

A Philosophically Inexpensive Introduction to Twin-Earth

Bryan Frances
Fordham University

I say that it's philosophically inexpensive because I think it is more convincing than any other Twin-Earth thought experiment in that it sidesteps many of the standard objections to the usual thought experiments. I also briefly discuss narrow contents and give an analysis of Putnam's original argument.

THE THOUGHT EXPERIMENT

When I was a child I sometimes played a joke on my older sister Leslie. My best friend Greg and I would say 'yes' when we meant *no* and we'd say 'no' when we meant *yes*. Obviously, this caused much confusion in Leslie until she caught on.

It's easy to see that the English language could have been different in that 'no' had always meant *yes* and 'yes' had always meant *no*. There's nothing magical in the symbols 'no' that they must mean something negative! If that had happened, if the words had always had meanings opposite to what they actually are, then when someone responded to the question 'Is there milk in the refrigerator?' with 'Yes' they would really mean *no*. There's nothing controversial or odd about that.

We just imagined a possible world in which the meanings of 'yes' and 'no' were switched from their real, actual meanings. Now I want you to imagine another possible world, one that switches the meanings of two other words.

But first we say something about the actual world. As things *actually* stand, in our (real) world, the terms 'walleye' and 'sauger' (pronounced SAW-GER) each stands for a common North American game fish, just like trout, salmon, cod, etc. You can trust me on this fact, as my father lives in Minnesota, fishes quite a bit, has caught lots of both kinds of fish, which are of course not the same, and insists on telling me about his fishing adventures. The two fish species are very

similar to the eye, just like gold and “fool’s gold” or elm trees and beech trees. Only an expert can tell the two kinds of fish apart (usually by the pattern of scales, if I remember right). Now imagine an alternative possible situation (or “world”) in which ‘walleye’ and ‘sauger’ are switched. So in this imaginary world ‘walleye’ has always been used to pick out sauger and ‘sauger’ has always been used to pick out walleye. So when someone in that world says ‘Walleye are big’ she has really said that sauger are big.

Now suppose that in this imaginary world my father Ron had read an article in an authoritative fishing magazine that contained the sentence ‘Walleye get bigger than sauger’. Clearly, what that sentence means in that world is that sauger get bigger than walleye (because the meanings of ‘walleye’ and ‘sauger’ are switched in that imaginary situation). So when the author wrote ‘Walleye get bigger than sauger’, she was expressing the idea that sauger get bigger than walleye. This is a false idea, but it’s an idea all the same. Suppose also that Ron believed what the article said. So in that imaginary situation he believes what the author said, namely that sauger get bigger than walleye—although he expresses his belief the same way as the author does in that world, with the sentence ‘Walleye get bigger than sauger’.

Now pretend that in this fantasy world Ron tells me his opinion. He says to me ‘Walleye get bigger than sauger’, thereby echoing the fishing article author. As before, what he means with that sentence is that sauger get bigger than walleye. That’s what he and the author believe—that sauger get bigger than walleye.

Let’s say that in that world I am agreeable about such matters, and so I take my father at his word. That is, I come to accept his belief. Like him, I believe that sauger get bigger than walleye and I express this belief just like everyone else does in that world, with the sentence ‘Walleye get bigger than sauger’.

Now let us leave the realm of imagination and return to the real world. A few years ago when I lived in Minnesota Ron actually told me that walleye get bigger than sauger. And since the only language he knows is English, he used the sentence that actually expresses that idea—the sentence ‘Walleye get bigger than sauger’. So as things actually stand today, I believe that walleye get bigger than sauger. In sum:

In the Imaginary World:

‘Sauger’ picks out walleye
‘Walleye’ picks out sauger
‘Walleye get bigger than sauger’ means that
sauger get bigger than walleye.
Ron (my father) believes that sauger get bigger than walleye.
Ron expresses that belief with ‘Walleye get bigger than sauger’.
Ron says to Bryan ‘Walleye get bigger than sauger’.
Bryan believes that *sauger get bigger than walleye*.

In the Actual World:

‘Sauger’ picks out sauger
‘Walleye’ picks out walleye
‘Walleye get bigger than sauger’ means that
walleye get bigger than sauger.
Ron believes that walleye get bigger than sauger.
Ron expresses that belief with ‘Walleye get bigger than sauger’.
Ron says to Bryan ‘Walleye get bigger than sauger’
Bryan believes that *walleye get bigger than sauger*.

