *P* without *constituting* knowledge that *P*, what would constitute it?

The question poses a *prima facie* awkward dilemma for a believer in the anti-sceptical powers of criteria. Suppose a subject does indeed know that *P*. Either the facts that constitute the truth of his claim lie within cognitive reach, so to speak, of the subject at the time he makes a justified claim to know that *P* – justified by his recognition of satisfaction of the appropriate criteria – or they do not. But neither alternative seems satisfactory. If the first is chosen, it immediately becomes unclear *why* criteria for *P* are treated as criteria for the claim to know that *P* as well: what is the point of this ‘convention’ when – provided the subject does indeed know that *P* – a further investigation will disclose a state of affairs which *constitutes* the fact? And how, in any case, can the first alternative cope with what seems certain to be a feature of any concrete example, viz., that the claim that *P*, and hence the claim to know that *P*, will always be open to ‘defeat’ by information not yet available? (It is intuitively evident, for example, that information serving to defeat the status of certain behaviour as a ground for the claim that another is in pain could emerge at any time after the behaviour episode.)

If, on the other hand, the state of affairs constituting the truth of the knowledge-claim lies beyond the cognitive reach of the subject at the stage when he, recognising satisfaction of the appropriate criteria, advances the claim, then surely a gap now opens on which the sceptic can pounce. For now it will be all the same to the subject at that stage whether the truth is that he knows that *P* or that he does not: his very best safeguards leave it open whether the knowledge-claim that they justify is actually true. The consequent sceptical query is this: how can our adoption of a *convention* that realisation of one state of affairs, viz., satisfaction of criteria for the knowledge claim, provides a reason for claiming the existence of another, quite distinct, state of affairs, viz.,

that in which the truth of the knowledge-claim consists, be *reasonable*? There are, trivially, plenty of cases in which such a convention would be foolish. People might have the convention that snowdrops by the end of January justify the claim that March will be dry; but such a convention would hardly be distinguishable from a superstition, and open to objection in exactly the same way. So how is the needed distinction to be drawn? An inductive, or significant statistical correlation between realisation of instances of the two types of state of affairs might serve to repel the charge of superstition, but once the proponent of criteria has allowed things to go so far that *that* sort of answer seems appropriate, he will have to respond to the sort of sceptic – the sceptic about induction – whose stock-in-trade it is precisely to query that sort of answer. And no sort of deployment of the notion of criterion is going to help; on the contrary, the response is owed in order to prop the notion up.

The root of the problem on both horns is, obviously, the logical distinctness – the fact that either may be realised without the order – of the states of affairs respectively constituting satisfaction of the criteria for, and the truth of, the relevant knowledge-claim. To repeat, if the truth-conferring state of affairs is accessible independently to the claimant, why settle for recognition that the criteria are satisfied? And if it is inaccessible to him, how can it be a priori reasonable to rely on satisfaction of criteria as an indication of its existence? So the troublesome logical distinctness will have to be prevented. But how? For, to stress, recognition of satisfaction of criteria for *P* cannot, consistently with their defeasibility, be held to constitute knowledge that *P* without contravention of the truism that knowledge entails the truth of what is known.

A possible thought is that really there *is* no such truism; that belief in the truism only arises because of a mistaken interpretation of what seems to be its contrapositive. Undoubtedly it is a feature of the concept of knowledge that the claim to know *P* must be withdrawn if the balance of evidence turns against *P*. But other explanations of this fact are possible besides appeal to the alleged truism. It would, say, equally be explicable if knowledge-claims had an appropriate token-reflexivity. Suppose, for instance, that what constitutes the truth of a knowledge-claim that *P* is some feature of the state of information of the claimant at the time of the claim: then subsequent information telling, on balance, against *P* may be inconsistent with possession of

that feature by the *later* state of information, so may require denial of the form of words formerly asserted. But the original knowledge-claim may no more be contradicted than would yesterday's claim, "It is raining", by today's assertion of "It is not raining."

Accordingly, a proponent of criteria might seek to identify the content of the claim to know *P* with, e.g., that of the claim that one's present state of information includes awareness of satisfaction of criteria for *P* and of no consideration defeating the warrant for *P* which that supplies. Such a proposal slips past the dilemma. But, leaving on one side its handling of the disputed truism, it is surely false to the intuitive concept of knowledge in another important respect. A rational, perfectly recollective being could never *lose* any item of knowledge, as we ordinarily conceive it; whatever he came to know he would always know. This is why one cannot unparadoxically advance a claim of the form that one formerly knew some statement to be true which one knows no longer: any adequate grounds for the first part of the claim would be treated as adequate for the claim to know the statement at the present time. In terms of the proposal, however, there is nothing paradoxical about such a claim; the situation will be merely that one's previous state of information criterially supported the claim that *P* whereas one's present, enlarged state of information does not.

That the proposal is revisionary is of no great importance unless there are independent grounds for thinking that the changes it enjoins are worth avoiding. But it wants, to say the least, a deal of explaining how the promise to save the possibility of knowledge from the depredations of the sceptic can be redeemed by a *revisionary* proposal. For the sceptic precisely contests our entitlement to knowledge-claims if knowledge is to be a *stable, truth-entailing* state. The proposal, in effect, gives the sceptic the point and recommends that we work with a more modest concept. Those who have seen the notion of a criterion as epistemologically important surely did not mean that its role would be, in this way, one of retrenchment.

