EPISTEMIC THEORIES OF TRUTH:

THE JUSTIFIABILITY PARADOX INVESTIGATED*

Vincent C. Müller & Christian Stein

Universität Hamburg

Abstract:

Epistemic theories of truth, such as those presumed to be typical for anti-realism, can be characterised as saying that what is true can be known in principle: $p \rightarrow \Diamond Kp$. However, with statements of the form "$p \, \& \, \neg Kp$", a contradiction arises if they are both true and known. Analysis of the nature of the paradox shows that such statements refute epistemic theories of truth only if the the anti-realist motivation for epistemic theories of truth is not taken into account. The motivation in a link of understandability ans meaningfulness suggests to change the above principle and to restrict the theory to logically simple sentences, in which case the paradox does not arise. This suggestion also allows to see the deep philosophical problems for anti-realism those counterexamples are pointing at.

In the manifold and vexed debate on realism an argument has surfaced that its proponents take to be a decisive move against theories that can broadly be labelled anti-realist. The argument, first sketched by Frederick Fitch in 1963 but only debated since William D. Hart mentioned it briefly in a paper he presented at the "Joint Session of the Aristotelian Society and the Mind Association" in 1979, is supposed to present us with truths that cannot be known, thus refuting a central thesis of anti-realism.

In this paper, we shall first develop a strong version of a variant of the argument against an anti-realist account of truth and then see what its philosophical significance comes to. Even though we find the argument to be sound, we do not think that it endangers the core of anti-realism. The anti-realist can live with it if he formulates his position on the issue of truth more cautiously – but in line with his initial intentions.

I.

Let us start with a short presentation of the argument as it was developed by Hart. Michael Dummett once formulated, what we shall call the anti-realist principle of knowa-

bility:

> A statement cannot be true unless it is in principle possible that it be known to be true. (Dummett 1973, 465)

Dummett stresses, that the constraint 'in principle' must be understood as 'in principle *by us*'; omniscient verificators are no option for the anti-realist. A formal version of the Knowability Principle is:

(K) $p \rightarrow \Diamond Kp$

Now, there are surely some truths, which are actually and contingently unknown, so:

(1) $p \ \& \ \neg Kp$.

If (K) holds for all sentences, it must be possible to know (1):

(2) $\Diamond K(p \ \& \ \neg Kp)$

Knowledge distributes over conjunction, therefore:

(3) $\Diamond(Kp \ \& \ K\neg Kp)$

Now the troubles for the anti-realist are obvious; in the second conjunct knowledge implies truth, so we have a contradiction:

(4) $\Diamond(Kp \ \& \ \neg Kp)$

As we have seen, inserting (1) into the anti-realist schema (K), yields the consequent (2). Since we know that this is false, we can deny the antecedent (1) via *modus tollens*, and arrive at:

(5) $\neg(p \ \& \ \neg Kp)$

This is equivalent to

(6) $p \rightarrow Kp$

It is thus shown that the principle of knowability implies that everything is known.

II.

In 1992, Wolfgang Künne presented a version of the argument that uses the weaker notion of *justifiability*. He applies it to truth theory, more precisely, to epistemic accounts of truth, which are intimately connected to anti-realism. It is against such an epistemic notion of truth, as expressed in the writings of Dummett and the later Hilary Putnam, that Künne's argument is directed. In the eyes of Putnam:

> … truth is an *idealization* of rational acceptability. We speak as if there were such things as epistemically ideal conditions and we call a statement 'true' if it would be justified under such conditions. (1981, 55)

The talk of epistemically ideal conditions is a metaphor to suggest a situation in which

we are ideally situated to decide the truth of a statement, no deception of any kind being involved (see Putnam 1990, vii–ix). Let us assume, for the sake of argument, that the anti-realist can define these conditions in a non-circular way. Again, truth is supposed to be tied to the cognitive capacities of human beings; any other sense is said to be inconceivable for us (Putnam 1981, 64), and this seems to be a central idea for an anti-realist theory of truth.

Künne gives the following formulation of the account of truth common to Putnam and Dummett, which we shall call the *Justifyability Principle*:

> (P/D) If a statement is true, a situation must be conceivable in which somebody would be justified to believe that things are as the statement says they are. (1992, 239).

Abbreviating "someone is justified to believe that" as "sjb" we can write it as follows:

> (J) $p \rightarrow \Diamond sjb\ p$

This principle appears to express the core of an epistemic notion of truth; so, if (J) can be shown to be false, any epistemic explanation of truth will be refuted.

