Two Thought Experiments Reviewed

TYLER BURGE

The issues raised by the anti-individualist arguments in Hilary Putnam’s “The meaning of ‘meaning’” and my “Individualism and the mental” are various and complicated. I shall be able to touch on only a few of these issues in this space. What makes matters more complicated is that Putnam’s interests and viewpoint differ from mine. Although a comparison of our views would be useful, I will have to concentrate entirely on those of Fodor’s remarks that concern my work. I begin by setting out two thought experiments, which I formulate in ways that raise a minimum of issues that are extraneous to my primary interests. Then I will state what I take the thought experiments to show. Finally, I will discuss Fodor’s criticisms.

The first thought experiment is a variant on Putnam’s original Twin Earth case. For our purposes, the differences are significant. Assume that Al has a variety of beliefs and occurrent thoughts involving the notion of aluminum. He thinks that aluminum is a light, flexible metal, that many sailboat masts are made out of aluminum, that someone across the street recently bought an aluminum canoe. These occurrent and state-like attitudes are correctly (truly) described with that-clauses containing ‘aluminum’ in oblique, not purely transparent position. That is, exchanges of expressions that are coextensive with ‘aluminum’ (apply to exactly the same quantities of aluminum) do not in general preserve the truth-value of these original attitude ascriptions.

Now, as a second step in the thought experiment, conceive of a physical duplicate of Al who lives on a fraternal twin of Earth. Call him Al'. Al' is bodily identical with Al. He undergoes the same stimulations on his bodily surfaces, excepting minor micro- and gravitational differences, engages in the same motions, utters the same sounds, has the same experiences—insofar as these stimulations, motions, and experiences are nonintentionally described. This physical and phenomenal similarity—virtual type-identity—is preserved from birth to the present. I will suppose also that there are internal-causal, functional, and syntactical similarities as long as they are specified nonintentionally.

Received January 25, 1982
and defined on the individual in isolation from his social and physical environment. Twin Earth looks very much like Earth to the naked eye. But on Twin Earth there is no aluminum. There is, however, a metal, not occurring on Earth that looks like aluminum and is put to many of the uses aluminum is put to here. The sign ‘aluminum’ is assigned to this stuff on Twin Earth (we shall use the sign ‘aluminum\text{\_2}’ for it). And wherever Al encounters aluminum, Al\_i encounters this other stuff.

To sharpen certain contrasts, I want to introduce some further differences between Earth and Twin Earth. Laymen like Al and Al\_i know of no features that could in principle be used to differentiate between aluminum and its look-alike, if the issue were ever to arise (it doesn’t). But there are substantial differences in the actions and attitudes in their respective communities toward the respective metals. Scientists here know aluminum to be an element; scientists on Twin Earth know the complicated formula, ZYX, of the look-alike. Differences in physical properties, including subtle macro-differences are reflected in the attitudes of the respective scientific, technological, and business communities. Differences also occur in some industrial uses. One can ramify the differences through the community as much as one wants as long as they do not produce significantly different physical impacts on Al and Al\_i’s bodily surfaces. Al\_i remains surface-identical and movement-identical with Al.

The thought experiment concludes with an observation. Al\_i does not think that aluminum is a light, flexible metal; he does not think that sailboat masts are made of aluminum; and so forth. He lacks attitudes that can be correctly (truly) described with ‘aluminum’ in oblique position. I take this to be obvious. Al\_i has never had contact with aluminum, nor contact with anyone who had contact with aluminum . . . ; and no one in his community uses a word that means what the word ‘aluminum’ means in English. Under these circumstances, we do not attribute thoughts like those we attributed to Al. We attribute different thoughts. So despite their physical similarities, Al and Al\_i have different propositional attitudes—differences that effect obliquely occurring terms in propositional attitude ascriptions.

We turn now to a second thought experiment. Assume that Bert has arthritis, knows it, and has a variety of occurrent and state-like attitudes truly describable with ‘arthritis’ in oblique position. He thinks his Aunt Mary’s arthritis was more severe than his own, that arthritis has crippled many people, and so on. At a certain time t, he mistakenly thinks that he has got arthritis in the thigh as well as in the fingerjoints and knee. Later he is apprised of the fact that arthritis is an inflammation of joints and cannot occur in the thigh.