The reader may for some time have been anxious to urge the following thought. The foregoing difficulties arise only because, granting that to recognise satisfaction of criteria for *P* may warrant a defeasible knowledge-claim that *P*, we have continued to press for a further account of the *truth-conditions* of that claim. But if a criterial semantics is, as proponents of criteria have suggested, something aimed to replace the classical truth-conditional line of goods, such a demand

will be illegitimate. There will be nothing in which knowledge *consists*, if that requires the possibility of an explanation of the truth-conditions of knowledge-claims. Rather the content of such claims, as of any statements (except effectively decidable ones), will be given by explaining the criteria for making them and the conditions under which those criteria should be considered defeated and the statements withdrawn.

This can seem like a smart reply; and it is, in any case, inevitable for a theorist who has it in view that the meaning of all statements, (save possibly the members of a base-class of effectively decidable ones) is to be conceived and understood by reference to criteria. What is harder to see is that the sceptic is really in any way disarmed by the move. To appreciate the remaining difficulties, reflect on the simplest form of the problem of induction. Most simply formulated, the problem is that of justifying belief in an unrestricted generalisation that might yet be falsified by unobserved cases. What is striking is that there is no need to appeal to the notion of the *truth* of such a generalisation in order to present the problem. Let us take it, simplistically no doubt, that such a generalisation is assertible if an appropriate variety of cases has been examined and only positive instances have been encountered; and that it is deniable if one has reason to think that there is at least one negative instance. Now, if those conditions are regarded as fixing the meaning of such generalisations, it is apt to seem that there can be no coherent sceptical query whether such a statement *ought* to be asserted in the appropriate circumstances; and the same holds for the claim to know the generalisation if it is held that the identical evidence criterially warrants its assertion as well.⁵ But the sceptic is not to be shaken off so easily. For the fact is that we do not responsibly make such assertions with an open mind about subsequent defeat. If I assert such a generalisation, believing myself fully justified, I will have determinate expectations about the way that future evidence-gathering will, or would, turn out. I shall be, and consider myself entitled to be, *surprised* if negative instances come to light. In this respect assertion is quite unlike the making of moves in, say, chess. I may, perfectly justifiably, advance a bishop, say, although fully expecting a situation to arise in which it will be necessary to retreat again. But I cannot *justifiably* assert an unrestricted generalisation, or indeed any statement, if I fully expect my present warrant to assert it to be overturned by subsequent developments – *or even merely consider that I am in no position to have a view on the likelihood of that eventuality.*

The classic form of the problem of induction trades, no doubt, on a distinction between the correct assertibility of *H* in a particular state of information and the truth of *H*; where the latter is taken to entail its correct assertibility in any state of information that can in principle be arrived at by enrichment of the original. Clearly *that* form of the problem cannot survive a successful attempt to describe the meaning of *H* without reference to any truth-conditions, properly so conceived. But, equally, some version of the problem is liable to survive so long as *any* notion of the correctness of *H* is in play distinct from its correct assertibility in an actual state of information. And such a distinction would appear to be implicit in the very notion of justified assertion. If the achievement of a total state of information in which the assertion of *H* is warranted is taken to entitle one to the expectation that subsequent states of information will *not* be of certain sorts, there cannot be any cogent objection to the introduction of a sense of *incorrectness* to go along with disappointment in that expectation. And the sceptic's question will then be, for instance: why does the warranted assertibility of *H* in our present state of information provide a reason for thinking that our state of information next Tuesday, say, will not show that our earlier assertion was, in just that sense, incorrect? And he can't be answered, of course, by insisting that it is merely part of the *content* of *H* that that is so, that to have reason for believing *H* is to have reason for believing in the *H*-favourable character of subsequent states of information. Such a response merely invites him to re-present his question as: how is it *reasonable* to associate a statement possessing those (ambitious) consequences with assertion-conditions of the relevant sort?

The example of induction is merely one illustration of a form of sceptical difficulty that must remain even after the adoption of a thorough-going anti-realist semantics based upon criteria. The content of a claim cannot be *exhausted* by the statement that a criterial basis for its assertion exists. The reason for saying so is not that it is otherwise inexplicable why it is *that* claim that has to be withdrawn if subsequent states of information are unfavourable – (various skirmishes to do with token-reflexivity would have to be fought through before that could be seen as a conclusive reason for the point). Rather, it is part of considering oneself fully justified in making a particular assertion – any assertion – that one anticipates that the sorts of development that would defeat that justification will not take place. To suppose that one had full reason for the assertion, yet no reason to discount such developments,

would just be to pre-empt the ongoing practical consequences of the assertion; subsequently, it would be all the same whether the assertion had previously been justified or not. But it is of the essence of asserting that one seeks to transfer information which can be *acted on*. Hence that, for example, certain sorts of behaviour are taken to justify, defeasibly and as a matter of convention, certain sorts of ascription of mental states to others cannot squeeze out all scope for scepticism about statements concerning other minds. Certainly, the character of the sceptical problem changes. There is no longer a problem of explaining how it can be reasonable to take what is 'outer' as evidence for what is 'inner'; (no problem conceived as it has to be conceived in order to make Argument from Analogy a relevant, if very weak, response.) The problem is rather to explain how it can be reasonable to suppose that a feature of one's present state of information – that it includes awareness of satisfaction of behavioural criteria for the claim that Jones is in pain and of no defeating circumstances – is going to survive into one's later states of information. Even now the temptation may remain to protest that it is precisely for the reasonableness of that supposition that the criteria are satisfied. But that is only bluster. The plain fact is that there is an inference from present to future – one among a range of alternative possible inferences, whose selection, the sceptic will insist, is therefore in need of defence.

