

HOWARD DARMSTADTER

INDETERMINACY OF TRANSLATION AND
INDETERMINACY OF BELIEF*

(Received 8 August, 1973)

In a number of writings over the last decade, W. V. Quine has advanced the thesis of the indeterminacy of radical translation.¹ Briefly, the thesis states that for any two natural languages there will exist several equally legitimate systems for translating sentences of the one language into sentences of the other. These differing systems will, for certain sentences, offer translations which differ radically among themselves. In particular, the translations may differ in truth value.

Quine's indeterminacy thesis raises some awkward consequences for our common sense notions regarding belief. Suppose a speaker sincerely assents to some sentence *S*, and suppose that two different but equally valid systems of translation offer sentences *A* and *B* of a second language as translations of *S*, where *A* and *B* differ in truth value. We would naturally think that if the speaker of the first language believes that *S*, then he also believes that *A*, if *A* is a correct translation of *S*. If *A* is a correct translation of *S*, then we feel that the belief that *A* and the belief that *S* are the *same* belief. But if Quine is right about the indeterminacy thesis – if the different systems of translations are equally legitimate – then the belief that *A* is the same as the belief that *B*, assuming (plausibly) the transitivity of the '... is the same belief as ---' relation. But *A* and *B* differ in truth value, so we seem forced to say that if Quine is right, two beliefs with different truth values can still be identical.

Far from regarding this result as a *reductio* of the indeterminacy thesis, Quine believes that it shows our common sense notions about beliefs to be fundamentally mistaken.

Where physical theories *A* and *B* are both compatible with all possible data, we might adopt *A* for ourselves and still remain free to translate the foreigner either as believing *A* or as believing *B*... The question whether... the foreigner *really* believes *A* or believes rather *B*, is a question whose very significance I would put in doubt. This is what I am getting at in arguing the indeterminacy of translation.²

In this paper, I shall try to offer a rigorous statement of Quine's thesis, and also indicate why I think that thesis to be unfounded.

I. INDETERMINACY OF BELIEF: THE QUINIAN ARGUMENT

I shall assume that the belief that S , where S is an English indicative sentence, is a certain kind of brain state or brain *configuration* or, which is the same, a certain set of brain configurations. A mapping B which associates each English indicative sentence S with the set of brain configurations $B(S)$ which is the belief that S will be called a *belief mapping* for English. I shall also assume that being an English speaker involves having a certain kind of brain configuration or, which is the same thing, having a brain configuration which is an element of a certain set of brain configurations, which set I shall call an *assent mechanism* for English. A person's assent mechanism determines the person's pattern of assent to the sentences of his language; standing before the window, my assent mechanism dictates my assent to the sentence 'You are standing before the window' but not to the sentence 'You are seated at the desk'. Should I move from window to desk, my assent mechanism will dictate the appropriate change in my pattern of assent.

Many conditions can be suggested for belief mappings, but in the following arguments, I shall only be concerned with one, which I shall call the *assent condition*: A belief mapping B for a language L will be said to satisfy the assent condition for L if and only if there exists an assent mechanism A for L such that for any speaker X of L in A , and any indicative sentence S of L , then generally X will assent to S if and only if X 's brain configuration is an element of $B(S)$. Where a belief mapping B and an assent mechanism A are related in this way, I shall say that B *describes* the pattern of assent of speakers of L in A ; in those cases where every speaker of L is in A , I shall simply say that B describes the pattern of assent of speakers of L .

As yet nothing has been said to indicate that there can not be radically different belief mappings for the same language. Imagine, for example, the following situation: Assume a language L , and two belief mappings, B_1 and B_2 , for L . By the assent condition, B_1 must describe the pattern of assent of speakers of L in some assent mechanism A_1 , and B_2 must describe the pattern of assent of speakers of L in some assent mechanism

A_2 . Let X be a speaker of L in A_1 , and Y a speaker of L in A_2 . Lastly, assume that for some sentence S of L , X is in $B_1(S)$ but not in $B_2(S)$, while Y is in $B_2(S)$ but not in $B_1(S)$. Thus, X and Y both assent to S .

