

The criticisms above relate solely to sentences for which conclusive verification or falsification is not ruled out. For sentences for which this is not the case — and there are a large number of these — different considerations apply from the point of view of both realist and anti-realist. The notion of assertibility the anti-realist applies to such sentences must be taken to be a defeasible one. Therefore, the anti-realist has to alter the account of negation so that a demonstration that a sentence will never be conclusively proved no longer counts as a refutation. This latter must in these cases require further a (defeasible) warrant for asserting that the sentence will never be warrantably assertible (in the defeasible sense of the latter notion). A sentence warrantably assertible at one time may cease to be so at a later time, indeed its negation may become warrantably assertible. This notion of assertibility could not, then, be taken as a mere terminological variant of the classicist's notion of truth which, for some sentences, is invariant over time. Hence, in this case, the anti-realist's position does amount to a rejection of the idea that truth conditions are applicable to the sentences in question.

The realist, in turn, must, of course, reject the idea that understanding the truth conditions of such absolutely undecidable sentences consists in an ability to recognize those conditions whenever they obtain. But this does not mean that the realist must reject a behaviouristic approach to explanation of speakers' understanding of such sentences. For such sentences will be semantically complex, constructed out of simpler components by means of sentential, and other, operators. If determinate extensions, such as truth functions, can be associated with such operators then, given that the simplest components have determinate truth values, so too will the complex sentences. Understanding of the extensions of the operators will, on a behaviouristic theory, be a matter of appropriately patterning responses to complex sentences with the operator dominant, given responses to the constituents. There is no *a priori* connection between such behaviour and behaviour which would manifest an attempted verification, or falsification, of the sentence. Hence there is no *a priori* reason to suppose that a sentence certified as possessing a determinate truth value be in any way decidable. Whether there are such sentences requires, of course, a much more extended investigation than can be given here.

University of Edinburgh,
David Hume Tower,
George Square, Edinburgh EH8 9JX

© ALAN WEIR 1983

NOZICK ON KNOWLEDGE

By B. J. GARRETT

IN his recent *Philosophical Explanations* (Oxford University Press, 1981) Robert Nozick has proposed the following insightful analysis of our concept of knowledge:

S knows that p if and only if:

- (1) p is true
- (2) S believes, via method or way of coming to believe M, that p
- (3) If p weren't true (and S were to use M to arrive at a belief whether (or not) p), then S wouldn't believe, (via M,) that p
- (4) If p were true (in changed circumstances) (and S were to use M to arrive at a belief whether (or not) p,) then S would believe, (via M,) that p.

And Nozick suggests that the possible worlds relevant to the assessment of the subjunctive conditionals in conditions (3) and (4) are those which are (in Lewis's terminology) 'closest to' or 'in the neighbourhood of' the actual world.

However, Raymond Martin ('Tracking Nozick's Sceptic: A Better Method', *ANALYSIS*, January 1983, pp. 28-33) believes that Nozick's analysis has 'a serious counterintuitive consequence' (p. 31). He gives the following example which purports to show that it is possible for an individual with a true belief that p to satisfy conditions (3) and (4) and yet not know that p:

Suppose that S bets on a horse at his local track. The bet S places is such that if S picks the winner in the first race or the winner in the second race, or both, S wins ten dollars; otherwise S loses. S picks Gumshoe in the first and Tagalong in the second. But S has the following peculiar method for arriving at his belief whether (or not) Gumshoe wins the first: S shields himself from any information about the results of either of the first two races, and then, after the conclusion of the second race, S presents his ticket to an automated cashier. (The cashier either pays S for his ticket or not, but cannot, due to the way it is constructed, provide S with any additional information about the results of either race.) If the cashier pays S for his ticket; then S believes that Gumshoe won the first; otherwise, S does not believe that Gumshoe won the first. On the occasion under discussion, Gumshoe wins the first and Tagalong finishes last in the second. Tagalong's performance in the second, moreover, was causally independent of Gumshoe's performance in the first: no matter how Gumshoe had performed in the first, Tagalong still would have finished last in the second. After the second race, with no information on the results of either race, S goes to the cashier and presents his ticket. The cashier gives S ten dollars in exchange for S's ticket. The cashier would not give S ten dollars in exchange for S's ticket unless *either* Gumshoe or Tagalong had won. But the cashier conveys no information to S about whether it was Gumshoe or Tagalong, or both, that won (pp. 30-31).