The focus of the earlier part of this section was on the still popular claim that criteria provide a new and powerful weapon for the defence of knowledge-claims against the sceptic. The truth is, however, that knowledge is something of a red herring in discussions of scepticism! It would not be too disturbing to have to cede the possibility of knowledge if we could at least defend the possibility of *reasonable belief* against sceptical assault. "Perhaps we do not, strictly speaking, ever *know* what another is feeling or thinking, what the sun will do tomorrow, etc . . . but can merely have (highly) reasonable opinions on such matters."⁶ However, the nastiest forms of traditional scepticism call into question not merely our title to knowledge-claims of problematic subjects matters but our right to any rationally supported opinions at all in those areas. These forms demand some sort of critical response; retrenchment is not an option. What (I am suggesting) must be recognised is that one such form of scepticism is endemic in the very nature of assertion: specifically, the *consequential* character of assertion, whereby the justifiability of an assertion in a particular state of information is *eo ipso* taken to license expectations about the character,

in relevant respects, of subsequent states of information. This guarantees a sceptically exploitable gulf whenever what is asserted does not admit of conclusive verification. And it is, to emphasise, an essential feature of asserting: it is impossible to see what it would be to put to practical use the 'assertions' of others if the convention was that no such expectations were licensed. So it is a feature that will be preserved by any worthwhile theory of meaning, including, if such a thing is possible, a criterial theory.

It is thus quite unclear how the promise of the Knowledge Feature can be fulfilled. On any account, the claim (to know) that Jones is in pain will be defeated by appropriately disconfirming, or equivocal, subsequent behaviour from Jones. It is part of being justified in making that claim about Jones that one also be justified in the belief that that possibility will not be realised. That the relationship between Jones' *present* behaviour and the justifiability of the claim, 'Jones is in pain' is one of convention cannot, from the sceptical viewpoint, makes the possibility any more remote. No doubt there are indeed things seriously amiss with the sceptical viewpoint. But if what I have said is right, it is doubtful whether the proponent of criteria occupies a position in any way advantageous for their disclosure.

2. DEFEASIBILITY AND THE MEANING FEATURE

In the lecture I argued that an approach to meaning that sees conditions of warranted assertion rather than truth-conditions as central would be bound to find work for something "interestingly akin to" Wittgenstein's notion of a criterion (as orthodoxy interprets it). The argument was, specifically, that at least some types of assertion-conditions for certain sorts of statements would have to be regarded both as having that status *a priori* and as defeasible. And it was suggested accordingly that difficulty in disclosing criteria for statements of the appropriate sort – in particular, statements about the past – would have to be regarded as calling into question the feasibility of a 'global' assertion-conditions conception of statement meaning. The doubt with which we are now concerned, however, is not whether criterial assertion-conditions can be found in *every* case where – according to this type of anti-realist conception of meaning – they ought to exist, but whether they *ever* can: whether the Meaning Feature and Defeasibility can be coherently combined at all. (If the doubt is sustained, it must, of course, if the

argument of the lecture was correct, immediately be put in question whether the meaning of statements can in general correctly be conceived as determined by conditions of warranted assertion.)

Let me recapitulate the essentials of that argument. The reason for thinking that the assertion-conditions theorist must, if he is to account for our understanding of enough of the statements which we want to regard as intelligible, regard some of them as associated with nothing but defeasible assertion-conditions, is straightforward enough. The most one can ask, if someone is to *show* that he understands a particular statement, is that he displays, as conclusively as merely finite samples of his behaviour can display, a practical grasp of the distinction between states of affairs that may reasonably be taken to warrant its assertion and states of affairs that may not. (Giving satisfactory verbal explanations of the content of a statement, and showing that he understands something of its deductive ancestry and offspring, supply grounds for attributing understanding only in so far as they are reliable indicators that someone possesses this type of practical grasp; and they are defeasible by behaviour that suggests that he does not.) The anti-realist, as is familiar, is impressed by the thought that a description of what understanding essentially is ought as little as possible to go beyond what someone is able to do who can supply the most one can reasonably demand by way of evidence for his understanding: there is to be in the notion of understanding as little theoretical residue as possible in excess of what is necessary for description of the essential skills of a language-master. But the classical truth-conditional conception of meaning breaks faith with this thought wherever it construes the kind of state of affairs that would render a particular statement true as something that can impinge only indirectly, or in part, on the faculties on one who understands the statement. Putative truth-conferring states of affairs which, for a particular agent, lie too far in the past for non-inferential recollection, or which lie in the future, or are spatially too distant from him to be visited, or are somehow infinite in extent, or are conceived to lie within the 'inaccessible confines of another's consciousness', cannot be the kind of states of affairs by reference to which his understanding of statements can be tested; testing must rather proceed by reference to conditions that we, the testers, can monitor and of whose obtaining our subject can be presumed to be perceptually aware. For any class of statements whose intelligibility he wishes to save the anti-realist has therefore no choices save *either* to describe truth-conditions for them of

this 'unproblematic', accessible sort, *or* to deny that any of their assertion-conditions amount to truth-conditions.