The above case (henceforth, the *Quinian case*) has some interesting features. For one, it is somewhat perplexing to say just what it is that X and Y believe. Both X and Y assent to S , but if we use either B_1 or B_2 as the belief mapping for the language, then either X or Y can be said to not really believe what he says he believes. But since there seems no non-ad hoc way of deciding between B_1 and B_2 , it is not at all obvious just which of our speakers is not aware of his own beliefs. Each speaker can readily 'attribute unimagined views'³ to the other. Indeed it seems that the question of what our speakers really believe is a question whose very significance might readily be put into doubt.⁴ The case constructed thus seems to do justice to many of the things Quine says about the indeterminacy thesis.

It should be obvious that in the Quinian case, A_1 cannot be the same as A_2 . For if B_1 describes the pattern of assent of all speakers of L in A_1 , and $A_1 = A_2$, then B_1 would also have to describe the pattern of assent of all speakers of L in A_2 . But Y is a speaker of L in A_2 , and therefore B_1 would have to describe Y 's pattern of assent. This would mean that, since Y assents to S , Y must be in $B_1(S)$. But this is what the Quinian case explicitly denied.

The different assent mechanisms used by X and Y might be said to involve different understandings of the language L . That is, we might wish to say that despite outward appearances, speakers of L were actually divided into (at least) two different language communities, corresponding to the assent mechanisms A_1 and A_2 . A_1 and A_2 can be thought of as containing a speaker's understanding of L , and as such they would contain different understandings. Quine would not only argue that this could happen, but that it could happen in such a way that speakers of L would be unaware of any linguistic cleavage. Speakers using A_1 would think that speakers using A_2 held different beliefs, not that they had a different understanding of the language. There would be no more hindrance to communication between speakers using A_1 and speakers using A_2 than could be convincingly accounted for by their holding different beliefs. In such a situation, no behavioral sense could be made of a distinction between persons using A_1 and those using A_2 meaning something different

by various sentences of *L*, and their merely holding different beliefs.⁵

II. INDETERMINACY OF BELIEF: THE ARGUMENT COUNTERED

In this section, I will try to show that the Quinian case is unlikely to occur. My argument will depend on showing that in the Quinian case, the two speakers will suffer from certain failures to predict each other's actions – failures of prediction that would cease if they should both come to use the same assent mechanism. While it remains possible that the two speakers could simply continue to suffer from such failures of prediction, it seems that this would be an unstable position; people would, one feels, gradually work their way out of the Quinian case so as to be able to produce more accurate predictions of behavior.

Consider speaker *X* of language *L*, who in the Quinian case uses the assent mechanism A_1 . Among the abilities we may assume *X* to possess is the ability to make moderately accurate predictions of the actions of other speakers of *L*. It seems plausible that *X* can make predictions of the actions of a speaker *Z* of *L* by assessing *Z*'s beliefs and goals. *Z*'s beliefs and goals can be assessed in large part by observing *Z*'s assents to various sentences of *L*, supplemented by various other clues as to where *Z*'s assents might be forthcoming. Once the requisite data is gathered about the pattern of *Z*'s assent, *X* can make moderately accurate predictions of *Z*'s actions by imagining what he (*X*) would do were he (*X*) to have the same pattern of assent as now characterizes *Z*. *X*'s ability to make projections of his own actions given various hypothetical patterns of assent is a long way from being foolproof – people often misjudge their own responses, and in some areas we may expect such misjudgments to be the rule – but I think it undeniable that people do make such projections, and that they are moderately accurate in making them.