Next conceive of a counterfactual situation in which Bert is identical to his actual self up through time t in all the ways we took Al and Al\_i to be. In the counterfactual situation, the community has not isolated arthritis for any special mention. (Incidentally, the disease is in actual fact not a physiological kind.) Counterfactually, the sign ‘arthritis’ applies to certain rheumatoid ailments, including forms of arthritis, but also including certain ailments of the muscles and tendons. Not only specialists but many layman use the sign in this way.

The thought experiment concludes with the observation that in the counterfactual situation, Bert lacks attitudes truly describable with ‘arthritis’
in oblique position. When he utters the signs ‘arthritis seems to have lodged in my thigh’, he expresses a belief that may be true, not one that is false. When before \( t \), he compares his disease with Aunt Mary’s, he is not thinking of his disease as arthritis (even though the disease he is thinking of is arthritis). No one in the counterfactual community thinks of any disease as arthritis.

What do the thought experiments show? They show that the intentional content of ordinary propositional attitudes, as indicated by obliquely occurring expressions in that-clauses, cannot be accounted for in terms of physical, phenomenal, causal-functional, computational, or syntactical states or processes that are specified nonintentionally and are defined purely on the individual in isolation from his physical and social environment. Intentional content, in the aforesaid sense, is not even supervenient on the nonintentional processes and states of an individual, insofar as these processes and states are ‘individualistically’ described. Thus individualistic functionalist, computationalist, or physicalist accounts of ordinary intentional content fail in a systematic manner.

The thought experiments also strongly suggest that conventional meaning is not illuminatingly seen as a “logical construct”, to use Fodor’s phrase, out of individuals’ propositional attitudes. They suggest that what a person’s propositional attitudes are is neither determined nor explicable independently of the socially determined meaning of his words. This is not to say that one could not fix social meaning given the propositional attitudes of all members of a community, together with their linguistic behavior. But note carefully that such “fixing” does not yield a logical construct. It is not yet a construction, and it yields no insight into “conceptual priority”. The propositional attitudes of the individuals in a community do not seem to be individuated independently of publicly accessible meaning. Meaning and propositional attitudes seem to be coordinate notions. In view of Quine’s work, this conclusion does not seem surprising. Nor do I think it should be rued. Philosophical attempts at logical constructions that reveal conceptual priorities seem to me to have yielded little insight. They represent, I think, an unpromising way of doing philosophy.

The thought experiments are not incompatible with the idea that there are token mental representations, unique to the individual thinker, in every case of an occurrent propositional attitude. They are also not incompatible with the view that such tokens enter into casual relations in virtue of their “syntax”, as opposed to their intentional features. These views constitute the principal tenets of what Fodor calls “Cognitive Science”. (In my view, the term is best taken to be like “Christian Science” not only in denoting a doctrine rather than a discipline, but also in being a proper name rather than a description.) These tenets are variants on a very old, very resilient philosophical interpretation of mentalistic ascriptions. I think that Fodor has provided a challenging defense of this philosophical viewpoint in earlier writing [6], [7]. As I said, the thought experiments are compatible with these tenets. I do doubt their truth. And these doubts are not unrelated to the thought experiments. But the relations are too complex to discuss here.

The two tenets of Cognitive Science make an ontological claim and a claim about the form of causal mechanism. Neither claim is incompatible with the thought experiments—which refute a view about how intentional content in
ordinary propositional attitudes is determined and how it is to be explicated. Fodor does, however, appear to hold such a view. He often writes that he expects computational relations among inner syntactical tokens to "explain" the content of propositional attitudes (see [6], pp. 75-77; and [7], pp. 231, 234-235, and 239-240). At the very least, Fodor fails to clearly distinguish a claim about what "accounts for" the causal efficacy of propositional attitudes (which I think is his primary concern) from a claim about what determines or "accounts for" their (opaquely, or obliquely specified) intentional content. Moreover, in his present piece, he resists the view that "the content of a mental representation is not a function of psychological variables as cognitive scientists understand such variables". (Although he does not say so, he seems to assume that such psychological variables must be individualistically specified.) And he seems to want to find some way of reinterpreting the thought experiments so as to avoid their antireductive, anti-individualistic consequences. On the other hand, much of his animus seems to be directed against inferences from the thought experiments, or closely related thought experiments, that I do not draw.