The first response is, in effect, that of the mathematical intuitionist: proof, for the intuitionist, is to be seen as what *constitutes* truth in pure mathematics. But with most of the relevant sorts of contingent statements, the sort of reductionism involved – of the past and future to the present, of the mental to the behavioural and/or physical, of the spatio-temporally infinite to the finite – would be bound, it seems, massively to misdescribe our ordinary conception of their meaning. And that sits ill alongside the anti-realist's motivation to describe more accurately than the realist in what an understanding of these statements consists. So for these statements he is well advised to prefer the second response: their meaning will be explained by reference to conditions of warranted assertion whose obtaining is not sufficient for their truth.

That does not immediately entail that such conditions will provide at most defeasible grounds for the assertion of such statements. From a realist point of view there would be no absurdity in the idea, for some particular statement, that although our best evidence could not be conclusive for its truth, we could never amass additional, overriding evidence suggesting either that it was false, or at least that its truth should not be assumed. (An example of that sort of situation might be a scientific hypothesis which, although indirectly corroborated as part of a successful theory, could not be *specifically* tested because the laboratory conditions required to do so would require the acceleration of particles up to the speed of light.) But from an anti-realist point of view, the circumstance that standard assertion-conditions for a statement were not sufficient for its truth would have to be manifestable in the *use* of the statement; and the only way, it seems, in which that could be manifest is if the obtaining of those conditions were, under certain further circumstances, acknowledged as insufficient for the correct assertibility for the statement. Hence whenever the anti-realist takes the second option, his assertion-conditions will be, in just the sense that has informed our discussion so far, defeasible.

The reason for thinking that, for the assertion-conditions theorist, every contingent statement must be associated *a priori* with at least some of its assertion-conditions is probably best presented by reductio. Let *P* be a contingent statement of definite content for which that is not the case; that is, for which it will betray no misunderstanding to demand, for *any* state of affairs presented – descriptively or ostensibly

– as warranting the assertion of *P*, *empirical* grounds for thinking that it does so. And let it be the case that *P* is indeed warrantably assertible in certain actual circumstances, Φ . How is the fact to be recognised? Only two forms of case seem possible. First, we may already possess empirical grounds for regarding the occurrence of Φ -circumstances as a good indication of the occurrence of certain circumstances of *another* sort that we already have independent reason to regard as warranting the assertion of *P*. But plainly this model cannot be *generally* applicable, on pain of infinite regress. The second case is that Φ -circumstances in general, or at any rate these ones, have been noted to satisfy some condition *F*, which we again have an independent empirical reason to suppose characteristic of circumstances that justify the assertion of *P*. But since this is an *empirical* reason, it must consist, it would seem, *either* in grounds for regarding *F*-circumstances as symptomatic for the existence of another sort of circumstance already justifiably regarded as warranting the assertion of *P* – in which case we revert to the first model; *or* in grounds for regarding *F* as tending to be coinstantiated with some yet further condition, whose instances we already have reason to regard as justifying the assertion of *P* – which again, on pain of infinite regress, cannot *always* be the story. Either way, then, *no finite* model is to be had of how the warranted assertibility of *P* in Φ -circumstances is to be apprehended. So it cannot be apprehended. Hence, generalising on Φ , we are not going to be able to make *any* application of the distinction between correct and incorrect assertion of *P*. So no-one will be able to manifest an understanding of *P* in the sort of way the assertion-conditions theorist demands; wherefore *P* cannot rank, on his view, as a statement of definite content, contrary to hypothesis.

All this argument does is to elaborate somewhat, for the special case of statements and the assertion-conditions conception of their meaning, what follows from denying the plausible thought that, in order for an expression to have a determinate meaning at all, it must at some level be a matter of *convention* what its correct use consists in: that any significant empirical investigation into whether it is correctly used in a particular case presupposes such a convention.⁷ The realist need not, of course, be hostile to that thought; but for him the relevant convention(s) for *P* will concern which states of affairs confer *truth* on *P*. It is therefore worth briefly pausing to inquire what the realist response should be to the argument sketched. Plainly he is at liberty simply to accept the argument if *P* happens to be *verifiable*; that is, is associated

with truth-conditions whose realisation we can recognise as such. But our interest is in the contrasting sort of case where the anti-realist finds the truth-conditional account problematic. If *P* is verification-transcendent, how should the realist view the status of the circumstances which we should reckon to warrant its assertion? If he accepts the argument, he will allow that certain of these circumstances have that status by convention. But what, in that case, is the function of the putative *further* conventional association with an (unrecognisable) truth-condition? What aspect of the use of *P* is the further convention needed to explain?⁸ To reject the conclusion, on the other hand, that certain of the assertion-conditions of *P* have that status a priori, is just to accept that *P* fits the premise for the reductio. So how, in that case, does the realist block the argument? Well, since for him the only conventions in the offing concern the assignment of verification-transcendent truth-conditions to *P*, he needs to show how these conventions render navigable either of the two routes towards recognition that Φ -circumstances warrant the assertion of *P* which the argument sketches. So *either* he must explain how we could arrive at empirical grounds for supposing Φ -circumstances – or any ascertainable circumstances – to be an indicator of a verification-transcendent state of affairs; *or* he must explain how we might gather empirical reason for supposing that circumstances satisfying *F* – or some other decidable condition – warrant a verification-transcendent claim. Either way, the evident problem is to make out how *experience* can teach us that the occurrence of certain factors within our experience reliably indicates the character of states of affairs beyond it.

