IV.—DISCUSSIONS

ON A SO-CALLED PARADOX

A PUZZLE that has had some currency from 1943 onward is concerned with a man who was sentenced on Sunday to be hanged on one of the following seven noons, and to be kept in ignorance, until the morning of the fatal day, as to just which noon it would be. By a faulty argument the man persuaded himself that the sentence could not be executed, only to discover his error upon the arrival of the hangman at 11.55 the following Thursday morning. What his faulty argument was is almost too familiar now to bear recounting (though I shall recount it); for the puzzle has kept recurring in the oral tradition and it has broken into MIND [1] in two variant versions, one relating to a surprise air-raid drill and the other to a surprise hour examination. The puzzle in each case is to find the fallacy. What is remarkable is that the solution, a solution which at any rate has contented me for nine years, seems seldom to have been clearly apprehended. There is a false notion abroad that actual paradox is involved. This notion has even brought Professor Weiss [1] to the desperate extremity of entertaining Aristotle's fantasy that 'It is true that p or q' is an insufficient condition for 'It is true that p or it is true that q.'

The plot, in each of its embodiments, is as follows: K knows at time t and thereafter that it is decreed that an event of a given kind will occur uniquely and within K's ken at time $t + i$ for some integer i less than or equal to a specified number n, and that it is decreed further that K will not know the value of 'i' until after (say) time $t + i - \frac{1}{2}$. Then K argues that $i \leqq n - 1$; for, if i were n, K would know promptly after $t + n - 1$ that i was n. Then, by the same reasoning with '$n - 1$' for 'n' he argues that $i \leqq n - 2$; and so on, finally concluding after n steps that $i \leqq 0$ and hence that the event will not occur at all.

It is notable that K acquiesces in the conclusion (wrong, according to the fable of the Thursday hanging) that the decree will not be fulfilled. If this is a conclusion which he is prepared to accept (though wrongly) in the end as a certainty, it is an alternative which he should have been prepared to take into consideration from the beginning as a possibility.

Thus K erred in his argument that $i \leqq n - 1$. Looking ahead at time t to the possible states of affairs at time $t + n - 1$, K discerned just two alternatives as follows: (a) the event will

[1] D. J. O'Connor, 1948, p. 358; L. J. Cohen, 1950, p. 86; Peter Alexander, p. 538; Michael Scriven, 1951, pp. 403 ff; Paul Weiss, 1952, pp. 265 ff.

have occurred at or before that time; (*b*) the event will (in keeping with the decree) occur at time $t+n$, and K will (in violation of the decree) be aware promptly after $t+n-1$ that the event will occur at time $t+n$. Rejecting (*b*) because of its violation of the decree, he elected (*a*). Actually K should have discerned not two alternatives but four, *viz.* (*a*) and (*b*) and two more as follows: (*c*) the event will (in violation of the decree) fail to occur at time $t+n$; (*d*) the event will (in keeping with the decree) occur at time $t+n$, and K will (in keeping with the decree) remain ignorant meanwhile of that eventuality (not knowing whether the decree will be fulfilled or not). He erred in not recognising that either (*a*) or (*d*) could be true even compatibly with the decree. The same fault recurred in each of his succeeding $n-1$ steps.

The tendency to be deceived by the puzzle is perhaps traceable to a wrong association of K's argument with *reductio ad absurdum*. It is perhaps supposed that K is quite properly assuming fulfilment of the decree, for the space of his argument, in order to prove that the decree will not be fulfilled. This, if it were all, would be good *reductio ad absurdum*; and it would entitle K to eliminate (*b*) and (*c*), but not (*d*). To suppose that the assumption of fulfilment of the decree eliminates (*d*) is to confuse two things; (i) a hypothesis, by K at t, that the decree will be fulfilled, and (ii) a hypothesis, by K at t, that K will know at $t+n-1$ that the decree will be fulfilled. Actually hypothesis (i), even as a hypothesis made by K, admits of two sub-cases: K's hypothetical ignorance and K's hypothetical awareness of the hypothetical fact.

Thus suppose that a mathematician at work on the Fermat problem assumes temporarily, for the sake of exploring the consequences, that Fermat's proposition is true. He is not thereby assuming, even as a hypothesis for the sake of argument, that he knows Fermat's proposition to be true. The difference can be sensed by reflecting that the latter would actually be a contrary-to-fact hypothesis, whereas the former may or may not be.

K's fallacy may be brought into sharper relief by taking n as 1 and restoring the hanging motif. The judge tells K on Sunday afternoon that he, K, will be hanged the following noon and will remain ignorant of the fact till the intervening morning. It would be like K to protest at this point that the judge was contradicting himself. And it would be like the hangman to intrude upon K's complacency at 11.55 next morning, thus showing that the judge had said nothing more self-contradictory than the simple truth. If K had reasoned correctly, Sunday afternoon, he would have reasoned as follows. "We must distinguish four cases: first, that I shall be hanged tomorrow noon and I know it now (but I do not); second, that I shall be unhanged tomorrow noon and know it now (but I do not); third, that I shall be unhanged tomorrow noon and do not know it now; and fourth, that I shall be hanged tomorrow noon and do not know it now. The latter two alternatives are the open possibilities, and the last of all would fulfil the decree. Rather than charging the judge with self-contradiction, therefore, let me suspend judgment and hope for the best."

W. V. QUINE

Harvard University