Imagine a lonely walk on a stormy day in spring. At time t_0 you admire a tree T full of blossoms, some of which are blown away by a gust of wind at t_1. Nobody counted the blossoms at t_0, and nobody is ever justified to believe that T had some particular number of blossoms at t_0 – but one can always guess what their number was at t_0: let us say the guess is 99. If the guess was right, the following holds:

> (Σ) (A) T has 99 blossoms at t_0 but
>
> (B) nobody is ever justified to believe that this is the case.
> (Künne 1992, 240)

Why should the anti-realist be prepared to accept (B)? Well, we just assume that it so happens that nobody *did* count the blossoms at t_0 and that no other attempts have been made to support any hypothesis about their number at t_0. As things contingently stand, nobody is ever justified to believe that T had a specific number of blossoms at t_0. Clearly, the anti-realist need not be tempted to admit that nobody *could have* been justified to believe that (A). (Σ) is just the conjunction two contingently true statements and there is no a priori reason to think that it could not be true, namely if both its conjuncts are true.

If (Σ) is true then according to the Justifiability Principle it must be possible that someone is justified to believe that (Σ). Let us spell out clause (B) of (Σ) as "¬sjb A". Then we have:

> (7) A & ¬sjb A   (Σ)

inserting (7) in Principle (J) yields:

> (8) (A & ¬sjb A) $\rightarrow \Diamond$sjb (A & ¬sjb A)   Instantiation of (7) in (J)

Künne claims that the consequent of (8) is *not* possible, that (Σ) is necessarily justification transcendent because someone who is justified to believe that (A) falsifies (B) and thus the conjunction. Let us see therefore what the assumption that someone *is* justified to believe (Σ) comes to. The important rule here is that if someone is justified to believe a conjunction then he is justified to believe each of its conjuncts, call it the *Distribution Rule*:

(D) sjb (p & q) → sjb p & sjb q

(Take each of the someone's to refer to the same person.) Now, assume that someone is justified to believe (Σ) as (8) says. Applying our rule to the consequence of (8), we get:

(9) ◊ (sjb A & sjb ¬sjb A)   cf. (3) above

In Hart's version of the argument, the corresponding step (3) is a crucial result: Knowledge entails truth, so we arrive at a formal contradiction. But here, speaking of justification, no such contradiction can be derived. Nevertheless, Künne, Crispin Wright (1986, 427) and others take this to be clearly impossible, a thought Dorothy Edgington expresses by saying that:

> … it is clear that no possible state of information could support the hypothesis: *p* and no one at any time has any evidence hat *p*. (Edgington 1985, p. 558)

Well, (9) does look paradoxical and it is hard to imagine the psychological state of the person in question, but is the situation *really* impossible? Why should it not be possible to have reasons to believe a statement *p*, and on the other hand some reasons to believe that nothing justifies the belief that *p*? This appears to be quite possible if we assume the customary weak concept of justification according to which one can be justified to believe things that are in fact false. So it appears that (Σ) does not refute (J). Nevertheless, sentences such as (Σ) must cause some disquiet for the epistemic truth theorist since it can be shown that it cannot be the case that (Σ) is true while somebody is justified to believe that (Σ). Our principle (J) does not demand that the statement is *true* in the epistemically ideal situation, but for precisely this reason it does not fully bring out the commitments of the epistemic truth theorist. He seems committed to a claim that *implies* (J), the claim that a situation is possible in which a statement *p* is true and somebody is justified to believe that *p*. More generally, the anti-realist holds that there are true statements and that, necessarily, all true statements can in principle be justified. Let us call this principle the *Strengthened Justifiability Principle*; it can be formulated as follows:

(J)* If *p* is true then a situation must be possible in which (*p* & somebody is justified to believe that *p*).

formally:

(J)* p → ◊ (p & sjb p)

In fact, some anti-realists accept the even stronger proposition that epistemically ideal situations are possible, in which somebody's being justified to believe that *p guarantees* the

truth of p – a view that brings them close to verificationism. So, although (J)* is stronger than (J), it is still a somehow cautious anti-realist who only holds (J)*. In effect, this principle is equivalent to the Knowability Principle above, making explicit the anti-realist connection between truth and justification.