For *X* to make projections of his own actions under various hypothetical patterns of assent, *X* must utilize some sort of psychological projection mechanism. I shall refer to this mechanism as an *empathizing* mechanism, in tribute to the role it plays in helping *X* to understand others by mirroring them in his own mind. Without speculating too much as to its form, it seems obvious that if this empathizing mechanism is to allow *X* to make accurate predictions, its output must be largely determined by the output of *X*'s assent mechanism. *X*'s assent mechanism dictates *X*'s

pattern of assent under a variety of types of brain configurations; X 's empathizing mechanism works (roughly) in reverse, projecting X 's actions (which are determined by X 's brain configuration) from hypothetical patterns of assent. If X 's assent mechanism dictates a certain pattern of assent P in a brain configuration of type C , then it would seem that X 's empathizing mechanism ought to project from the hypothetical pattern of assent P those actions which would be taken in brain configurations of type C .

In the Quinian case, X uses assent mechanism A_1 , and this mechanism determines an empathizing mechanism which X uses to make predictions regarding the actions of speakers of L . I shall condense this by simply saying that X uses A_1 to predict the actions of speakers of L . Similarly, Y uses A_2 to predict the actions of speakers of L . The question arises: Will X be able to make as accurate predictions of Y 's actions, and of the actions of other speakers of L who use assent mechanism A_2 , as he can of the actions of those speakers of L who like himself use assent mechanism A_1 ? After all, when X predicts on the basis of another speaker's pattern of assent P that that speaker will perform a certain type of action α , it is presumably because X 's assent mechanism would cause X to have a pattern of assent P when in a brain configuration of type C , and people in brain configurations of type C generally perform actions of type α . But when Y has a pattern of assent P , it may be that Y 's (different) assent mechanism is such that it causes Y to have a pattern of assent P when his brain configuration is of type C^* ; and it may be that persons with brain configurations of type C^* do not generally perform actions of type α .

Y 's assent mechanism is different enough from X 's that for some sentence S of L , X would assent to S when in $B_1(S)$, while Y would assent to S when in the (different) set $B_2(S)$. But it might be argued that this in itself need not cause X to make any sort of systematic mistake in predicting Y 's actions; for it might be that X and Y are still in very similar brain configurations when they assent to the same sentences. For example, while $B_1(S) \neq B_2(S)$, it might still happen that $B_1(S) \cap B_1(S_1) \cap B_1(S_2) \cap \dots \cap B_1(S_n) = B_2(S) \cap B_2(S_1) \cap B_2(S_2) \cap \dots \cap B_2(S_n)$, for sentences S, S_1, S_2, \dots, S_n of L . (The two complex sets of the preceding equation will henceforth be abbreviated ' $B_1(S, S_1, S_2, \dots, S_n)$ ' and ' $B_2(S, S_1, S_2, \dots, S_n)$ ' respectively.) In such a case, if Y would assent to all of S, S_1, S_2, \dots, S_n , then one could explain to X that any failure to accurately predict

Y 's actions was due to a failure to realize that Y believed all of $S, S_1, S_2, \dots, \dots, S_n$. Had X predicted Y 's actions based on Y 's assent to all of S, S_1, S_2, \dots, S_n , rather than on Y 's assent to S alone, he would have predicted Y 's actions based on Y 's being in $B_1(S, S_1, S_2, \dots, S_n)$; the fact that Y uses A_2 while X uses A_1 could not lead X to make any mistake, for $B_2(S, S_1, S_2, \dots, S_n)$ would be the same set as $B_1(S, S_1, S_2, \dots, S_n)$. X might misjudge Y 's actions when X bases his judgments on Y 's pattern of assent to some proper subset of S, S_1, S_2, \dots, S_n , but X would always be able to trace his misjudgment to lack of complete information about the full pattern of Y 's assents. Such misjudgments could thus always be attributed to differences in Y 's beliefs rather than to differences in Y 's understanding of L .

