

PHILOSOPHY 1 LECTURE NOTES ON POSSIBLE WORLDS, REFERENCE, TWIN EARTH

Possible worlds. Various course authors say things like “this claim holds only in remote possible worlds.” What sort of a thing is a possible world?

Talk of possible worlds derives from work on modal logic (the logic of possibility and necessity) by Saul Kripke. The idea amounts to a framework for talking about possibility.

Possible worlds are complete ways that things might be. It is a common thought that things might have gone differently than they actually went. Your birthday party went well, say, but it might have gone badly if some of your friends had come to the party in a really grouchy mood. Bush won the U.S. Presidency in the last election, but Kerry might have won. A possible world is a complete way things might have gone, past, present, and future, down to the last detail, everywhere in the universe. One such world is the actual world. Along with the actual world there are huge numbers of complete ways things might have gone differently. Two possible worlds might differ from each other just in the duration of a blink of a dog’s eye (plus the differences in causes and effects needed to sustain this minor difference). We can’t actually describe any particular possible world in full detail or anything remotely resembling that, but the idea seems coherent. We can also talk about partial descriptions about how things might have gone in terms of subsets of possible worlds—there is for example the set of possible worlds in which you were never born. These possible worlds differ from each other in myriad ways but agree on this point, your not being born.

A possible world is a full description of how the history of the universe might have gone that is logically possible. The laws of logic set the only constraint on possibility in this wide inclusive sense. A possible world might include events that are physically impossible—not consistent with the fundamental laws of nature (whatever these turn out actually to be). If there are laws of metaphysics that rule out some logically possible things as metaphysically impossible, then metaphysically impossible but logically possible things are included in possible worlds. In ordinary conversation, you might say that something is impossible, meaning it is inconsistent with things we already know to be true of the actual world. It is in the colloquial sense impossible for my seven brothers to come to my funeral if I don’t have seven brothers, for example. But this is logically possible—there is no latent inconsistency in supposing I have seven brothers—so there are possible worlds, huge numbers of them, in which I have seven brothers and they come to my funeral.

There is another notion of possibility that comes up sometimes in the writings of course authors. This is epistemic possibility. This is the idea of what is possible relative to the speaker’s epistemic state—what is possible for all the speaker knows. At the end of Meditation 1, Descartes seems to be saying that for all the meditator knows, arithmetical and geometrical claims such as the claim that all triangles have exactly three sides might be false. Since I know very little physics, for all I know, quantum thermodynamics might be entirely false, an elaborate hoax perpetrated by jokers masquerading as physicists.

With the device of possible worlds in hand, we have a useful way of talking about what is necessary and contingent. Statements make claims about how things are. For each statement, there is a set of possible worlds at which the statement is true. If the statement is necessarily false, it is true at no possible worlds. If a statement is necessarily true, it is true at all possible worlds. It is impossible that the statement is false; at no possible world is the statement false. A

contingent statement is one that is neither necessarily true nor necessarily false. A contingent statement is true at some possible worlds and not others.

One statement entails or implies another statement if and only in every possible world where the first statement is true, the second statement is also true. “There are husbands” entails “there are wives.”

An essential property of an individual is a property an individual has, such that in any possible world in which that property of the individual is not present, the individual is not present in that possible world. The idea of an essential property contrasts with that of an accidental property. An accidental property is one that an individual has, such that there are possible worlds in which that property of the individual is not present, but the individual exist in that possible world. If being bald is an accidental property of a person, there are possible worlds in which he is not bald.

Reference. How do words relate to the world? How does the term “dog” get to pick out a certain class of animals, often pets and sometimes wild, closely related to wolves, sometimes called “man’s best friend”? The theory of reference addresses this issue.

Consider how proper names can be used to refer to particular people. A speaker can use a name such as “Einstein” to single out a particular individual and then say something about that individual even though neither the speaker nor the speaker’s audience are acquainted with that individual. One view is a description theory of reference for proper names. On this view, a name has a definite description attached to it, this is the sense of the name. So, the name “Einstein” might be associated with the description, the physicist who developed general and special relativity theory. The name “Einstein” refers to a particular individual in virtue of the associated definite description denoting *x*. A definite description, such as “the *F*,” denotes a particular individual if and only if “the *F*” applies to that individual and nothing else. So on this view I only successfully use a proper name to single out a particular individual only if I have associated with that name a definite description that uniquely applies to that individual.

