EPISTEMIC OPACITY

Eddy M. ZEMACH

The problem of the alleged opacity of contexts governed by epistemic operators can be formulated as follows. Suppose that

(1) Abe repeatedly says that his next door neighbor is a decent man.

If we have every reason to believe that when Abe says that his neighbor is a decent man he intends to convey to us his sincere beliefs, and we have no reason to believe that Abe intends to fool us or cover up his real beliefs, then we are justified in concluding that

*(2) Abe believes that his neighbor is a decent man.

I.e., we are justified in asserting that

(2) $B[a, H(ixFx)]$.

Again suppose that

(3) Abe repeatedly says that the Major is not a decent man.

If again we have every reason to believe that Abe is expressing his real opinions, we may conclude that

*(4) Abe believes that the mayor is not a decent man.

I.e., that

(4) $B[a, \neg H(ixGx)]$.

If Abe believes that the major is not decent, he does not believe that the mayor is decent. Hence (4) implies

(5) $\neg B[a,H(ixGx)]$.

However, suppose that, unbeknownst to Abe, his next door neigbor is the town's mayor, i.e.,

(6) ixFx = ixGx.

By the principle of the indiscernability of identicals, (5) implies

(7) $\neg$B[a, H(ixFx)]

which is the denial of (2). But if we are justified in asserting (2) and (7) we are justified in asserting

(8) B[a, $\neg$H(ixFx)] & $\neg$B[a, H(ixFx)]

i.e., we are justified in asserting a logical contradiction. This is absurd; one can never be justified in asserting that p & $\neg$p.

In order to solve this problem philosophers have been willing to pay most staggering prices. Some, like Frege, said that in belief contexts 'ixFx' does not refer to ixFx. Quine has argued that belief contexts are referentially opaque, hence they cannot be quantified into, and the substitutivity of identicals does not there hold. Chisholm, too, gives up the substitutivity of equireferential terms in belief contexts and, moreover, argues that 'B[a, H(ixFx)]' has two senses, such that only in one of them (*in sensu diviso*) does 'ixFx' refer to ixFx, while in the other sense (*in sensu composito*) it does not. Swartz, who offers another version of the Chisholmian solution, also distinguishes between referential terms "in referential position" and "not in referential position." Others went so far as to reject the principle of the indiscernability of identicals *in toto*. (1)

In view of all the above, I believe I can offer a solution to the puzzle at a bargain price. I shall keep the *strongest* version of Leibniz principle — substitutability *salva veritate* of equireferential terms in *all* contexts, deny that there are any "opaque" contexts, that there are any "composite" senses or

(1) A fairly complete bibliography of the literature on the subject is to be found in Robert J. SWARTZ, "Leibniz Law and Belief", *Journal of Philosophy* LXVII (1970): 122-137.

any "non referential position" for referential terms to stand in, and yet avoid the contradiction (while retaining the intuitive advantages gained by imposing the above restrictions on quantification and substitution).

(5), I claim, is strictly false. We must remember that (5) is not asserted by Abe. It is asserted by us, and the term 'ixGx' used in it is a referential term in *our* language referring to whatever is *in fact* ixGx, that is, by (6), to ixFx. In using the term 'ixGx' we do not imply that Abe himself understands this term or that he would be ready to use it in order to refer to ixGx. E.g., Abe may be ignorant of English, and hence would never use the term 'ixGx' in referring to ixGx. Using the term 'ixGx' is just *our* way of referring to this entity. Hence it is just false to say that Abe does not believe concerning the entity which we, in fact, denote by using the referential term 'ixGx', that it is H. On the contrary, we know, by (2), that Abe does believe that ixGx, i.e., the entity which is in fact denoted in our language by 'ixFx' is H. If Abe blesses his neighbor, and his neighbor is the town's mayor, then Abe has blessed the mayor (i.e., the entity which is in fact referred to in our language as 'the mayor'). Similarly if Abe believes that his neighbor is honest, and his neighbor is the town's mayor, then Abe believes that the mayor (i.e., the entity which is in fact referred to in our language as 'the mayor') is honest.

But, if (5) is false, how can it follow from (4), which is true ? My answer is that indeed it does not follow from (4) at all. (4) states that Abe holds a certain belief, i.e., the belief that ixGx is not H. This, however, does not say a single thing about what *other* beliefs Abe may or may not hold; especially, it does not say a thing about Abe's holding, or not holding, the belief that ixGx is H. We may agree that if someone holds the belief that Fx and also the belief that not Fx, that person is in trouble. But *we* are not in trouble for saying that this, in fact, is the balance of his beliefs. The conjunction of (2) and (4) is not self-contradictory at all.

But how can one hold both beliefs, i.e., that Fx and that not Fx at the same time ? Well, one explanation may be that one is just very irrational and makes no attempt to formulate a consistent set of beliefs, such that he holds both p and q to be true although q is p's denial. This may be easier to imagine in case q is not *patently* a denial of p; it only implies not-p, or amounts to not-p. Similarly one may hold that Fx and $\neg$Gx although the property G is (or implies) the property F. Finally, one may hold that Fx and $\neg$Fy although $x = y$. Holding contradictory beliefs may be a psychological, or practical, problem, but it is not a logical one. If one does *not* (yet ?) know that his beliefs contradict each other, his problem is mainly practical; if he knowingly entertains those contradictory beliefs, the man is surely irrational. But the logician is not practically troubled nor irrational if he states that this is the sorry predicament of someone else.