Now though S satisfies condition (4) it is not clear that condition (3) is satisfied. Martin holds that it is:

... if it weren't true that Gumshoe won the first and S were to use M to arrive at a belief whether (or not) Gumshoe won the first, then (since the cashier would not have paid S for his ticket) S wouldn't believe, via M, that Gumshoe won the first ... (p. 31)

But, of course, this is true only if the possible worlds in which Gumshoe does not win the first and Tagalong wins the second are sufficiently 'far' from the actual world as to be irrelevant to the assessment of the claim that S knows that Gumshoe won the first. But if, as Nozick suggests, the interpretation of the tracking conditionals is to be left at the level of intuitive understanding then there appears to be no reason why such worlds should be excluded. Indeed a Nozickian will argue that it is precisely S's insensitivity in such worlds to the fact that Gumshoe does not win the first (i.e. S's failure to track the fact that not-p) which ensures that S's belief does not constitute knowledge. Hence unless we are told why appeal to such worlds in this context is illegitimate we have no reason to believe that Martin has produced a genuine counterexample to Nozick's thesis.

However, consider the following example which is not open to the above objection. Suppose that a person X arrives at a belief about whether (or not) the father of a particular person A is a philosopher on the basis of being told the profession of the father of another person B (in fact A and B are brothers but X is unaware of this). X adopts the following method for arriving at his belief whether (or not) A's father is a philosopher. If B's father is a philosopher (and X is informed of this) then X will believe that A's father is a philosopher. If B's father is not a philosopher then X will not believe that A's father is a philosopher. B's father is indeed a philosopher and X is subsequently informed of this.

In this example all four of Nozick's conditions appear to be satisfied. A's father is a philosopher; X believes that A's father is a philosopher; if A's father were not a philosopher then, since B's father would not have been a philosopher, X would not believe that A's father was a philosopher; if A's father (in changed circumstances) were a philosopher then, since B's father would have been a philosopher, X would believe that A's father was a philosopher. Hence, on Nozick's analysis, X knows that A's father is a philosopher. But this is surely false — it was just good luck that, on this occasion, X's method happened to be applied to two brothers. (Note that I have here assumed the necessity of causal origin — A and B necessarily have the same father.)

However, there are counterexamples to Nozick which do not rely upon any essentialist principles. Suppose that a person Y arrives at a belief about whether (or not) the number of people in Paris is the sum of two odd primes in the following way. If Y is

told that the number of people in Paris is even then Y will believe that the number of people in Paris is the sum of two odd primes; if Y is told that the number of people in Paris is not even then Y will not believe that the number of people in Paris is the sum of two odd primes. Further suppose that (necessarily) all and only even numbers are the sum of two odd primes, but that Y does not know this, and that the number of people in Paris is even. Y is informed that the number is even. Here again all four of Nozick's conditions are satisfied. The number of people in Paris is the sum of two odd primes; Y believes that the number of people in Paris is the sum of two odd primes; if the number of people in Paris were not the sum of two odd primes then, since the number of people in Paris would not have been even, Y would not believe that the number of people in Paris was the sum of two odd primes; if the number of people in Paris (in changed circumstances) were the sum of two odd primes then, since the number of people in Paris would have been even, Y would believe that the number of people in Paris was the sum of two odd primes. But, via his method, Y does not know the contingent truth that the number of people in Paris is the sum of two odd primes because he does not know that every large even number is the sum of two odd primes.

Nor can Nozick's analysis satisfactorily account for our knowledge of necessary truths. Nozick holds (p. 186) that conditions (1), (2) and (4) are individually necessary and jointly sufficient for S to know that p, where p is a necessary truth. (Condition (3) is here waived since Nozick is unclear how subjunctive conditionals with a necessarily false antecedent are to be handled.) However, consider the following example. A certain mathematician, Z, believes Fermat's Last Theorem (the as yet undecided conjecture that there is no number n such that $x^n = y^n + z^n$ for positive integers x, y, z and $n > 2$) to be true on the grounds that no counterexample has been discovered. And suppose, for the purposes of this example, that Z's belief is true. Since conditions (1) and (2) are satisfied and condition (3) is not applicable, it follows, on Nozick's analysis, that Z knows that Fermat's Last Theorem is true if condition (4) is satisfied. And condition (4) is indeed satisfied — in any possible world Z would believe, via the same method, that Fermat's Last Theorem is true. But Z does not know that Fermat's Last Theorem is true. His belief to that effect was reached via an unreliable method without the aid of any reliable sub-method.