You might be wondering, what about the many people and animals who are all named “Einstein”? Many individuals can share a name. This is true. But according to the description theory of reference for proper names, We use a name to refer successfully to a particular individual on a particular occasion only if on that occasion we associate a sense with the name that uniquely singles out one and only one individual.

Different persons may communicate successfully associating a name with a different sense, each sense denoting the same individual. When I speak to you about Einstein, a dog we both know, I associate the name with the description “smelly dog owned by our mutual friend who lives upstairs” and you associate the name with the description “beautiful dog who lives in this building” but both may be descriptions that apply to one and only one object, a particular dog.

A description theory of reference has also been devised for theoretical terms or natural kind terms—terms such as “water,” “lemon,” “heat,” “gold,” and so on and so on. Take *water*. We associate the term with a list of properties—being a liquid, turning to ice at cold temperatures, turning to steam when heated to a very high temperature, fills lakes and rivers, slakes thirst, is essential to life, and so on. There is a cluster of properties, and some particular batch of stuff does not have to possess every one of the properties to qualify as water. If something possesses enough of the properties in the cluster that defines a term, the term applies. There can be fuzzy borderline cases—how many is “enough”? On this view “water” means “the stuff that has these

properties ____ [a list is supplied]”. The term “water” refers to whatever has (enough of) those properties.

Now consider the fact that chemists have discovered that water has a certain chemical structure; water is H₂O. The description theory denies that the reference of the term “water” is fixed by this fact. After all, we used the term “water” (and cognate terms in other languages) before the chemists made their discovery about its chemical structure.

The description theory of theoretical or natural mind terms has the consequence that different people using a term like “water” but associating the term with different lists of properties are not using the term with the exact same meaning invariant from speaker to speaker. But this may not be a problem for the theory. What about the fact that people can use terms without having the foggiest idea what the associated list of properties is that gives the term its unique meaning? For example, even though I know nothing about “beech tree” and “elm tree” except that both are trees, I can use “beech” to refer to beeches and “elm” to refer to elms. How so? The suggestion is that my sparse list of properties associated with elm includes the property of being called “:elm” by the tree experts. This property does serve to separate beeches from elms in my usage.

Twin Earth. Hilary Putnam invented the Twin Earth story to explain what is wrong with the description theory of reference for theoretical terms. Twin Earth might be thought to be a planet like earth but in a distant part of our universe, or it might be thought to be a thing that exists in another possible world, not in the actual world. We imagine Twin Earth to be a planet that is for the most part a replica of earth. A counterpart of you exists on Twin Earth who is just like you; same goes for just about everything on earth. However there is this difference. On Twin Earth the watery stuff, the stuff that falls from the sky, is essential for sustaining life, slakes thirst, is liquid at medium temperatures and turns to ice at low temperatures and ice at high temperatures, happens not to have chemical structure H₂O but rather (we stipulate) XYZ. Also, on Twin Earth, there is H₂O, but it is black and tarry, not at all like water as it manifests on earth. The Twin Earthians call their watery stuff “water.”

Question: In our earth language, what is the stuff the Twin Earthians call “water”? Putnam thought, and most of us agree after thinking about it, that what they call “water” is not actually water, and that the black tarry stuff on Twin Earth is definitely water.

What’s the point? This is supposed to show the description theory is wrong. The term “water” in our speech does not mean “the stuff that has the following properties_____” —the watery stuff. The stuff that has those properties on Twin Earth is not water. What the example reveals is that all along, when we used the term “water” here on earth, even before the chemists made their discovery, we used the term for the kind of stuff that satisfies the description “slakes thirst” and so on with certain exemplars in view. We used water to refer to whatever kind of stuff, picked out by our description, has that fundamental nature. “Water” is whatever kind of stuff that wet stuff over there is, people said, pointing to examples. That stuff turns out to be H₂O. When people introduced the term into the language, they were presupposing that the examples they were pointing to—the stuff in the lakes, the stuff in the rivers—shared a common nature. This presupposition might have been wrong, in which case our usage would have come undone (this happens). But with water, it turns out that there was a common kind of stuff, the natural kind of the exemplars. (Complication: of course, we knew our exemplars of water were impure, mixed with other stuff like dirt. But we supposed there was a kind of stuff, in the exemplars, that could be separated from the impurities and had the properties in our canonical list. We used the term to refer to that stuff, whatever shared the fundamental nature of that stuff. That stuff is there on Twin Earth (in the story we have told), but oddly it is black and tarry there. Oddly, too, there is

some other stuff on Twin Earth that is as it were stuff that has the properties we associate with the cluster in the description theory account of the meaning of “water.” But this stuff, though superficially like water, is not water, not H₂O.