Let me make a very abbreviated pass at sorting some of this out. In practice, cognitive psychologists, including those who embrace the tenets of Cognitive Science, make nontrivial use of intentional attributions. They attribute propositional attitudes. So one would think that intentional states and processes would count as psychological variables. But since such states and processes are specified by reference to intentional content (i.e., in terms of obliquely occurring terms in that-clauses), content is trivially a function of psychological variables. It is only if one invests 'psychological variable' with some special, philosophical, individualistic meaning that the thought experiments show that content is not a function of psychological variables. Insofar as that content is what is adverted to in use of ordinary English discourse about propositional attitudes, it is not individualistically determined.

From here, the issues become complicated. For one may or may not take different views toward ordinary propositional attitude discourse and idealized psychological ascriptions. One might hold that some idealized psychological explanation uses nonindividualistic notions. I am sympathetic with this view but only given a certain reservation: nonintentional specifications of computational relations are parasitic on ordinary intentional attributions. I see no clear reason to believe that this parasitism will ideally be dispensed with. Alternatively, one might hold that idealized psychology must purge nonindividualistic elements from current psychology. Having taken this position, one might either seek some individualistic revision of ordinary intentional discourse or claim that intentional explanation must be dispensed with altogether in psychology.

I think that Fodor is right in seeing this last eliminationist alternative as poorly motivated. Thus I do not take the thought experiments to place the notion of intentional content "in jeopardy"—even in psychology, much less for ordinary, macrodescriptive purposes. At the very least, Fodor is right in holding that the eliminationist viewpoint cannot reasonably argue from the theory or practice of "computational psychology".

The other "individualistic" strategy for interpreting idealized psychology is revisionist rather than eliminationist. Fodor does not explicitly adopt this
strategy either. Those philosophical attempts in this vein that have so far appeared strike me as patently inadequate. Perhaps, though, psychology—or some important part of it, such as a descendent of “computational” approaches—will yield individualistic notions that are both intentional and adequate to sophisticated explanatory purposes. I believe, however, that much of cognitive psychology will remain intentional, nonindividualistic, and intellectually worthwhile.  

All of these remarks about idealized psychology are rather speculative. My original claim about what the thought experiments show is, I think, not speculative. The intentional content of ordinary propositional attitudes cannot be accounted for in terms of physical, phenomenal, causal-functional, or syntactical states or processes that are specified nonintentionally and individualistically. Cognitive psychology, including computational approaches, now makes nontrivial use of ordinary intentional language (somewhat refined and made more precise). So the same point holds for intentional content, as specified in actual psychological theory.

Fodor does not adopt either of the individualistic responses to the thought experiments that I mentioned four paragraphs back. He appears to question parts of the thought experiments themselves. But the issue is somewhat unclear. He concentrates on Putnam’s counterpart of the first thought experiment and criticizes inferences from it that I am not concerned to defend. Moreover, he is none too careful in his reading of the argument of my “Individualism and the mental”. And he significantly misconstrues my background assumptions. In the remainder of this paper I will try to set forth the basics of my position by reference to some of Fodor’s remarks.

In the first place, our thought experiments are not primarily about meaning. They are about propositional attitudes. Here they differ from Putnam’s first thought experiment. Considerations about meaning play a role in setting the circumstances in which we say what the protagonists think. But I do not know a precisely formulable inference from a general principle about meaning to our conclusions about attitudes. For example, I am not assuming, absurdly, that whenever someone sincerely utters words that mean that p, he believes that p. The thought experiments depend on no such inference. Nor do they depend on an antecedently formulated theory of “concepts”. Our judgments about attitudes stand on their own. If one (nontheoretically) understands propositional-attitude discourse and knows how to apply it, one should realize that, in the situations we specify, it is plausible and certainly possible that the protagonists have the attitudes I say they have.

We turn now to Fodor’s central question: “What de dicto belief is a speaker of Twin English expressing when he says, ‘water₂ is wet’, ‘aluminum₂ is flexible’, ‘cancer is worse than arthritis’?” Fodor puts this question to Putnam. I shall presume to answer it from our viewpoint. There is great potential in the expression ‘de dicto’ for misunderstanding. For present purposes, I think Fodor would agree that we could replace the expression with the stipulation that we are interested in oblique, or not purely transparent, occurrences in belief ascriptions.

