Possible Worlds Semantics

M&E Core

September 4th, 2018

1. Consider the following claims:[1]

   (a) The first star visible at night = the first star visible at night.

   (b) The first star visible at night = the last star visible in the morning.

   (1a) is *a priori* and *analytic*. (1b) is not. Discovering (1b) constituted an advance in our knowledge. Not so for (1a). So (1a) and (1b) must differ in meaning.

2. It turns out (let's suppose) that all creatures that have hearts also have kidneys, and *vice versa*. A creature with a heart is called a 'cordate', and a creature with a kidney is called a 'renate'. So, it turns out that all cordates are renates, and all renates are cordates. Now, consider the following claims:

   (a) All renates are renates.

   (b) All renates are cordates.

   (2a) is *a priori* and *analytic*. (2b) is not. Discovering (2b) was an advance in our knowledge. Not so for (2a). So (2a) and (2b) must differ in meaning.

3. Frege dealt with sentences like these by distinguishing two different aspects of meaning: *sense*, on the one hand, and *reference*, on the other. Since Frege, another method for distinguishing the meanings of these expressions has become pervasive: a *possible worlds* semantics.

4. To set the stage for the possible worlds semantics, consider the following presentation of Frege's puzzle:

---

[1] We suppose that 'star' means any body visible in the night sky.

(a) On analogy with a standard semantics for first-order logic, let's suppose that the logical form of (2a) is $(\forall x)(x$ is a renate $\rightarrow x$ is a renate), the logical form of (2b) is $(\forall x)(x$ is a renate $\rightarrow x$ is a cordate), and let's suppose that the meaning of a predicate like 'is a renate' is just the set of things which are renates.

(b) We can use *interpretation brackets* '⟦ ⟧' as a function from expressions of the language to the meanings of those expressions. Then, the standard assumption of first-order logic is that the meaning of a predicate like 'is a renate' is given by the set of things which satisfy the predicate, which we can notate as follows:

$$⟦ \text{ is a renate } ⟧ = \{x \mid x \text{ is a renate }\}$$

And similarly, we will say that the meaning of the predicate 'is a cordate' is given by the set of things which satisify the predicate:

$$⟦ \text{ is a cordate } ⟧ = \{x \mid x \text{ is a cordate }\}$$

But, since we've seen that all and only renates are cordates,

$$\{x \mid x \text{ is a renate }\} = \{x \mid x \text{ is a cordate }\}$$

it follows that the meaning of 'is a renate' is the same as the meaning of 'is a cordate'.

$$⟦ \text{ is a renate } ⟧ = ⟦ \text{ is a cordate } ⟧$$

So, given compositionality of meaning, (2a) and (2b) will have the same meaning,

$$⟦ \text{ All renates are renates } ⟧ = ⟦ \text{ All renates are cordates}⟧$$

(On a referential Fregean semantics, this meaning will just be *the True*.)

5. Let's call these kinds of meanings—*e.g.*, the set $\{x \mid x$ is a renate$\}$ for the meaning of 'is a renate'—*extensions*. Then, the problem is that expressions which plainly differ in meaning—'is a renate' and 'is a cordate', *e.g.*—nevertheless have the same extensions. Possible worlds semantics solves this problem, not by admitting a new aspect of meaning for 'is a renate', *in addition* to its extension (as with Frege), but rather by *generalizing* extensions.

   (a) Begin with the notion of a *possible world*. There are metaphysical debates to be had about what exactly a possible world is, but, at the least, a possible world is a *maximal* way things could be. One way for things to be is that I don't wear socks on September 4th—that's a possible way for things to be, but it's not a possible world, since it's not a *maximal* way for things to be. At the least, a possible world should settle the truth-value of every sentence in the language of interest.

   (b) Predicates like 'is a renate' don't just have extensions *at the actual world*. They will also have extensions *at every possible world*. Call something which specifies the extension of 'is a renate' at every possible world the *intension* of 'is a renate'.