The term “water” in our speech refers to the stuff that we causally interact with on the occasions of picking out by pointing the stuff we define via the surface description when the term is introduced into the language. And what goes for “water” goes for other theoretical or natural kind terms as well. Neglecting this important role of causal interaction in the process that determines what the term “water” refers to is a mistake that the description theory of the reference of theoretical terms makes. Hence the theory that replaces the description theory (or maybe it is better to say, supplements it) is dubbed the “causal theory of reference.”

There is another objection that the Twin Earth story generates for the description theory of reference. The objection is that on reflection we understand the term “water” to be a rigid designator. “Water” rigidly designates the kind of stuff that we interacted with when we introduced the term into the language. A rigid designator names the same things in every possible world. Water, for example, rigidly refers to the stuff that has a particular nature—whatever turns out to be the nature of the stuff we were interacting with when we pointed to some lakes and rivers and declared we would use the term “water” to refer to the stuff *over there*, that slakes thirst, is liquid, odorless and colorless, and so on. The chemists discovered that the stuff has a fundamental nature; it is H₂O.

Many expressions are non-rigid designators: they refer to different things in different possible worlds. “The person alive in 2006 who has used the most steroids” is a non-rigid designator. Say the description fits Barry Bonds. This if so is only contingently so. Barry Bonds might have become a lawyer and never used steroids, or he might have become a baseball player but one who never used steroids, or a baseball player who used a lot of steroids but less than some cyclists, and so on. Even if he fits the description in the actual world it is possible that he did not, or in other words he does not fit the description in many possible worlds. “The person alive in 2006 who has used the most steroids” might be a description true of Fred, a Russian cyclist, for example.

It is conceivable that we might use a theoretical term to refer to a type of being, and successfully use the term to refer, even though all of the descriptions that we think apply to the things that we are causally interacting with when we introduce the term into the language, in fact do not apply. For example, we use the term “tiger” to refer to beings that have the nature of the beings in front of us when we associated the term “tiger” with *those things*, picking them out by a list of superficial descriptions—predator, yellow and black striped, big teeth, look somewhat like big cats, etc. It could conceivably turn out to be the case that the beings we refer to when we use the term “tiger” actually share none of the properties we identified with them. The true nature of the beings we call “tigers” turns out to be that they are disguised robots placed on earth by Martians to spy on us as part of a War of the Worlds strategy. Just suppose. This illustrates again both the importance of causal interaction in fixing the reference of the term “tiger” and the fact that the term “tiger” rigidly designates the beings that have the nature, whatever it is, of the things we picked out by the canonical superficial descriptions when we introduced the word. “Tiger” rigidly designates whatever in any possible world has the nature of those things, the particular things that served as our exemplars. “Tiger” rigidly designates whatever has that nature, and we are imagining that nature is Martian robot with feline disguise, in which case “tiger” would rigidly designate Martian robots with feline disguises in any possible world in which tigers exist at all.) If some possible world has mammal predators that look and act tigerish, those things would not be tigers.) (You might resist the claimed moral of this paragraph by denying that we succeeded in fixing any reference when we introduced the term “tiger” if actually nothing

answered to our initial surface descriptions, if nothing in front of us, pointed at by us when we said “tigers are defined as the things that have the nature of the yellow predator big-cat-like creatures over there,” actually was yellow predator big-cat-like etc.

Consider the fact that it was an empirical discovery that water was actually H₂O. Water is identical to H₂O. But then “Water = H₂O” is necessarily true, it is true in all possible worlds. In any possible world in which water exists, it is H₂O. The term “water” rigidly designates H₂O. It is not an a priori truth that water = H₂O. We did not learn this truth by pure reasoning independently of observation/experience. We learned this truth a posteriori—by observation/experience. “Water = H₂O” is then an a posteriori necessary truth. It seems A.J. Ayer was wrong to identify truths learned a priori with the analytic truths with the necessary truths and truths learned a posteriori with the synthetic and contingent truths. Some think Ayer runs together distinctions that are logically distinct and independent.

Why bother? We introduce the Twin Earth example and the lessons about reference that philosophers have drawn from it because when we get to the mind and body readings toward the end of the course, it will be important or at least is thought by some authors to be important to keep these lessons in mind and also to keep straight the various kinds of possibility we have distinguished.