So understood, the question should be answered by Fodor’s “first gambit”: Al₁ believes that aluminum₂ [oblique occurrence] is flexible. The
protagonist Twin Earthian lacks the thought that aluminum [oblique occurrence] is flexible. He lacks any scientific competence, or (we may suppose) any exposure to a term like 'ZYX' that is closely associated with scientific usage. So it would be wrong to take him to be expressing the belief that ZYX [oblique occurrence] is flexible. Insofar as we lack a term that has the uses of ‘aluminum₂’, we should introduce one—‘aluminum₂’ would do—making sure that these uses are understood.

Fodor replies to the “first gambit” by saying that we do not know which belief the belief that [aluminum₂ is flexible] is. It seems to me that this reply does not raise any real problem for our point of view. One needs only to master Twin-Earthians’ usage to know which beliefs they express. I have set out the thought experiments in such a way as to suggest that this usage might be as rich and different from our use of the Earthian counterpart term as one likes. For example, ‘water₂’ or ‘aluminum₂’ might differ from ‘water’ or ‘aluminum’ not only in their referents, but also in their usage among businessmen, laymen, and so forth. I do not think it necessary to describe details of Twin-Earthian usage to make it clear that the meaning and use of these terms differ from those of ‘water’ and ‘aluminum’—and that we could master this usage and meaning, and associate them with ‘water₂’ or ‘aluminum₂’ for our own purposes. Similar points apply to ‘arthritis₂’. Under these circumstances, I see no philosophically relevant problem in understanding what beliefs we attribute with propositional attitude discourse that places ‘aluminum₂’ (etc.) in oblique position.

This answer seems to me to need no further defense at present. But since Fodor’s other “gambits” may be a source of distraction I will remark on them. I am in broad agreement with his rejection of the third gambit, the invocation of indexicality (cf. note 6). What of the second?

The idea is that Twin Earthians believe that sailboat masts (on Twin Earth) are made of aluminum; Bert counterfactually believes that arthritis has lodged in his thigh. This line may have had some initial plausibility in reference to Putnam’s original Twin-Earth experiment, though it seems to me quite mistaken there too. But it lacks any plausibility as applied to our thought experiments. One only needs to understand how we use propositional attitude discourse to realize this. Bear in mind here that we are discussing ordinary propositional attitude notions. One can go on later to inquire “what to do about them”.

There is no ordinary means of explaining how Twin Earthians might have acquired the beliefs attributed by the “second gambit”. No one on Twin Earth has ever had contact with aluminum. No word on Twin Earth, at least no word used by the protagonist, means what ‘aluminum’ does. In the arthritis case, it is again hard to see why Bert, in his counterfactual situation, could be seen as having acquired the relevant beliefs since no one on Twin Earth ever isolated arthritis for special consideration.

Further, there are no grounds for attributing to the Twin Earthian what are patently false beliefs. There are no aluminum sailboat masts on Twin Earth; and arthritis can no more occur in the thigh in the counterfactual situation than it can actually. There is no ground for seeing Al or “counterfactual” Bert as being mistaken about these matters.

The second gambit is no better off if one claims that Al thinks (de re) of
ZYX that it is aluminum, or of ZYX that sailboat masts are made of aluminum [oblique occurrence]. Under the circumstances I described, we just do not attribute these false beliefs, or indeed any attitudes by using discourse that places ‘aluminum’ or ‘arthritis’ in oblique position. What is more, in the arthritis case, there is no difference in the res that Bert holds beliefs of, between actual and counterfactual circumstances. So there should be no temptation to think that the difference between actual and counterfactual situations can be explicated purely by reference to differences in the res that his attitudes apply to.

Fodor’s own positive solution to the Twin-Earth problems, in terms of pragmatically governed shifts in the domain of discourse, is a variant of the view discussed in the previous paragraph. The proposal is subject to the same criticisms. It is not plausible, when one considers the actual sentences of propositional-attitude attribution that would have to be asserted, that the same terms (‘aluminum’, ‘arthritis’, etc.) can remain in oblique position when we attribute propositional attitudes to the protagonists in actual and counterfactual situations. I also think that the modal problems Fodor raises for his view are more serious than his replies indicate. Moreover, the proposal will not apply to the second thought experiment since there is no shift in the domain of discourse. The change is in the language spoken, and in the usage and attitudes among those in the protagonist’s social environment.