   (c) More formally, we can let an intension be a *function* from possible worlds to extensions. And we can represent the meaning of 'is a renate' and 'is a cordate' with their *intensions*. As a bit of notation, if 'P' is a predicate, then we can let '$[\![P]\!]^w$' denote the *extension* of the predicate 'P' at the possible world w. And, still using '$[\![P]\!]$' to stand for the *meaning* of the predicate 'P', we can say that the meaning of 'P' is its intension,

   $$[\![P]\!] = \lambda w.[\![P]\!]^w$$

   i. Here, '$\lambda w.[\![P]\!]^w$' denotes a *function*—the function *from* possible worlds w *to* 'P's extension at w.

   ii. In general, we can denote a function f as follows: $\lambda x.f(x)$. The way to read '$\lambda x.f(x)$' is like this: 'the function from x to $f(x)$.' It is thus a name for a function. For instance, $\lambda x.x$ is the identity function, $\lambda x.x + 1$ is the function *plus one*, and $\lambda x.\sqrt{x}$ is the square root function.

6. To get a sense of how the notion of an *intension* can be applied in general, let's consider a simple language.

   (a) The vocabulary of this language is simple: there are two *singular terms*, a and b, there are two 1-place predicates, P and Q, and there is one 2-place predicate of identity, =.

   (b) The syntactic rules are the familiar ones:
      i. If $\ulcorner \tau \urcorner$ is a singular term and $\ulcorner \Pi \urcorner$ is a one-place predicate, then $\ulcorner \Pi\tau \urcorner$ is a sentence.
      ii. If $\ulcorner \tau_1 \urcorner$ and $\ulcorner \tau_2 \urcorner$ are terms, then $\ulcorner \tau_1 = \tau_2 \urcorner$ is a sentence.

   (c) Now, an *extensional* semantics for our simple first-order language would work like this: we have a set of entities $\mathcal{D} = \{u, v\}$, known as the *domain*.
      i. The meaning of a singular term is some entity from the domain. For instance, it could be that

      $$[\![a]\!] = v \quad \text{and} \quad [\![b]\!] = v$$

      ii. The meaning of a one-place predicate is a *set* of entities from the domain. For instance, it could be that

      $$[\![P]\!] = \{v\} \quad \text{and} \quad [\![Q]\!] = \{v\}$$

      iii. The meaning of a *sentence* is either 0 (the false) or 1 (the true). The meaning of a sentence is determined compositionally from the meaning of its parts, according to the familiar rules:
         A. if $\ulcorner \tau \urcorner$ is a term and $\ulcorner \Pi \urcorner$ is a one-place predicate, then

         $$[\![\Pi\tau]\!] = \begin{cases} 1 & \text{if } [\![\tau]\!] \in [\![\Pi]\!] \\ 0 & \text{if } [\![\tau]\!] \notin [\![\Pi]\!] \end{cases}$$

         B. if $\ulcorner \tau_1 \urcorner$ and $\ulcorner \tau_2 \urcorner$ are terms, then

         $$[\![\tau_1 = \tau_2]\!] = \begin{cases} 1 & \text{if } [\![\tau_1]\!] = [\![\tau_2]\!] \\ 0 & \text{if } [\![\tau_1]\!] \neq [\![\tau_2]\!] \end{cases}$$

      For instance, given the stipulations above, $[\![a]\!] \in [\![Q]\!]$, so

      $$[\![Qa]\!] = 1$$

      Similarly, $[\![a]\!] = [\![b]\!]$, wherefore

      $$[\![a = b]\!] = 1$$

(d) To get an *intensional* semantics, we may take our extensional semantics and outfit it with an additional set of *worlds*, $\mathcal{W} = \{w_1, w_2\}$. For each world $w \in \mathcal{W}$, we provide an *extensional* interpretation $[\![\ ]\!]^w$. Finally, we let the *meaning* of any expression $\ulcorner \xi \urcorner$ just be its intension—a function from worlds $w \in \mathcal{W}$ to the expression $\ulcorner \xi \urcorner$'s extension in the world w, $[\![\xi]\!] = \lambda w.[\![\xi]\!]^w$.

i. As before, at a world w, the extension of a singular term is some entity from the domain $\mathcal{D}$. For instance, it could be that

$$\begin{array}{ccc} [\![a]\!]^{w_1} = v & & [\![a]\!]^{w_2} = v \\ [\![b]\!]^{w_1} = v & \text{and} & [\![b]\!]^{w_2} = u \end{array}$$

However, the *meaning* of a singular term isn't its extension, but rather its *intension*—a function from worlds to its extension in that world. For instance, the meaning of 'b' is a function from worlds in $\mathcal{W}$ to entities in the domain $\mathcal{D}$.