The plausibility of arguing that

(9) $B(a, \neg p) \supset \neg B(a, p)$

is false may be enhanced by pointing out the similarity between the formal features of Epistemic and Modal logic, especially the similarity between 'Know' and 'Necessary' on the one hand side, and 'Believe' and 'Possible' on the other hand side. Thus while the truth of

(10) $(\Box \neg p) \supset (\neg \Box p)$

lends support to our considering

(11) $K(x, \neg p) \supset \neg K(x, p)$

as true, the falsity of

(12) $(\Diamond \neg p) \supset (\neg \Diamond p)$

should incline us to hold that (8) is false too. Both '$\Box \neg p$' and '$K(x, \neg p)$' imply '$\neg p$', and hence they cannot be asserted if '$\Box p$' or '$K(x, p)$', which imply 'p', are also asserted. But '$\Diamond \neg p$' and '$B(x, \neg p)$' do not imply ($\neg p$), and hence they can be conjoined with '$\Diamond p$' and 'B(x, p)' respectively, with no contradiction involved.

Can we not, however, be justified in asserting (5) independently of (4) ? Suppose, e.g., that

(13) Abe repeatedly says that he does not believe that the mayor is a decent man.

Would we not then be justified in asserting (5) ? The answer is, again, No. Even if we suppose Abe is absolutely honest, the only conclusion we may derive from (12) is not (5) but rather

(14) $B\{a, \neg B[a, H(ixFx)]\}$

i.e.,

*(14) Abe believes that he does not believe that the mayor is a decent man.

One cannot derive (5) from (13) in the same way that one cannot derive

(15) The major is not a decent man

from (3). Abe may believe p (e.g., that blacks are inferior) although he does not believe he believes p. But if all we may derive from (13) is (14), then surely no contradiction does ensue. (2) and (14) do not contradict each other.

Let me now turn to examine Quine's proof that some epistemic contexts are referentially opaque (*Word and Object*, pp. 148-9) (2). Quine defines a number, dp, as
"the number x such that $(x = 1)$ & p or $(x = x)$ & $\neg p$"
and proceeds to say: "We may suppose that poor Tom ... is enough of a logician to believe a sentence of the form 'dp = 1'

(2) An earlier version of this examination was included in my «Reference and Belief", *Analysis* 30: 11-15 (1969).

when and only when he believes the sentence represented by 'p'."

I shall stop here since this passage already contains the mistake which vitiates Quine's proof. Indeed we may suppose that if Tom is a good logician, if he is familiar with the above definition of dp, and believes that p, then he believes that dp equals 1. I.e.,

(16) $B(t, p) \supset B(t, dp = 1)$.

But the converse is *not true*. We have no right to say that

(17) $B(t, dp = 1) \supset B(t, p)$

i.e., that if Tom believes that dp is equal to one he believes that p. 'dp' is here *used* by us to refer to a certain number concerning which Tom believes it is equal to one, and we have no right to suppose that Tom refers to this number in the same way we do. E.g., since

(18) dp = the number mentioned by Quine on p. 148 l. 3 up of *Word and Object*

is true, Tom may believe for some strange reason that the number mentioned by Quine on p. 148 l. 3 up of *Word and Object* is equal to one. This fact may be expressed by us (using another term to denote that same number) by saying (truely):

(19) $B(t, dp = 1)$.

But surely this does not imply anything about Tom's believing, or not believing, the sentence represented by 'p'.

Now Quine's proof runs as follows:

(20) Tom believes that Cicero denounced Catiline.

(21) Tom believes that d(Cicero denounced Catiline) = 1.

Let 'p' represent any true sentence. Then

(22) dp = d (Cicero denounced Catiline)

and, by substitution,

(23) Tom believes that dp = 1.

So far the proof is flawless. But from (23) Quine derives, by assuming the illegitimate and invalid (17),

(24) Tom believes that p

i.e., Tom believes any true sentence. But since (17) is invalid, (24) is invalid too, and the proof of the existence of opaque contexts collapses.

Finally, I would like to suggest a way to express that which Quine, Chisholm, Hintikka, and most other epistemic logicians mistakenly referred to as the non-referential, or opaque, use of referential terms. As a first approximation, we may write, instead of (2),

(25) B[a, T 'H(ixFx)']

This, as it stands, will not do, since it may be true that Abe believes that his neighbor is a decent man, but (since he does not know any English) he does not believe that the sentence 'Abe's neighbor is a decent man' is true. However, all we need in order to rectify this is to replace the regular quotes here by something like Sellars' dot quotes.

(26) B a, T • H(ixFx) •

says that Abe believes that a sentence (in his language), whose translation into the language of the one who asserts (26) is the sentence symbolized by 'H(ixFx)', is true. E.g., if the language of the one who asserts (26) is English, then when he asserts (26) one says that Abe believes that a sentence whose

English translation is 'Abe's neighbor is a decent man' is true. *Inside* the quotation marks there are no individual terms, referential or otherwise, and hence (obviously) "they" cannot be quantified over or substituted *salva veritate* by equi-referential ones.

The Hebrew University of Jerusalem Eddy M. Zemach