Assuming its soundness, then, the argument does more than disclose a commitment to criteria on the part of the assertion-conditions theorist. In addition it brings into relief a dilemma for the realist. To recapitulate: what, in his view, is to be the relation between the assertion-conditions of verification-transcendent *P* and the truth-conditions he regards as determining its content? If the former are regarded as merely contingent symptoms of the latter, the realist can expect to encounter grave difficulties when he comes to try to explain how we are to gather evidence that such indeed is the relationship in a particular case. But if certain of the assertion-conditions are regarded as having the status by convention, it is a very nice question why truth-conditions have any genuine semantic role to play at all. (And it is, independently, quite unclear with what right we could regard the occurrence of one

type of state of affairs as a *conventional* indication of the occurrence of another, logically independent kind.)

So much, then, for rehearsal of the reasons why the assertion-conditions theorist has work for defeasible but conventional assertion-conditions, conditions apparently exemplifying the fifth and second features in our original list. What is the problem in combining these features supposed to be? The crucial consideration is again the *consequential* character of assertion, stressed above. When someone asserts *P*, even on grounds which are admitted to be inconclusive, he sets himself against the subsequent defeat of those grounds. In particular, the discovery of information, additional to his original grounds, which requires withdrawal, or even denial, of his original assertion must be a possibility which recipients of that assertion are, for practical purposes, intended to be entitled to ignore. Nevertheless, defeat always *is* a possibility where criteria are concerned. And it will be in the lap of the gods both whether it occurs in any particular case, and *how often* it happens that a particular type of criterial ground for *P* is subsequently overturned. It is to be expected, no doubt, – if only for ‘evolutionary’ reasons – that we will in general have selected criteria so as to minimize the frequency of defeat. But that is not to say that it is certain that we have been successful in any particular case, or that success will last. So, in the absence of argument to the contrary for particular cases, it has to be a possibility that a type of criterion for a particular assertion be defeated often enough to shake our confidence in the propriety of that assertion when made solely on that ground. It would appear to follow that no type of ground, even one conventionally associated with *P*, can be *necessarily* ‘good evidence’ for *P* if it is regarded as a defeasible ground: for if it is defeasible, it may be defeated; and defeated moreover, so frequently that, bearing in mind the consequential character of any assertion of *P*, one would rightly become reluctant to assert *P* on its basis.

This simple point is worth relating to two formulations suggested in my [8]. The first, taking the criterial relation as obtaining between statements, was the following *Criterial Schema*:

If an agent was verified each of $\{D_1, \dots, D_n\}$, and possesses no reason to doubt *P* and no information which would explain, without the need to suppose *P*, why $\{D_1, \dots, D_n\}$, are true, then it is reasonable for him to believe *P*.⁹

Here it is to be understood that $\{D_1, \dots, D_n\}$ are to be decidable state-

ments – (whatever that means) – and P a verification-transcendent statement. Thus knowledge of the truth of $\{D_1, \dots D_n\}$ will be at best an inconclusive ground for P . It is also to be the case that each of $\{D_1, \dots D_n\}$ is *P-predictive*: that is, is a statement understanding which does not presuppose an understanding of P . (The intended effect of this restriction was to ensure that $\{D_1, \dots D_n\}$ should not entail that it is reasonable to believe P *independently* of the provisos that the agent possesses no information telling against P and no information which would explain, without the need to suppose P , why those statements are true.)

Now, the thought was that $\{D_1, \dots D_n\}$ will collectively constitute a criterion for a particular P just in case their introduction, with P , into the Schema results in a noncontingent truth. The fact is, however, that is it actually quite unclear whether that can ever be so. For, if upon nothing else, the truth of such an instance of the Schema has to be contingent at least upon the frequency with which verification of $\{D_1, \dots D_n\}$ is subsequently associated with the development of considerations that overturn their evidential status. Any alleged example of the criterial relation is going to be controversial, but, to make the point vivid, let $\{D_1, \dots D_n\}$ be as meticulous a description as you like of Jones's very typical behaviour and appearance as he sits miserably in the Dentist's waiting room awaiting treatment for a (claimed) badly infected molar; and let P be 'Jones has toothache'. Then it is plausible that the Schema is at least true: that is, that someone who knows of the truth of the appropriate $\{D_1, \dots D_n\}$ and has no reason to doubt that Jones has toothache, nor any better explanation of the facts recorded by $\{D_1, \dots D_n\}$ than to suppose that Jones has toothache, is indeed in position reasonably to believe that Jones has toothache. But it is also plausible to suppose that it is contingent that this is so: that developments could have taken place as a result of which $\{D_1, \dots D_n\}$ would not have had this power, even though both provisos – that the agent had no independent reason to doubt that Jones had toothache, nor a better explanation of his behaviour than to suppose that he did – were met. For it might have been, e.g., that – in order to avoid compulsory games, or whatever – there had been a very great deal of toothache simulation going on; but only sufficient to give cause for diffidence in drawing any firm conclusion about Jones – so, consistently with the provisos, not enough to provide a reason for doubting the veracity of his behaviour, or for proposal of some alternative explanation of it.