So in the two worlds I believe different things. That’s because it’s as plain as day that the imaginary belief is true just in case *sauger get bigger than walleye*, whereas my actual belief is true just in case *walleye get bigger than sauger*. One belief’s truth condition is just the reverse of the other belief’s truth condition. One says that A is bigger than B and the other says that B is bigger than A.

Now comes the clincher: there’s nothing preventing me from being physically identical in the two worlds. E.g., I need never see, smell, or touch either fish in order to come to have beliefs about them, so my experiences with the fish themselves can be identical (since there are none!). In addition, my fishing conversations with my father can be exactly the same physically:

“Dad, which fish gets bigger?”
“Walleye get bigger than sauger!”

“Okay. Whatever you say. Walleye get bigger than sauger.”

So although I hear the same words from my father in the real and fantasy worlds, I acquire different beliefs—his beliefs. In fact, all my encounters with ‘walleye’ and ‘sauger’ can be identical in the two worlds.

Of course, I hear *enough* about the two kinds of fish, using ‘walleye’ and ‘sauger’, to form beliefs about them—including beliefs that differentiate them. In each world I’ve heard all of the following sentences (as well as many others) asserted by authoritative sources:

Walleye are game fish.

Sauger are game fish.

We caught four walleye, three trout, and one sauger.

Walleye get bigger than sauger.

Only experts can tell walleye from sauger; they do so based on the scales I think.

Last night we had walleye for dinner.

We can, if we wish, suppose that what has been asserted with these sentences is true in each world. This means the two worlds differ in many physical ways, but these differences need not ever impinge on my body. I have had *enough* interaction with fish, ‘walleye’, and ‘sauger’ to take on my father’s beliefs.

The thought experiment is intended to prove two theses:

Linguistic Anti-Individualism:

Bryan is physically identical in the two worlds but the meanings of his words are different. So what we mean by our words isn’t fixed by what’s going on in our brains or body.

Mental Anti-Individualism:

Bryan is physically identical in the two worlds but his beliefs are different. So what we think or believe isn’t fixed by what’s going on in our brains or body.

When philosophers talk about anti-individualism they mean the mental one.

Thus: physical duplicates—people who have exactly the same physical qualities, brain processes, sensory experiences, utterances, etc.—can have different thoughts (e.g., me in the actual world and me in the imaginary world). Throughout their lives they experience or have identical visual fields; they utter identical words with identical pronunciations; they see the very same objects at the exact same times from the very same perspectives, etc. Surely everything in their respective and physically identical local environments seems precisely the same to them from the inside, from their own cognitive perspective. But how could this be compatible with them thinking different things? A very surprising result.

NARROW CONTENT

So perhaps I have different thoughts in the two worlds. More precisely: although in the actual world I believe that walleye get bigger than sauger, in the counterfactual world I believe that sauger get bigger than walleye. In the actual world my belief is true whereas in the counterfactual world my belief is false—even though the things the beliefs are about (the two kinds of fish) are identical in the two worlds.

Does this mean that I have different thoughts in the two worlds? Well, they surely differ in their truth conditions. Here's what that means. My imaginary belief is true just in case sauger get bigger than walleye; that's the condition in the world that will make that belief true. But my actual belief is true just in case walleye get bigger than sauger. So those two beliefs are true under different conditions. That's what we mean by saying that they differ in their truth conditions. So if the Anti-Individualist is right then the thoughts definitely differ in one important way. So any theory that tries to understand a thought's truth conditions internally is doomed to fail. Fair enough. Still, one might think that the two beliefs don't differ in any psychologically important way. The properties of my beliefs that are important when it comes to understanding me do not differ across worlds.

Here's an analogy that motivates this position. You probably do not believe, like I do, that baseball was invented in 1863. But pretend that you do. Let's say that you're right; baseball was invented that year. Your evidence is that you read it in a couple baseball history books and have heard it orally from experts as well. So you know that baseball was invented in 1863. Now consider an alternative possible world in which your language is exactly the same (no word switching or other funny stuff) but baseball was invented in 1864, not 1863. Everything

else about baseball is the same; it just started one year later by the very same people and in the same way. For some reason, the experts have made an error and have logged the invention of baseball at 1863. So the books you read in this possible world have the exact same words—with the exact same meanings—in both worlds. The only difference is that in the real world the books are right whereas in the alternative world they're wrong. So in both worlds you believe (justifiably so) that baseball was invented in 1863. The actual belief amounts to knowledge whereas the counterfactual belief does not. But if you're interested in understanding me, you won't pay any attention to this difference! The fact that in just one world I know that baseball was invented in 1863 is not at all to the point if you're interested in understanding my psychology and behaviour. The knowledge difference has nothing to do with me; it concerns only the world outside of me. The knowledge difference is, we might say, explanatorily irrelevant when it comes to understanding me.