Now, insert (Σ) for p into (J)* and you get:

(10) (Σ) $\rightarrow$ $\lozenge$[(Σ) & sjb (Σ)]

If we assume that (Σ) is the case, we can isolate the consequent by *modus ponens*:

(11) $\lozenge$ [(Σ) & sjb (Σ)]   (10), (Σ), *modus ponens*

Spelling out (Σ), that comes to:

(12) $\lozenge$ [A & ¬sjb A & sjb (A & ¬sjb A)]

Applying the distribution rule (D) to the second conjunct, we arrive at a contradiction:

(13) $\lozenge$ (A & <u>¬sjb A & sjb A</u> & sjb ¬sjb A)   (12), (D)

If we take it that (Σ) is true and accept (J)* we get a contradiction, so the (J)* is refuted — the epistemic account of truth seems to be not even extensionally adequate.

III.

The anti-realist has several possibilities to respond to this argument. He may claim for example that (Σ) is not actually a sentence that we can understand. We agree with Künne that (Σ) is perfectly understandable as our linguistic intuitions make us believe; we know the words, know how they are combined an can explain the sentence formed by them.

Another reaction may be to doubt the Distribution Rule. There appears to be no formal proof for such a rule, so let us just appeal to intuition here; does it not sound extremely plausible to claim that if someone is justified to believe the conjunction p & q, he also is justified to believe each of the conjuncts?

Some anti-realists have reacted to Hart's argument by pointing out that it makes use of moves that are invalid in intuitionistic logic. The version presented here does not seem to involve any such moves, however.

Joseph Melia (1991) accepts Hart's argument, but argues from an anti-realist point of view against the principle of knowability. His argument is that statements such as (Σ) can be excluded since they are of a kind whose truth value is affected by answering the question of whether they are known or not – their being known makes them false. Melia's point is a little misleading, since statements as thought of here do not *change* their truth value. Even though statements like (Σ) speak of their own justification, the proposition they express is true or false independently of whether one actually tries to justify them. While we agree with Melia that these statements do have pathological traits and that the anti-realist principle could make room for such sentences, Melia does not really offer an argument why the anti-realist *can* consistently allow for such statements. This

brings us to our proposal, at last.

IV.

In order to see whether the anti-realist really is committed to the Strengthened Justifiability Principle we shall look at his or her *motivation* for holding it: the motivation clearly lies with insights in the philosophy of language an anti-realist takes to be deep and important. We will argue that this motivation is not challenged by (Σ).

Very briefly, anti-realists tend to put questions of meaning in terms of understanding (Dummett 1993, 35) and they argue that a statement cannot be understood unless we can conceive of a situation in which it would be reasonable or even compelling to assert it. We shall call this the 'Understandability Principle' and write it as follows:

(U) $\Diamond Up \rightarrow \Diamond sjb\ p$

The relevant connection to the concept of truth can be made explicit by remembering that not only anti-realists would think that there are no statements which are *principally* beyond understanding:

(T) $p \rightarrow (\Diamond Up)$

The anti-realist comes to defend an epistemic theory of truth by combining the two principles (U) and (T):

(AR) $(p \rightarrow \Diamond Up)\ \&\ (\Diamond\ Up \rightarrow \Diamond sjb\ p)$

from which we can infer the familiar (J), which says that "$p \rightarrow \Diamond\ sjb\ p$". Furthermore, the anti-realist assumes that the statement in question can be true in the justification condition, which takes us to the Strengthened Justifiability Principle (J)*.

Now, we do not just understand (Σ) as a whole but we do understand its *components* and their connection; and we thereby understand the whole. So the requirement for an understanding of (Σ) must be that an epistemically ideal situation is conceivable for each of the *components* (A) and (B) only. We therefore suggest to restrict the Understandability Principle (U) and thus the Principle (J)* to logically simple statements. Let us say that a statement is logically simple, if it contains neither a logical connective like 'and', or 'if … then' nor a quantifier. We take it that logically simple statements are more basic in terms of understandability and also in terms of truth – after all, the truth-value of complex statements is a function of the truth-values of the simple components. The truth of logically complex statements can be defined in terms of the truth of atomic statements. This is also in the spirit of the Dummettian concept of meaning theory which demands to specify the meaning of a sentence in accordance with its logical composition (e. g. 1980, 60).

There are also grounds independent from (Σ)-type statements why one ought to restrict the Principle (U) and thus (J)* to logically simple sentences. The idea that a situation must be *conceivable* where it is reasonable to assent to a statement becomes difficult, if we are confronted with extremely long and complex conjunctions, for instance. As for

the epistemic side, a conjunction of very many statements that have nothing to do with one another may well be understandable while there is not one single conceivable situation in which someone would be justified to believe the conjunction.