There is, to be sure, scope for debate about whether toothache is a plausible candidate for possession of behavioural criteria. But it should be obvious that the point depends only upon the *structure* assigned by orthodoxy to the notion. If criterial grounds are always to be defeasible, then they must, if the world turns out sufficiently awkward, be defeasible in whole classes. The point is merely a consequence of the logical independence of the conditions for satisfaction and defeat respectively of a given (type of) criterion: if things can go badly at all, they can go badly often enough to put the utility of the very practice (of asserting *P* on the basis of that criterion) in doubt – which need not be to say, however, that *P* itself now becomes doubtful when that discredited (type of) criterion is satisfied.¹⁰

The second proposal, conceived for a view of the criterial relation which holds that its domain comprises not statements but worldly states of affairs, was this:

The occurrence of a particular (probably ostensibly explained) type of state of affairs provides *conventional support* for a particular statement, *P*, if and only if it would standardly be considered acceptable in certain circumstances to assert *P* on the basis of knowledge of such a state of affairs; and we would not require, in order for someone to be credited with a full understanding both of *P* and of the type of state of affairs in point, that he know what it would be *empirically* to investigate whether occurrences of that type of state of affairs really did provide, other things being equal, a reason for believing *P* and to find that they did not.¹¹

The criteria for *P*, so conceived, will be states of affairs whose occurrence provides conventional support for *P* in this sense. (And the “other things being equal” should be amplified in a manner appropriately corresponding to the provisos of the earlier Schema.) Again, though, it would seem that an empty notion has been characterised if criteria, so conceived, are to be defeasible. For anyone who understands both that a certain type of (ostensibly characterisable) state of affairs conventionally supports *P* and that this support is defeasible by developments which, again, he understands and can recognise if they occur, *must* understand how experience could lead one to the conclusion that occurrences of that type of state of affairs did not, after all, provide an adequate reason for believing *P*, or asserting it. The possibility involved is, once again, just that of the world turning

awkward, of too frequent accompaniment of states of affairs of the appropriate kind by subsequent defeating developments.

How should a proponent of criteria respond to these thoughts? A first response would quarrel with the supposition that, because it is possible for criterial support to be defeated in any particular case, it is possible for this to happen often enough to undermine confidence in the basis which the particular type of criterion supplies for the relevant sort of statement. The idea is familiar in discussions of scepticism. Concepts like pretence, or illusion, are themselves rooted, it is suggested, in behavioural, or perceptual, evidence. To suppose such phenomena sufficiently widespread is, in consequence, inevitably to weaken, or even to sever completely, those roots; and so to undermine the content of those very suppositions. The suggestion would be, accordingly, that when we get down to cases, the idea that a particular type of criterion for *P* might, if the world proved awkward, be defeated sufficiently frequently to make us hesitate to rely on it, would prove self-destructive in the same sort of way.

But this is surely a false hope. Matters might be different if defeat of a criterion for *P* had to take the form of the accumulation of additional information warranting the *denial* of *P*: then there might indeed be a tension between the supposition that such defeats might be common and the obtaining of conditions necessary for *P* to possess a sufficiently definite content to render such denials *intelligible*. But defeat of a criterion does not require that; it requires only development of a state of information in which the assertion of *P* should not be made, or, if previously made, should be withdrawn. So there is no prospect of a successful appeal to the conditions of intelligibility of *P* in order to impose an a priori restriction on the frequency of the occasions in which criteria for *P* may be overturned: *worldly awkwardness may precisely take the form of an erosion of those conditions*.

A second response would be that nothing more serious is signalled than a need for further complication in the formulation of the notion of criterial support. We need, for example, it might be proposed, a further proviso in the antecedent of the Criterial Schema. Evidently it is necessary to hypothesise that the agent has no reason to think that the type of ground which $\{D_1, \dots, D_n\}$ affords for statements of the type of *P* has been so often defeated as to inhibit reasonable belief in *P* on the basis of $\{D_1, \dots, D_n\}$. On reflection, indeed, more is wanted: it is

necessary to hypothesise that the agent should have no reason for remaining *agnostic* – keeping a perfectly open mind – about the likelihood of developments which defeat the support afforded by $\{D_1, \dots, D_n\}$ for P . No weaker proviso, it appears, will do. I may, for example, have verified an appropriate $\{D_1, \dots, D_n\}$ in the case of Jones's putative toothache, have no reason to doubt that Jones has a toothache, and be in possession of no information which I can use to explain Jones's behaviour without involving the supposition that he has a toothache and yet *still* have no reason to believe in his toothache *if* I am in possession of any sort of information whose effect is that it is reasonable to keep an open mind about the likelihood of developments which would subsequently force me to withdraw the assertion 'Jones has toothache' if I were to make it now. Such information would not have to concern a spate of unreliable toothache behaviour. I might merely have some special information about Jones which cautions circumspection, without giving me reason to doubt that he has toothache or to explain his behaviour in some independent way. So something like the following modified Schema would seem to be suggested:

If an agent

- (i) has verified $\{D_1, \dots, D_n\}$, and
- (ii) has no reason to doubt P , and
- (iii) is in possession of no information which would explain, without the need to suppose P , why $\{D_1, \dots, D_n\}$ are true, and
- (iv) has no reason for expecting, or even for retaining an open mind about the likelihood of, subsequent informational states which would force withdrawal of the assertion of P if he were now to make it,

then it is reasonable for him to believe P .