Will filling in more information about Y 's assents always enable X to make fairly accurate predictions about Y 's course of action under various circumstances, even though X uses A_1 rather than A_2 ? For such to be always the case, it must be that whatever the complete pattern of Y 's assents, the intersection of the sets under B_1 must be the same as the intersection under B_2 . That is, for any S_1, S_2, \dots, S_n , if Y would assent to all and only S_1, S_2, \dots, S_n , then $B_1(S_1, S_2, \dots, S_n) = B_2(S_1, S_2, \dots, S_n)$. But this is impossible. For let b be a brain configuration in A_2 , and S_1 a sentence such that $b \in B_2(S_1)$ but $b \notin B_1(S_1)$.⁶ Now let S_1, S_2, \dots, S_n be the set of sentences such that $b \in B_2(S_i)$, for $1 \leq i \leq n$. Then $b \in B_2(S_1, S_2, \dots, \dots, S_n)$, but $b \notin B_1(S_1, S_2, \dots, S_n)$, since $b \notin B_1(S_1)$. Therefore, $B_1(S_1, S_2, \dots, S_n) \neq B_2(S_1, S_2, \dots, S_n)$.

It would seem that the use of different assent mechanisms by different speakers of the language would result in some failure on the part of the speakers to accurately predict each other's actions. To eliminate such failures, one may reasonably expect speakers to come to adopt the same assent mechanism. Such a conclusion can not be said to have been definitely established – the arguments are far too weak for that. Nonetheless, I do think it fair to say that the conclusion has been made somewhat probable. What now remains is to extend these results from the one-language to the two-language case.⁷

III. INDETERMINACY OF TRANSLATION

In the preceding section I argued that all the speakers of a language L

would come to use the same assent mechanism A . But this conclusion does not imply that there must exist a unique belief mapping B which describes the pattern of assent of every speaker of L ; it may happen that there are two belief mappings for language L , B_1 and B_2 , both of which describe the pattern of assent of every speaker of L but are such that for some sentence S of L , $B_1(S) \neq B_2(S)$. The argument of the preceding section only implies that any brain configuration which is an element of $B_1(S)$ but not of $B_2(S)$, or vice-versa, must be a brain configuration which is not an element of A , i.e., a brain configuration which is not that of a speaker of L . This allows us to formulate another version of Quine's thesis of the indeterminacy of radical translation.

Suppose there are two belief mappings for a language L , B_1 and B_2 , which describe the pattern of assent of all speakers of L , where A_L is the assent mechanism used by all speakers of L . Further suppose that there also exist two belief mappings for a different language N , B_3 and B_4 , which describe the pattern of assent of all speakers of N , where A_N is the assent mechanism used by all speakers of N . Then it may happen that for some sentence S^* of N , $B_3(S^*) = B_1(S_1)$ while $B_4(S^*) = B_2(S_2)$, where S_1 and S_2 are radically different sentences of L .⁸ If we attempt to translate sentence S^* of N by finding a sentence S of L such that the belief that S^* is identical with the belief that S , then we can translate S^* as either S_1 or S_2 . In such a case, there will be two radically different systems for translating sentences of N into sentences of L .

While the above type of indeterminacy may be possible, it does not seem to represent a stable position. As long as speakers of L remain relatively isolated from speakers of N , it may be that one can use either system of translation indifferently. But once there exists a sizable body of bilinguals who must learn to predict the actions of each community and of other bilinguals, then there will be pressures on the bilingual community to adopt a common system of translation. To see this, imagine two bilinguals, X and Y , and imagine that X has internalized the system of translation which correlates a sentence S_L of L with a sentence S_N of N if and only if $B_1(S_L) = B_3(S_N)$, while Y has internalized the system of translation which correlates a sentence S_L with S_N if and only if $B_2(S_L) = B_4(S_N)$. X and Y would assent to various sentences of N if and only if they would assent to their translations in L according to their own internalized systems of translation; thus, X would assent to S^* if and only

if he would assent to S_1 , while Y would assent to S^* if and only if he would assent to S_2 . Now suppose that X is trying to predict Y 's actions, where X knows that Y would assent to S^* . X will also assume that since Y would assent to S^* , Y would assent to S_1 as well. But this may very well be false, in which case X will credit Y with the wrong pattern of assent to sentences of L . Once this happens, it seems likely that X will be apt to make some sort of systematic error in predicting Y 's actions.