$$[\![b]\!] = \lambda w.[\![b]\!]^w$$

We can represent this function by showing where it maps every world in $\mathcal{W}$:

$$[\![b]\!] : \left\{ \begin{array}{l} w_1 \mapsto v \\ w_2 \mapsto u \end{array} \right.$$

ii. As before, at a world w, the extension of a one-place predicate is a set of entities from the domain $\mathcal{D}$. For instance, it could be that

$$\begin{array}{ccc} [\![P]\!]^{w_1} = \{v\} & & [\![P]\!]^{w_2} = \{v\} \\ [\![Q]\!]^{w_1} = \{v\} & \text{and} & [\![Q]\!]^{w_2} = \{u\} \end{array}$$

However, the *meaning* of a one-place predicate isn't its extension, but rather its *intension*—a function from worlds to the predicate's extensions in that world. For instance, the meaning of 'Q' is a function from worlds in $\mathcal{W}$ to subsets of $\mathcal{D}$.

$$[\![Q]\!] = \lambda w.[\![Q]\!]^w$$

We can represent this function by showing where it maps every world in $\mathcal{W}$:

$$[\![Q]\!] : \left\{ \begin{array}{l} w_1 \mapsto \{v\} \\ w_2 \mapsto \{u\} \end{array} \right.$$

iii. Finally, as before, at any world w, the extension of a *sentence* is a truth-value—either 1 or 0. This truth-value is determined at each world in the usual way. So, for instance,

$$\begin{array}{ccc} [\![P\,b]\!]^{w_1} = 1 & & [\![P\,b]\!]^{w_2} = 1 \\ [\![Q\,b]\!]^{w_1} = 1 & \text{and} & [\![Q\,b]\!]^{w_2} = 0 \end{array}$$

while

$$[\![a = b]\!]^{w_1} = 1 \quad \text{and} \quad [\![a = b]\!]^{w_2} = 0$$

However, the *meaning* of a sentence isn't its extension, but rather its *intension*—a function from worlds to its extension in that world. For instance, the meaning of '$Q\,b$' is a function from worlds in $\mathcal{W}$ to $\{0, 1\}$. We can represent this function by showing where it maps every world,

$$[\![Q\,b]\!] : \left\{ \begin{array}{l} w_1 \mapsto 1 \\ w_2 \mapsto 0 \end{array} \right.$$

(e) We can use the tables below to summarize some of the information contained in this simple intensional semantics.

	w_1	w_2
$[\![a]\!]$	v	v
$[\![b]\!]$	v	u
$[\![P]\!]$	$\{v\}$	$\{v\}$
$[\![Q]\!]$	$\{v\}$	$\{u\}$

	w_1	w_2
$[\![Pa]\!]$	1	1
$[\![Qa]\!]$	1	0
$[\![a = a]\!]$	1	1
$[\![a = b]\!]$	1	0

7. (a) An intensional semantics can distinguish the meanings of (1a) and (1b)

1a) The first star visible at night = the first star visible at night.

1b) The first star visible at night = the last star visible in the morning.

in precisely the way that our simple semantics above distinguished the meanings of '$a = a$' and '$a = b$'. Though 'the first star visible at night' and 'the last star visible in the morning' co-refer at the *actual* world, there are possible worlds at which they refer to different entities. So, while (1a) is true at every possible world, (1b) is false at some possible worlds. Since the meanings of 'the first star visible at night' and 'the last star visible in the morning' isn't just their referents, but rather a function from possible worlds to referents, (1a) and (1b) can be true at different worlds, and thereby have different meanings.

(b) Similarly, an intensional semantics can distinguish the meanings of (2a) and (2b)

   2a) All renates are renates.

   2b) All renates are cordates.

in precisely the way that our simple semantics distinguished the meaning of 'Pa' and 'Qa'. Though everything in the *actual* world which is a renate is also a cordate, there are possible worlds at which there are creatures with hearts but not kidneys, and there are possible worlds at which there are creatures with kidneys but not hearts. Since the meaning of 'is a renate' and 'is a cordate' isn't just the set of renates/cordates, but rather a function from possible worlds to sets of renates/cordates at those worlds, (2a) and (2b) can be true at different worlds, and thereby have different meanings.