We turn now to the “fourth gambit”. I find it quite baffling that Fodor attributes to me the view that on Twin Earth “water\textsubscript{2} is wet” is used to express the belief that XYZ is wet, or the “corresponding solution” for the examples I discuss. I have never held or expressed such a view of Putnam’s thought experiment, or of any of the cases I constructed. I take it that if the protagonist lacks a term that is closely associated with specialized knowledge and lacks the specialized knowledge, we do not attribute attitudes using the term in oblique position. Since the protagonist on Earth in Putnam’s example lacks the term H\textsubscript{2}O and lacks knowledge of chemistry, we should not attribute the belief that H\textsubscript{2}O (oblique occurrence) fills rivers and streams. Similar remarks apply to ‘XYZ’ and the Twin-Earthian counterpart.

Perhaps the misunderstanding derives somehow from Fodor’s misstatement of my contract example. This example was one of many, and in some respects it is not a central case. But Fodor’s criticisms largely rest on a misreading of it. So I shall discuss it. The root of the trouble lies in the statement that I assume (or ask someone to assume) that the fact that contracts need not be written is constitutive of our concept of contract. In fact, I assumed only that it is not “constitutive of our concept” of contract that contracts must be written. (I use the scare-quoted terms only in a loose nontheoretical way, and gave similar warning in [4], p. 81.) I explicitly disavowed any reliance on a philosophical view of concept-constitution ([4], p. 88). I share Quine’s distaste for using such notions for philosophical gain. What I did claim was that a person who thought that contracts must be written would be mistaken and that we could imagine a society, where the word ‘contract\textsubscript{2}’ was explicitly and strictly confined to written agreements, in which a physically identical person, described individualistically, would not hold that mistaken belief.

The rest of Fodor’s criticism flows from this initial mistake and is made
irrelevant by it. The claim that, on my view, Jones’ denial that verbal contracts bind is “explicitly contradictory”, the points about the principle of charity, the substitutivity of identity, and so forth, all miss the mark because of the initial misstatement of my view. Similarly, his claim that I place excessive weight on the notion of same language is misdirected, or at least needs sharpening. I nowhere make heavy use of the notion of same language or same language community. I am quite free to allow considerable pragmatic latitude in using these notions (cf. [4], pp. 91-92). All I rely on is the fact that individuals actually regard themselves as responsible to linguistic or conceptual norms that might be applied to them by others. This much seems implicit in the notion of interpersonal agreement and disagreement.

I shall conclude by briefly discussing Fodor’s claim that I cannot reasonably maintain the five assumptions he lists. I think that this claim is right. But I do not find it a source of discomfort.

1. I reject the first assumption. I agree with Fodor that is is untrue that verbal contracts bind is constitutive of the meaning of ‘contract’. More generally, as I have said, the thought experiments do not depend on any philosophical views about concepts or concept constitution.

2. I am not happy with the second assumption. I doubt that the meaning of a word is rightly seen as a construct out of anything, and I would not take the notion concept as a good starting point for a construction in any case. I do suppose that there is something commonsensical and right about saying that two words that express the same concept are synonymous, as long as the saying is not pressed. (The converse is more problematic.) But synonymy does not play a critical role in the thought experiments, nor does ‘same concept’. I rely only on judgements about nonsynonymy, and nonequivalence among obliquely occurring expressions.

3. I would not accept without serious qualification the statement that when Jones says ‘contract’ he expresses the same concept that we do when we say ‘contract’. This is partly because of wariness about ‘same concept’. But it is partly because, on a perfectly ordinary reading, the statement is not true. I have discussed parallel subtleties in [4], pp. 100-102. What I do accept as true is that Jones might really believe (mistakenly) that there are no verbal contracts [oblique occurrence]—and that he might have other, true beliefs specified with ‘contract’ in oblique position (such as that one should not sign a contract without reading the fine print) that we share with him.