It is, however, unclear whether the Schema, so emended, can be satisfactory. The intended point is to capture, for a given verification-transcendent P , a three-fold distinction among the possible substitutions for $\{D_1, \dots, D_n\}$: substitutions that would generate a non-contingent truth, substitutions which would generate a contingent truth, and substitutions that fail to satisfy the Schema. Otherwise we cannot effect the aimed-for distinction between information that criterially supports P , information that contingently (symptomatically) supports P , and information that has no bearing on the issue. In the presence of the new

proviso, however, it now becomes unclear whether we can contrive any substitutions for $\{D_1, \dots, D_n\}$ that will *fail* to satisfy the Schema. Suppose, as (iv) requires, that I have now, at t , no reason for expecting, or even for keeping an open mind about the likelihood of (*agnosticism* about), subsequent informational developments which would require me to withdraw a claim that P made at t . Suppose, too, that I have every reason to expect a lengthy, healthy, normally progressive cognitive future. Now clearly, if I subsequently arrive at a total state of information which, on balance, fails to justify the belief that P , I *will* then be in a position where I ought to withdraw any former claim that P . Accordingly, letting Q = 'some of my future total states of information will, on balance, fail to justify the belief that P ', the suppositions entail:

1. I have at t no reason to believe Q .
2. I have at t no reason to be agnostic about Q .
3. I have at t every reason to expect to enjoy a lengthy series of, by the large, increasingly superior total states of information.

It is plausible, however, that, for any statement R of determinate content, reflection on my present total state of information must *either* provide reason to believe R , *or* provide reason to doubt R (= believe its negation), *or* – because my information is irrelevant, or equivocal – provide reason for agnosticism about R . Hence, given 1 and 2,

4. I have at t reason to believe the negation of Q = 'all my future total states of information will, on balance, not fail to justify the belief that P '.

In presence of 3, and given that the relevant double negation elimination is unexceptionable, this is tantamount to

5. I have reason at t to anticipate enjoyment of a lengthy, largely progressive series of total states of information all of which will, on balance, justify the belief that P .

And if that is not for it to be reasonable for me to believe P *now*, at t , *why* is it not?

So (iv) threatens to guarantee the truth of the consequent of the Schema on its own; with the result that the Schema would fail to impose any interesting condition on the relation between $\{D_1, \dots, D_n\}$ and P . I shall not here attempt the detailed consideration of the situation necessary to make it clear whether the difficulty is not merely one of

formulation. What should be clear, at least, is that there *is* a difficulty. The proponent of criteria has to construct the Schema, or an analogue for ostensively given states of affairs, in such a way that (i) is supplemented with other provisos so that the three appropriate types of substitutions for $\{D_1, \dots, D_n\}$ are possible for suitable cases of P . Clearly, then, the supplementary provisos must not, singly or in combination, guarantee the truth of the consequent; if they do, the Schema will have *only* noncontingent instances. Equally, however, satisfaction of all the provisos at t must not always be consistent with the agent's simultaneous possession of reason to expect, or at least to keep an open mind about the likelihood of, his subsequent informational states being such as to force withdrawal of an assertion of P made at t . Otherwise the Schema will have *no* noncontingent instances. The difficulty is that if we try to meet the second condition in the most natural and immediate way, we seem to spoil our prospects of meeting the first.

* * * *

We arrive at a position, then, where there seem to be both good grounds for supposing that an assertion-conditions conception of meaning must have recourse to criteria and every possibility that the notion, standardly interpreted, is empty. If, as was suggested above, realism has no satisfactory response to the argument that seems to foist criteria on the assertion-conditions theorist, we are therefore on the brink of a paradox: neither truth-conditional nor assertion-conditional approaches hold out any prospect of a satisfactory account of the semantics of ordinary statements.

Let me briefly indicate how the paradox may perhaps be forestalled. The trouble, expressed somewhat simplistically, is that it is hard to see how defeasible grounds can be *necessarily* good grounds: there just is a tension, curiously unremarked in the literature, between the second and fifth features. But is it criteria in this sense to which the assertion-conditions theorist is committed (assuming the soundness of the earlier argument)? Certainly the argument commits him to holding that if P has communally acknowledged assertion-conditions at all, then some of them must have that status *a priori*. But is that the same thing? On reflection, it would seem that it is not. The force of the argument was the empirical support can be forthcoming for the claim that certain conditions justify the assertion of P only if P is associated with other

assertion-conditions for which empirical support neither exists nor could intelligibly be sought. That is to say: it will be no part of understanding *P* to have any conception of how the status of these assertion-conditions might be empirically confirmed. But that claim is not, as the threatened *modus tollens* implicitly presupposes, the same as the claim that there will be no conception of how their status might be empirically *disconfirmed*: on the contrary, the argument has been that their very defeasibility must serve to supply such a conception. And it is the second claim which is needed if criteria are to supply “necessarily good evidence”.