So we might say that in light of anti-individualism the truth conditions of our thoughts are explanatorily irrelevant as well. Just as in the case of knowledge, the difference in my beliefs has nothing to do with me; it concerns only the world outside of me.

So what about my beliefs is explanatorily relevant to understanding me and my behaviour? It must be some psychological property other than truth conditions; call it narrow content. The term 'narrow' is used to reflect the idea that the property in question is internal to me, so I have it in every world in which I'm physically duplicated. The term 'content' is used because whatever it is, it is very much like truth-conditional content. The task: produce a theory of narrow content and improve the argument given above for its existence.

WORDS, CONCEPTS, EXTENSIONS, PROPERTIES, REFERENTS

By keeping the following distinctions in mind you can avoid making lots of annoying mistakes when thinking about this Twin-Earth stuff.

We have our word 'dog', which is a linguistic string of three symbols. If I write on the wall 'Tom is a dog, a dog larger than our cat Fred', I have used the word 'dog' twice. We say that the single word 'dog' has two occurrences in that sentence on the wall. The occurrences are called tokens. The word itself, which has two occurrences there and many elsewhere, is called the word type.

Just to make things complex, we should distinguish between word tokens and uses of tokens. If I write ‘yes’ on a card and walk around with it, flashing it to people I want to say ‘yes’ to, then I have one token—the ink pattern on the card—that has many uses, where the uses are the individual flashes of the card.

Then there is our concept of a dog. This is a mental thing. It is part of thoughts, the thoughts about dogs. Much of the philosophy of mind is devoted to figuring out what concepts are. Then there is the concept’s extension, which is the set of dogs: the set of all things that “fall under” the concept.

In addition to those is the species, the dog species, which might be thought of as the referent of ‘dog’ and is somehow made up of all the individual dogs. For our purposes we can think of the species as a property.

PUTNAM’S MASTER ARGUMENT

Now we move on to Putnam’s original argument. I assume you have already read it (with reference to his *Philosophical Papers* vol 2).

- A. There is a possible world with the following two features. (i) Oscar₁ (on Earth) believes the intension/content of ‘water is wet’, call it C_E , as that sentence is used on Earth; similarly, Oscar₂ (on Twin-Earth) believes the intension/content of ‘water is wet’, call it C_{TE} , as that sentence is used on Twin-Earth. In order to distinguish the two uses of the sentence, let’s pretend that ‘water’ as used by Oscar₁ on Earth with H₂O, has an invisible subscript, so it may be written (with anti-invisible ink) ‘water_E’ (‘E’ for ‘earth’). Similarly, pretend that ‘water’ as used by Oscar₂ on Twin-Earth with XYZ, has an invisible subscript so it may be written ‘water_{TE}’. (ii) Oscar₁ and Oscar₂ are physically identical.
- B. The extension of ‘water_E’ differs from the extension of ‘water_{TE}’ (in the world described in (A)).
- C. Now for the sake of argument adopt **Assumption (II)**: Necessarily, intensions/contents determine extensions; so if the extension that word *A* has in a world differs from the extension that word *B* has in that world, then the intension/content that *A* has in that world differs from the intension/content that *B* has in that world. (Assumption (II) is on p. 219;

we'll get to Assumption (1) in premise (H).) So, for instance, if the extensions of 'water_E is wet' and 'water_{TE} is wet' differ in the world in question), then their intensions/contents differ.