The consequences of these counterexamples are clear. Unless Nozick can resolve the problems which arise for his theory from certain situations in which a true belief that p is based upon a belief that q (where the believer is unaware of the logical or conceptual relation which holds between p and q) and offer an account of subjunctive conditionals with necessarily false antecedents which excludes beliefs about certain true quantified mathematical statements based upon the no-counterexample method from constitut-

ing knowledge, then his unified account of knowledge will have to be supplemented with a condition to the effect that a true belief must be justified or reached via a reliable method if it is known.

The University,
St Andrews, Fife KY16 9AL

© B. J. GARRETT 1983

POST-HYPNOTIC SUGGESTION AND THE EXISTENCE OF UNCONSCIOUS MENTAL ACTIVITY

By DONALD LEVY

FOR as long as psychoanalysis has existed, its central concept, that of unconscious mental activity, has been the object of hostile scrutiny by philosophers. Freud replied to such doubts with two sorts of arguments, the first of which he based on the phenomenon of post-hypnotic suggestion. He tended to employ this argument when explaining his views to non-psychoanalysts and beginners in analysis, perhaps just because it seems to establish his fundamental concept without appealing to any of the distinctive clinical data familiar to practitioners of psychoanalysis.¹ One implication of this argument's independence of psychoanalysis is that the psychoanalytic concept of the unconscious will not be undermined if the argument fails. This needs to be emphasized because the main task of this paper is to show that the argument does fail. Here is the argument as Freud states it

... a person is put into a hypnotic state and is subsequently aroused. While he was in the hypnotic state, under the influence of the physician, he was ordered to execute a certain action at a certain fixed moment after his awakening, say half an hour later. He awakes, and seems fully conscious and in his ordinary condition; he has no recollection of his hypnotic state, and yet at the pre-arranged moment there rushes into his mind the impulse to do such and such a thing, and he does it consciously though not knowing why. It seems impossible to give any other description of the phenomenon than to say that the order had been present in the mind of the person in a condition of latency, or had been present unconsciously, until the given moment came, and then had become conscious. But not the whole of it emerged into consciousness: only the conception of the act to be executed — the order, the influence of the physician, the recollection of the hypnotic state, remained unconscious even then.

But we have more to learn from such an experiment. We are led from the purely descriptive to a *dynamic* view of the phenomenon. The idea of

¹ See, e.g. 'A Note on the Unconscious in Psycho-Analysis' (1912), *Standard Edition*, volume XII, written for presentation before the Society for Psychical Research of London, and 'Some Elementary Lessons in Psycho-Analysis' (1938), *Standard Edition*, volume XXIII.

the action ordered in hypnosis not only became an object of consciousness at a certain moment, but the most striking aspect of the fact is that this idea grew *active*; it was translated into action as soon as consciousness became aware of its presence. The real stimulus to the action being the order of the physician, it is hard not to concede that the idea of the physician's order became active, too. Yet this last idea did not reveal itself to consciousness, as did its outcome, the idea of the action; it remained unconscious, and so it was *active and unconscious* at the same time (*Standard Edition*, volume XII, p. 261).

By the end of the first paragraph, the existence of unconscious ideas has been proved only in the trivial sense that forgotten ideas have been shown to exist, which is merely another way of saying that people sometimes forget things they were aware of earlier. Freud's claim at the end of paragraph one

(a) the idea of the order remained unconscious

is equivalent to

(b) the subject did not remember the order.

The crux of the argument really lies in the second paragraph. For if the hypnotic subject did not perform the action ordered, but merely became conscious of the idea of the act, say, then the sense in which the idea of the order was unconscious would be merely that the order had been forgotten, and no more than that. So Freud is correct in emphasizing in paragraph two, that the argument depends upon some idea — the idea of the act ordered — 'becoming active'. Freud sees the need to show that some idea, while remaining forgotten, 'becomes active'. But the idea of the act does not illustrate this, since it becomes conscious, too. So Freud argues

(c) the real stimulus to the act is the order of the physician therefore,

(d) the idea of the physician's order became active, too.

Unless (d) can be proved, no important sense in which ideas are unconscious has been shown. Does (c) imply (d)? To see why this is doubtful, we need to examine the peculiar meanings of the expressions 'real stimulus' and 'becoming active', when applied to ideas. The former, it will be argued, represents Freud's commitment in this proof to an unacceptable theory of causality, especially mental causality, while the latter expression is ambiguous.

The premise that might at first seem to justify the inference from (c) to (d) becomes clear when we ask what more must be said than

(e) the real stimulus of an act must be active at the same time as the act (or contiguous with it), or must cause something else which is active at the same time as the act (or contiguous with it).