(c) In this way, a possible worlds semantics is able to solve (at least some instances of) Frege's puzzle without needing to introduce *sense*. All meaning is reference—it is just reference, not only at the actual world, but additionally at *possible* worlds.

8. We've just seen that, on a possible worlds semantics, the meaning of a sentence $\ulcorner\phi\urcorner$, $[\![\phi]\!]$, is a function from possible worlds to $\{1,0\}$. Given such a function, we could construct a corresponding *set* of worlds—the worlds which get mapped to 1 by the function $[\![\phi]\!]$.

(a) Let's call the set of worlds mapped to 1 by $[\![\phi]\!]$ the *proposition* expressed by the sentence $\ulcorner\phi\urcorner$. And let's denote it with $\ulcorner\langle\phi\rangle\urcorner$.

$$\langle\phi\rangle \stackrel{\text{def}}{=} \{w \in \mathcal{W} \mid [\![\phi]\!]^w = 1\}$$

(b) We'll say that a proposition $\langle\phi\rangle$ is true at a world w iff $w \in \langle\phi\rangle$.

(c) One nice thing about this semantics is that it allows us to relate the the logical connectives $\neg, \wedge, \vee$, and $\rightarrow$ to standard set-theoretic operations. Note:

   i. $\langle\phi \wedge \psi\rangle = \langle\phi\rangle \cap \langle\psi\rangle$.

   ii. $\langle\phi \vee \psi\rangle = \langle\phi\rangle \cup \langle\psi\rangle$.

   iii. $\langle\neg\phi\rangle = \langle\phi\rangle^c$.

   iv. $\langle\phi \rightarrow \psi\rangle = \langle\phi\rangle^c \cup \langle\psi\rangle$

(d) In a possible worlds semantics, *propositions* play the theoretical roles which Frege associated with the senses of sentences (thoughts). They:

   i.   Explain (some) differences in cognitive significance (*i.e.*, (2a) and (2b)).

   ii. Are the primary bearers of truth and falsehood.

   iii. Are what's preserved in translation.

   iv. Are the objects of belief—that is, they are the referents of *that*-clauses in belief reports.

9. How much difference in cognitive significance can a possible worlds proposition account for? That depends. Suppose that we take 'Hesperus' to be synonymous with the definite description 'the first star visible at night', and we take 'Phosphorus' to be synonymous with the definite description 'the last star visible in the morning. Then, (1a) and (1b) will be synonymous with (**A**) and (**B**)

A) Hesperus = Hesperus

B) Hesperus = Phosphorus

Then, the possible worlds semantics would solve Frege's puzzle applied to proper names.

(a) Notice that this solution depended crucially upon our supposition that proper names like 'Hesperus' and 'Phosphorus' were synonymous with some kind of definite descriptions. We will revisit this assumption next week.

10. However, even if we grant that 'Hesperus' is synonymous with 'the first star visible at night'—and that, more generally, each proper name is synonymous with some definite description—it still looks as though there will be sentences with different cognitive significances which express the very same possible worlds proposition. For instance:

(a) On this theory, there can be only one necessarily true proposition. Since '$2 + 2 = 4$', '997 is a lucky prime', and 'torturing babies for fun is wrong' are all necessarily, true, they will all express the very same possible worlds proposition—they each express the set of all possible worlds.

(b) Similarly, on this theory, there can be only one necessarily false proposition. Since '2+2 = 5', 'Not every commutative ring has a prime ideal', and 'torturing babies for fun is obligatory' are each not only false, but *necessarily* false, they will all express the very same possible worlds proposition—they all express the empty set.

(c) Just as there are co-extensional predicates which nevertheless differ in meaning—like 'is a renate' and 'is a cordate'—there are *co-intensional* predicates which nevertheless appear to differ in meaning—like 'is triangular' and 'is trilateral'.

(d) Similarly, consider the sentences 'Socrates exists' and 'the singleton set containing Socrates exists'. These will be true in exactly the same worlds. But they appear to be saying different things.

11. The differences in meaning between (1a) and (1b) (and (2a) and (2b)) were *intensional* differences in meaning. In order to bring out their differences, we had to move to the level of intensions. If you think that there is meaning more fine-grained than this—if you think that the co-intensional sentences in (10) above have different meanings, then you think that there are *hyper-intensional* differences in meaning.