- D. By (B) & (C) $C_E \neq C_{TE}$.
- E. Thus, by (A) & (D) Oscar₁ and Oscar₂ differ in their intensions/contents C_E and C_{TE} even though they are physically identical.
- F. Now make the **Narrow → Supervenient Assumption**: Necessarily, if Oscar₁ and Oscar₂ are physically identical, then they don't differ in their narrow psychological states. That is, it's necessary that *all narrow states are supervenient*. A state is narrow, by stipulation of 'narrow', if and only if a person could have it even if she were the only thing that existed in the whole universe. This Narrow → Supervenient Assumption is only implicit in Putnam as far as I can see.
- G. Thus, by (A) & (F) Oscar₁ and Oscar₂ don't differ in their narrow psychological states.
- H. Thus, by (E) & (G) C_E and C_{TE} are not narrow psychological states. This contradicts Methodological Solipsism or what Putnam also calls **Assumption (I)**, the thesis that intensions/contents are narrow psychological states. (Assumption (I) is on p. 220.)

The argument (A)-(H) is valid. The premises are (A), (B), (C), and (F). (A) is supposed to be unproblematic. He's right about that, although we still need to unpack what 'intension/content' amounts to. In the bulk of Putnam's essay (B) is established by Kripkean arguments regarding natural kinds. But I suspect that most people have realized that this is unnecessary; the premise can be established with much less work. In fact, Putnam *himself* shows exactly how to do this on pp. 226-7 with his remark on switching 'elm' and 'beech'! We could use any of many, *many*, examples that can be as realistic and convincing as one wishes; more on that below. Premise (F), the Narrow → Supervenient Assumption, is supposed to be unproblematic. It isn't, as you'll see below, but we can get around that problem, as we'll also see below. So, since Putnam has concluded that every premise except (C) is true, he concludes that if (C) is true, then (H) is true. That is, if assumption (II) is true, then assumption (I) is false. That is, *either assumption (I) or assumption (II) is false*; you can't have both. He later in the essay expressed a preference for (II) over (I); externalists have agreed with him in this regard, at least for contents construed as truth conditions.

Here's why the Narrow → Supervenient Assumption is doubtful. Arguably, even if a belief content doesn't supervene (because physical twins need not share it), it might still be narrow: some nearly divine genius could acquire it in an otherwise empty universe if given enough

sophisticated theorizing over zillions of years. That is: there *might* be beliefs that are narrow but not supervenient. Now, no one to my knowledge has *proven* that there really are such beliefs! All we know is this: just because you know that a belief is anti-supervenient doesn't mean you can *automatically* infer that it is wide. That is, it's not a matter of definition or logic alone that an anti-supervenient state has to be wide. Another way to put it: the inferences 'If anti-supervenient then wide' and the logically equivalent 'If narrow then supervenient' could conceivably be wrong. That's why the Narrow \rightarrow Supervenient Assumption is contestable.

However, regardless of what one thinks of that criticism, one could alter the Twin-Earth argument so that it ignores narrowness altogether and just focuses on supervenience. This is what people have in fact done since the time of Putnam's essay. They construe the basic externalist argument this way:

1. There is a possible world with the features (i) & (ii).
2. The extension of 'water_E' differs from the extension of 'water_{TE}'.
3. **Assumption (II)**: Necessarily, intensions/contents determine extensions; so if the extension that word *A* has in a world differs from the extension that word *B* has in that world, then the intension/content that *A* has in that world differs from the intension/content that *B* has in that world.
4. By (2) & (3) $C_E \neq C_{TE}$.
5. Thus, by (1) & (4) Oscar₁ and Oscar₂ differ in their intensions/contents C_E and C_{TE} .
6. By (1) and the definition of 'supervenient', Oscar₁ and Oscar₂ don't differ in their supervenient psychological states.
7. Thus, by (5) & (6) C_E and C_{TE} are not supervenient states. This contradicts a subtle variant of Assumption (I), **Assumption (I*)**, the thesis that intensions/contents are *supervenient* states or conditions.

Thus, even if (A)-(H) doesn't quite establish the falsity of Assumption (I) (because of lingering uncertainty regarding the Narrow \rightarrow Supervenient Assumption), it does establish the falsity of Assumption (I*). And the falsity of (I*) is obviously an externalist thesis.

When we take 'intensions/contents' to pick out *truth conditions*, then most everyone accepts assumption (II), even if they think that (II) is false for other kinds of content.

As far as I'm concerned, the primary thesis of externalism is without question the denial of (I*) for truth conditions. Then, the secondary thesis is the denial of (I*) for belief contents, where belief contents are the properties of beliefs we appeal to when explaining thought and behavior in ordinary life.