

# Ramsey/Lewis Method of Defining Terms

Let's say we want to explain what the different parts of a car are. Suppose we have a theory that says how the different parts of a car interact with each other, and with things our audience already understands, like air and gasoline. The theory might look something like this:

Car Theory: ...and **the carburetor** mixes *gasoline* and *air* and sends the mixture to **the ignition chamber**, which in turn...and that *makes the wheels turn*.

The **bold** terms are names for parts of the car, with which our audience may not be familiar. The *italicized* terms are names for things and phenomena we'll suppose our audience already understands.

Now, given this Car Theory, how might we go about explaining to people what a carburetor and an ignition chamber and the rest are? We can't just define a carburetor as something that interacts with the ignition chamber in such-and-such ways, because our audience doesn't yet know what an ignition chamber is.

What we can do is the following. First, we transform our Car Theory into an existentially quantified sentence, quantifying out all the bold terms our audience doesn't yet understand.

$\exists \mathbf{x}_1 \mathbf{x}_2$  (...and  $\mathbf{x}_1$  mixes *gasoline* and *air* and sends the mixture to  $\mathbf{x}_2$ , which in turn...and that *makes the wheels turn*.)

This is called the **Ramsey Sentence** for our Car Theory (after the philosopher and mathematician Frank Ramsey).

Next we can define what it is to be a carburetor and an ignition chamber as follows:

A carburetor = an  $\mathbf{x}_1$  such that  $\exists \mathbf{x}_2$  (...and  $\mathbf{x}_1$  mixes *gasoline* and *air* and sends the mixture to  $\mathbf{x}_2$ , which in turn...and that *makes the wheels turn*.)

An ignition chamber = an  $\mathbf{x}_2$  such that  $\exists \mathbf{x}_1$  (...and  $\mathbf{x}_1$  mixes *gasoline* and *air* and sends the mixture to  $\mathbf{x}_2$ , which in turn...and that *makes the wheels turn*.)

In this way, we explain what a carburetor is, in terms of how it interacts with ignition chambers and with other things, without presupposing that our audience already knows what an ignition chamber is. In the same way, we explain what an ignition chamber is, in terms of how it interacts with carburetors and with other things, without presupposing that our audience already knows what a carburetor is.

In addition, we've explained what a carburetor and an ignition chamber are in

terms of the *causal roles* they play, as specified in our Car Theory. *Any* pair of things which play the appropriate causal roles count as a carburetor and an ignition chamber. The details of their physical construction are not important. In other words, carburetors are **multiply realizable**. To be a carburetor, it doesn't matter what you're made out of; only that you do the right job. (The same goes for ignition chambers.)

So this method of defining terms gives us the two benefits that we relied on Turing Machines for, earlier. It lets us:

- i. define things like carburetors in terms of how they interact with other things, like ignition chambers, without presupposing that the notion of an ignition chamber is already understood
- ii. define carburetors in such a way that they can be realized by different physical mechanisms

How might we apply this method of defining terms to the mind? Well, suppose we have a theory about how our various mental states are causally related to each other, and to input and output:

Mental Theory: ...and **pain** is caused by *pin pricks*, and **pain** causes **worry** and *the emission of loud noises*, and **worry** in turn causes *brow-wrinkling*...

As before, the **bold** terms are names for mental states with which our audience may not be familiar. The *italicized* terms are names for various sorts of sensory stimulation, and behavioral output, which we'll suppose our audience already understands.

Now, we take the Ramsey Sentence for our Mental Theory:

$\exists \mathbf{x}_1 \mathbf{x}_2$  (...and  $\mathbf{x}_1$  is caused by *pin pricks*, and  $\mathbf{x}_1$  causes  $\mathbf{x}_2$  and *the emission of loud noises*, and  $\mathbf{x}_2$  in turn causes *brow-wrinkling*...)

Next we define what it is to be in pain, and to be worried, as follows:

A person is in pain =  $\exists \mathbf{x}_1 \mathbf{x}_2$  (...and  $\mathbf{x}_1$  is caused by *pin pricks*, and  $\mathbf{x}_1$  causes  $\mathbf{x}_2$  and *the emission of loud noises*, and  $\mathbf{x}_2$  in turn causes *brow-wrinkling*...) & the person has  $\mathbf{x}_1$ .

A person is worried =  $\exists \mathbf{x}_1 \mathbf{x}_2$  (...and  $\mathbf{x}_1$  is caused by *pin pricks*, and  $\mathbf{x}_1$  causes  $\mathbf{x}_2$  and *the emission of loud noises*, and  $\mathbf{x}_2$  in turn causes *brow-wrinkling*...) & the person has  $\mathbf{x}_2$ .

The functionalist thinks that all of our mental states can be defined in this way. Anything which has states which play those causal roles counts as having a mind, and whenever it's in the first of those states, it's in pain, and when it's in the second of those states, it's worried. It does not matter what the intrinsic make-up of those states is. In humans, they are certain kinds of brain states. In Martians, they would likely be different sorts of states. In an appropriately-programmed computer, they would be electronic states. These would be

different physical realizations of the same causal roles. The functionalist identifies our mental states with the causal roles. How those roles are realized is not important.

### Common-Sense vs. Scientific Functionalism

Where does the theory that the functionalist uses to define our mental states come from?

- The **common-sense functionalist** says that the theory is an *a priori* theory, made up of platitudes about our mental states that everyone who has the concepts of pain, belief, and so on, tacitly knows, or at least, is in a position to recognize as true. (This is also sometimes called analytic functionalism. Block simply calls it "Functionalism.")
- The **scientific functionalist** says that the theory is an *a posteriori* theory, which we only learn as a result of scientific investigation of how our minds work (the sort of investigation they do in cognitive science labs). (This is also sometimes called empirical functionalism. Block calls it "Psycho-functionalism.")

Note that not *every* causal fact about a mental state has to enter into the functionalist's definition of pain. Different carburetors can have causal properties which play no role in making them carburetors: my carburetor might shimmy a bit, while yours whistles. These facts are irrelevant to their being carburetors. In just the same way, our mental states might have some causal properties which play no role in making them the mental states they are.

It is a very difficult matter to know which of a mental state's causal properties ought to enter into the definition of that mental state, and which are merely accidental.

---

URL: <http://www.courses.fas.harvard.edu>

Last Modified:

Copyright © The President and Fellows of Harvard College

---