If the second claim cannot be made good, there is still, therefore, the possibility of a ‘thinner’ notion of criterion, apt for the needs of the assertion-conditions theorist. So that particular research programme may be able to survive even if criteria, as interpreted by orthodoxy, do not. The thin notion will retain that aspect of the Meaning Feature concerned with “content fixing” while eschewing the bit about “necessarily good evidence”.

In is another question whether the thin notion of criterion can assist in the exegesis of Wittgenstein’s (steadfastly unselfconscious) employment of ‘criterion’ and its cognates in the *Investigations*; or whether it can yet have a useful part to play in combatting certain forms of scepticism. What is sure is that those who would have us believe that Wittgenstein discovered in criteria, as standardly interpreted, a tool of the greatest epistemological and semantic importance owe a treatment of the two difficulties with which this paper has been concerned.¹²

NOTES

¹ Wright [8].

² See e.g., Hacker [4], ch. X; McFee [7]; and Baker [3] and [4].

³ The objection was put by John McDowell in his paper read at the second University of Keele Conference on the Philosophy of Language and Logic in March 1981 [5]. Cf. McFee [7], p. 596. The initial development of the objection below owes a lot to the first section of McDowell’s [6].

⁴ Cf. *Brown Book*, p. 57.

⁵ This is something that formerly struck me as obvious; see the paragraph joining pp. 227–28 of Wright [8].

⁶ As Russell used to like to suppose.

⁷ Cf. Wright [8] pp. 231–33 and 235–37.

⁸ Wright [9] p. 120 seems to me now somewhat over-sanguine about the availability of criteria to realism.

⁹ Wright [8] p. 230. – The original formulation was actually: “...and possesses no information telling against *P* and no information which would explain, without the need...”. The present reformulation is intended to resolve an ambiguity in “no information telling against *P*”; viz., that between “no reason to doubt *P*”, i.e., no reason to believe not *P*, and “no reason not to believe *P*”, i.e., no reason *either* to doubt *P* or, more weakly, to suspend belief in *P*.

It may seem that the original Schema would not be vulnerable to the difficulty about to be adumbrated in the text (see also note 10), if the ambiguity were to be resolved in favour of the second, weaker interpretation instead. And that is so. For what threatens to be, however, as it seems to me, a decisive objection to the Schema under the weaker interpretation, keep reading.

¹⁰ To amplify a little: It may not seem evident that the provisos really are met in the toothache case. Would not knowledge of all that attempted deceit have to provide *some* reason for doubting Jones? Perhaps; but if so, the fault lies with a rather weak example. For it is quite clear, in the abstract, that considerations serving to discredit a particular type of ground for an assertion do not *have* to take the form of reasons to doubt that assertion or to propose alternative explanations of the obtaining of the ground: it may just turn out that the ground is *unreliable* – and that it has proved so need provide no reason for *any* view about the status of the assertion this time, or about the correct explanation of this instance of the ground. Most discredited *symptoms* – early holly berries as a sign of a hard winter, premature baldness as a sign of virility, etc., etc. – are in this situation. The heart of the present difficulty for the Schema is that criteria for *P* must also, it appears, be liable to just this form of discredit as indicators of *P*’s continuing assertibility.

¹¹ Wright [8] p. 235.

¹² Some of these negative thoughts were caused by a re-reading of the *Postscript* to Albritton’s [1]. But I hesitate to attribute them to him; and he certainly takes no blame for any faults in my formulation. My thanks, and apologies, are owing to Gordon Baker, Peter Carruthers, Peter Hacker and Graham McFee, who all sent me extensive comments on an earlier draft attempting (unsuccessfully) to persuade me that First Thoughts had been best.

REFERENCES

- [1] Albritton, R.: 1966, ‘On Wittgenstein’s Use of the Term “Criterion”’ in G. Pitcher (ed.), *Wittgenstein: The Philosophical Investigations*, Macmillan, New York.
- [2] Baker, G. P.: 1974, ‘Criteria: A New Foundation for Semantics’, *Ratio* **16**, 156–189.
- [3] Baker, G. P.: 1977, ‘Defeasibility and Meaning’, in P. M. Hacker and J. Raz (eds.), *Law, Morality and Society*, Oxford University Press, Oxford.
- [4] Hacker, P. M.: 1972, *Insight and Illusion*, Oxford University Press, Oxford.
- [5] McDowell, J. H.: 1981, ‘Anti-Realism, Truth-Value Links, and Other Minds’, read at the 1981 Keele Conference on the Philosophy of Language and Logic.
- [6] McDowell, J. H.: 1982, ‘Criteria, Defeasibility and Knowledge’, Henrietta Hertz Philosophical Lecture for the British Academy, November 1982.
- [7] McFee, G.: 1981, ‘On the Interpretation of Wittgenstein’, *Mind* **90**, 592–599.
- [8] Wright, C. J. G.: 1978, ‘Anti-realist Semantics: The Role of *Criteria*’ in *Idealism: Past and Present*, Royal Institute of Philosophy Lectures, 1978–79 (published as a

supplement to *Philosophy* by Cambridge University Press, 1982).

- [9] Wright, C. J. G.: 1980, 'Realism, Truth-Value Links, Other Minds and the Past', in *Papers on Language and Logic*, proceedings of the 1979 Keele Conference, Keele University Library; reprinted in *Ratio* **22**, 1980, 112–132.

Dept. of Logic and Metaphysics
Univ. of St. Andrews
Fife
Scotland