It should be noted that X will make no systematic mistake in predicting Y 's actions in those cases where Y is not a bilingual: If Y is a speaker of N but not of L , then X will simply use A_N to predict Y 's actions. But suppose Y is a bilingual, and X tries to simultaneously use Y 's assents to sentences of L and to sentences of N to predict Y 's behavior. X can be thought of as using either A_N or A_L to predict Y 's actions. If he uses A_L , he will use Y 's assents to sentences of L , but he will also want to use translations of Y 's assents to sentences of N , and here the trouble will begin. If X is careful, and only tries to predict Y 's actions by using Y 's assents to sentences of one language at a time, there will be no systematic error. But to make predictions this way is to treat Y as two different unilingual persons; X would not be able to make use of all the potentialities for increased accuracy that might be gained by combining Y 's assents to the sentences of the different languages. This potentiality could be realized, and no systematic mistake would be made, if X and Y used the same internalized system of translation. So it would seem that a community of bilinguals could greatly increase their potential for accurately predicting each other's actions by coming to use the same internalized system of translation.

X and Y could come to agree in their internalized systems of translation by both adopting X 's system or by both adopting Y 's system (there may be other systems as well). Prior to their adopting a common system, neither S_1 nor S_2 will be the 'real' translation of S^* , and we may remain free to speculate (if such is our wont) whether S^* really means S_1 or S_2 . If the situation described in this section could actually occur, then part of what Quine is arguing for in stating the thesis of indeterminacy of radical translation is correct. But Quine also seems to argue that indeterminacy will remain a permanent possibility even after we have come to agree on a single system of translation. Thus, he seems to feel that we could independently adopt a system of translation different from the one

in general use. Under this different system, people may seem to hold different beliefs, but nothing will happen, according to Quine, to upset our confidence in our translations.⁹ But of course something will happen; if we attempt to use our new system of translation with bilinguals, we will be unable to meld together their assents to sentences of different languages in a way which will enable us to produce moderately accurate predictions of their future courses of action. Quine seems to assume that because we may be free to translate a newly discovered language in any of various ways, we will still retain that freedom even after a sizable body of bilinguals has come into existence. *Before* we have settled on a system of translation, the question whether the foreigner *really* believes S_1 or believes rather S_2 is a question whose significance can be put in doubt. But once we have a system of translation, the question becomes a significant one: If we accept one answer rather than another, we may end up with significant difficulties in attempting to predict the actions of bilinguals.

University of Massachusetts, Boston

NOTES

* I am indebted to Martin Andic, Howard Cohen, Robert Shope, and William Walters for their comments and criticisms. A special debt is owed to Gilbert Harman, who gave generously of his time despite various misgivings about my conclusions. His criticisms somewhat delayed and greatly improved this paper.

¹ The thesis finds its fullest exposition in Chapter II, 'Translation and Meaning', of Quine's *Word and Object*, MIT Press, Cambridge, and Wiley, New York, (1960), pp. 26-79. Further clarification is offered in W. V. Quine, 'On the Reasons for Indeterminacy of Translation', *The Journal of Philosophy*, LXVII (March 26, 1970), 178-183.

² 'On the Reasons for Indeterminacy of Translation', pp. 180-181, Quine's italics.

³ *Word and Object*, p. 78.

⁴ 'On the Reasons for Indeterminacy of Translation', pp. 180-181.

⁵ In this connection see especially *Word and Object*, p. 77.

⁶ Given the conditions of the Quinian case, there must be a brain configuration b and a sentence S of L such that either $b \in B_1(S)$ and $b \notin B_2(S)$, or $b \notin B_1(S)$ and $b \in B_2(S)$. We may without loss of generality assume the latter, letting S_1 be the sentence in question.

⁷ As was argued in my doctoral dissertation 'W. V. Quine on Translation', Princeton University (1967), the one-language case is the crucial case for Quine, since the two-language case can be fulfilled in completely trivial ways. See Chapter I, esp. pp. 27-29 and 52-57.

⁸ 'May happen' in the sense that I can see no reason why it couldn't happen. But further progress in the sciences of psychology and linguistics might provide empirical reasons for such a situation's being impossible.

⁹ *Word and Object*, p. 78.