The anti-realist did not claim to give an analysis of our concept of truth, he aimed at a philosophical explanation of wider concern. So it is no objection that we did not arrive at a definition for truth – even the initial principle (J) was not formulated as a definition. But have we not arrived at two concepts of truth, one for simple statements and another one for complex statements? Well, there is a difference between the truth of complex and of simple statements, but the two cases are systematically related. For a statement to be true is either to be justifiable *simpliciter* or to be composed in a straightforward way out of justifiable statements. There seem to be no special problems for sentences containing intensional operators, because logically simple statements may contain intensional operators. And for logically complex intensional statements the same holds as for logically complex extensional ones – their truth value is a function of the truth values of their components. So our proposal is not confined to extensional language.

Clearly, our suggestion needs to be elaborated upon, particularly since the simplicity in question here is an epistemic simplicity that is supposed to translate into semantic simplicity – and it is far from clear that this must coincide with a notion of logical simplicity. Still, the suggestion is in the spirit of the anti-realist conception of meaning: (Σ) is intelligible and can be true, because (A) is intelligible and not justification transcendent, because (B) is intelligible and not justification transcendent and because the meaning of the logical sentence connective 'and' is straightforward. So, the anti-realist doctrine should concern logically basic semantic units only. It does not matter that one can use these logically atomic components to make up statements that are justification transcendent.

To conclude: The knowability paradox does demand a re-reading of (J)* but it does not offer itself as a *coup de grâce* for epistemic notions of truth. It merely forces the anti-realist to elaborate on its theses concerning understanding and on the connection of those to theses on the nature of truth.

2775 words

References

Dummett, Michael: (1978) *Truth and other Enigmas*, London: Duckworth.

— (1973) *The Interpretation of Frege's Philosophy*, London: Duckworth [2]1980.

— (1991) *The Logical Basis of Metaphysics*, London: Duckworth.

— (1993) *The Seas of Language*, Oxford: Oxford University Press.

Edgington, Dorothy: (1985) 'The Paradox of Knowability', *Mind* 94, 557–68.

Fitch, Frederick D.: (1963) 'A Logical Analysis of some Value Concepts', *The Journal of Symbolic Logic* 28/2, 135–142.

Hart, William D.: (1979) 'The Epistemology of Abstract Objects—Access and Inference', *Proceedings of the Aristotelian Society* Suppl. Volume 53, 153–165.

Künne, Wolfgang: (1992) 'Bolzanos blühender Baum', in: Forum für Philosophie Bad Homburg (eds.): *Realismus und Antirealismus*, Frankfurt/Main: Suhrkamp, 224–244.

— (1994) *Theories of Truth*, forthcoming.

MacIntosh, J. J.: (1984) 'Fitch's Factives', *Analysis* 44, 153–158.

Mackie, John Leslie: (1980) 'Truth and Knowability', *Analysis* 40, 90–92.

Melia, Joseph: (1991) 'Anti-Realism Untouched', *Mind* 100, 341–342.

Percival, Philip: (1990) 'Fitch and Intuitionistic Knowability', *Analysis* 50, 182-187.

— (1991) 'Knowability, Actuality and the Metaphysics of Context Dependence', *Australasian Journal of Philosophy* 69, 82–97.

Putnam, Hilary: (1969) 'Logical Positivism and the Philosophy of Mind', in: *Mind, Language and Reality, Philosophical Papers II*, Cambridge: Cambridge University Press 1975, 441–451.

— (1981) *Reason, Truth and History*, Cambridge: Cambridge University Press.

— (1990) *Realism with a Human Face*, Cambridge, Mass.: Harvard University Press.

— (1994) 'Sense, Nonsense and the Senses: An Inquiry into the Powers of the Human Mind', *The Journal of Philosophy* XCI, 445–517. (The Dewey Lectures 1994.)

Williamson, Timothy: (1982) 'Intuitionism Disproved?', *Analysis* 42, 203-207.

— (1987a) 'On Knowledge of the Unknowable', *Analysis* 47, 154–158.

— (1987b) 'On the Paradox of Knowability', *Mind* 96, 256-261.

— (1988) 'Knowability and Constructivism', *Philosophical Quarterly* 38, 422-432.

— (1993) 'Verificationism and the Non-Distributive Knowledge', *Australasian Journal of Philosophy* 71, 78–86.

Wright, Crispin: (1986) 'Can a Davidsonian Meaning-Theory be Construed in Terms of Assertibility?', in: *Realism, Truth and Logic*, Oxford: Oxford University Press [2]1993, 403–432.