4. I accept the principle of charity in the form: Do not attribute simple, explicitly contradictory beliefs (like a belief that contracts are sometimes verbal and are never verbal). But this form of the charity principle is not relevant to our thought experiments. The mistake Jones is charged with in the previous paragraph is not an explicit contradiction. It is a mistake any rational person could make. I discussed just these matters in [4], p. 100, and in [3]. As far as I can see, Fodor overlooked this discussion. The crux of my view is that there is nothing impossible about certain sorts of ignorance. One can believe that aluminum occurs in masts without knowing about features of aluminum that would distinguish it from other metals that one does not often encounter.
One can believe that arthritis is lodged in one's thigh, even though occurrence in joints is a distinguishing feature of arthritis. One can think that contracts must necessarily be written, even though being written is not a fundamental feature of contracts. I find these claims rather humdrum. But they do run against a very comprehensively articulated and deeply entrenched philosophical tradition (cf. [4]). The suggestion that alternatives to this tradition be pursued is the thought experiments' main claim to philosophical interest.

5. In earlier work [3], I have explicitly rejected assumption 5, the general intersubstitutability of synonymous expressions in oblique belief contexts (cf. also [4], note 4). This rejection is abetted by giving up the traditional viewpoint challenged by the thought experiments. But it is not essential to our view of the thought experiments since they nowhere rely on pairs of synonyms.

Attention to basics has prevented my discussing the relevance of the thought experiments to idealized cognitive psychology, except by the way. Although I see this matter differently from Fodor, I have some sympathy with his resistance to what I regard as too cavalier a rejection of (or instrumentalism about) propositional attitude discourse. My aim here has been to give some sense of the power and simplicity of the thought experiments and to point out their effect on certain reductionistic philosophical programs.

NOTES

1. The reader is to be warned that an accurate assessment of the present discussion cannot dispense with a careful reading of the original papers in which the thought experiments are given. The two thought experiments that follow are set out in [5] and [4], respectively. Actually both are stated in [4] but only the second is discussed in detail; the first occurs only in footnote 2. Putnam's argument appears in [8].

2. As far as I can see, it does not matter whether Twin Earth scientists know that aluminum exists. Suppose that they do not.

3. There may be projects that could be called "Gricean", but that do not involve attempts at "logical construction", with which I could have more sympathy. They would have to be specified more precisely before they could be fruitfully discussed. I do not see why Fodor is attracted to Gricean "logical construction". It does not seem to be needed by Cognitive Science, as he explains it. Nor do I see any strong motivation for it. There is the point that some (animal) propositional attitudes are independent of socially accessible meaning. There is the point that people normally know their own intentions (obliquely specified). And there is the point that propositional attitudes serve in the explanation of behavior. But I do not think that these points support the logical construction view. Fodor does not say how they, or other points, might be supposed to.

4. In making this remark, I must issue a caution about it. A lot of philosophy hangs on what one expects from psychology and what one takes to be intellectually worthwhile. I cannot undertake to develop my views on these issues here, except to say that I am not monistic about either "science" or the tasks of the intellect.

5. To be sure, one might ask for a fuller description of the situations. I have tried to comply, at probably unnecessary length, in [4].
6. The problem is that a belief may be de re, with respect to either some aluminum or some other entity, yet the term ‘aluminum’ may still occur obliquely. (‘Aluminum’ may do double duty both as a referring term and as an attitude characterizer.) The argument in Fodor’s note 11 seems to me to be flawed because it overlooks this point. I agree, however, that ‘water,’ ‘aluminum,’ ‘arthritis,’ and so forth are not indexical, and have so argued in [5], pp. 103-107. There are other things that Fodor says about de dicto attitudes that I do not accept, although some of these differences are terminological (cf. [2], [4], [5]). I will not go into these matters here.

7. This reply is backed by a criticism of Putnam’s lexical theory that I do not fully understand. But since I need not defend that theory, I shall ignore the criticism.

8. I think that Fodor is calling attention to an interesting and genuine pragmatic phenomenon, but that he is misapplying it to the cases at hand. There are other misapplications as well. The Marco Polo example is easily accounted for in terms of underlying tense constructions. The John the Baptist example (note 18) is best accounted for in terms of the indexicality of proper names (cf. [1]). I believe that indexicality is also involved in the Chinese cookie example (note 18). In order to understand the pragmatic phenomenon, we need a more careful discussion of its scope and limits.

REFERENCES


Department of Philosophy
University of California at Los Angeles
Los Angeles, California