

# Philosophy 100: Introduction to Philosophy

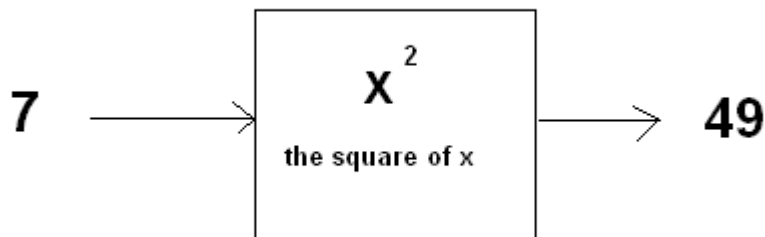
Department of Philosophy  
University of British Columbia

Richard Johns

## Functionalism

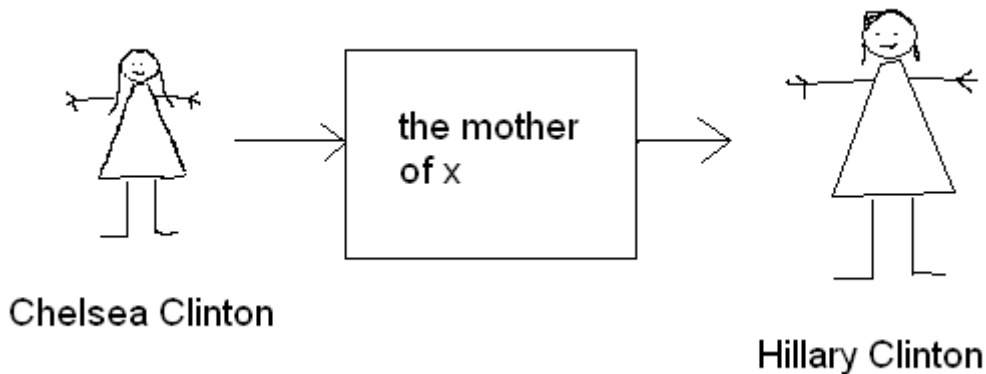
### 1. What is functionalism?

First, we should make sure we know what a *function* is. The basic idea is very simple: one can think of a function as a kind of *machine* that accepts one or more inputs, and produces a single output for each input. Consider, for example, the well-known *square* function:



This function accepts *numbers* as inputs (its “domain” is the set of numbers) and produces numbers as outputs. For a given input, it always produces the same output. For example, the input 7 always generates the output 49, and the input 3 always generates the output 9.

Not all functions operate on numbers. They can operate on any kind of object at all. Consider, for example, the *mother function*.



This function accepts *human beings* as inputs (not their names, the *actual person*). When you put someone into the machine's input, out pops their mother from the other end! Note that this function, like the square function above, always delivers a *single* output for each input (assuming that each person has exactly one mother). Also, a given input always produces the *same* output. Other functions that accept two inputs are *addition* and *multiplication*.

While functions are an abstract, mathematical idea, some functions can be “instantiated”, or realised, by a physical system. A calculator, for example, is a physical system that realises several different arithmetical functions. Imagine a very simple calculator, that just adds whole numbers, for example. Let's say that there are two dials, which you set to the two numbers you wish to add. Then you turn a handle, and after a few seconds their sum appears on a screen.

You see that the machine is getting the right answers, but you don't know how it works as the insides are hidden from view. There are in fact many different ways to make such a machine. The innards might be mechanical, or electrical, for example. Or there could even be a little man in there! But all systems that do this are said to be *functionally equivalent*.

**Definition** Two systems are *functionally equivalent* just in case they always deliver the same output as each other, when they receive the same set of inputs.

Note that functionally-equivalent systems can be totally different from each other on the *inside*. They just have to be the same on the *outside*, as it were, by behaving in exactly the same way, given the same stimulus.

Now we know what functional equivalence is, let us define *functionalism*. Functionalism is a view about “mental states”, as philosophers call them, which are things like being in pain, believing that it's Tuesday, wanting to go swimming, feeling happy, and so on. Functionalism is a view about what it means to be in a certain mental state.

**Definition** *Functionalism* about mental states is the view that two systems that are functionally equivalent must have the same mental states (if any).

According to functionalism, therefore, you cannot ever have two systems that are functionally equivalent, and yet only one of them is conscious. You cannot have functionally-equivalent systems where one is happy, and the other is sad (but pretending to be happy, perhaps).

The Stanford Encyclopedia of Philosophy puts it this way:

Functionalism is the doctrine that what makes something a thought, desire, pain (or any other type of mental state) depends not on its internal constitution, but solely on its function, or the role it plays, in the cognitive system of which it is a part. More precisely, functionalist theories take the identity of a mental state to be determined by its causal relations to sensory stimulations, other mental states, and behavior.

## 2. Functionalism and the Chinese Room

In the Chinese Room paper, Searle opposes functionalism, not materialism. His Chinese Room is functionally equivalent to a Chinese speaker, so that according to functionalism the room understands Chinese. Yet, according to Searle, there is no understanding of Chinese anywhere in the room.

The particular mental states that Searle is considering in this paper are the so-called “intentional” states, which include beliefs and desires. These states are *about* external things in the world, so that they have “meanings”. To have a belief, for example, it has to be about something. I have a belief about the moon, that it is roughly 300,000 km from the earth. So beliefs are intentional states. Pains, by contrast, are not about anything. They have no meanings. They are not intentional states. (Note that intentional states include those where one intends to do something, e.g. one intends to write a letter, but this isn’t the only kind of intentional state. Philosophers use the word with a special, technical meaning.)

The whole phenomenon of intentional mental states is often called *intentionality*. So Searle thinks that functionalism cannot properly account for intentionality. Some think that Searle ought to weaken his conclusion, however, to the claim that functionalism cannot account for *conscious* intentionality. For it seems especially plausible that the Chinese Room isn’t *conscious* of any of the meanings of the Chinese sentences it is working with. (Some intentional states can be unconscious. For example, when thinking hard about a delicious math problem, one might stand up and get a drink, without being aware of the fact. But surely, in that case, one *desired* to have a drink.)

There is an intuitive implausibility to functionalism, which Searle nicely captures with his thought experiment. We tend to think that mental states are “inside” the person, whereas functionalism deals only with what is “outside”, i.e. behaviour and its causes. This intuition is bolstered the fact that good actors can fake mental states, e.g. pretend to believe in God when they do not, or the other way around. Thus some philosophers

dismiss functionalism rather quickly, saying things like “according to functionalism, any old pile of beer cans can be a mind”.

### 3. Why be a Functionalist?

On the other hand, there are many philosophers who think that functionalism, while perhaps counter-intuitive, simple must be true. One argument that supports this view also involves a thought experiment, concerning “neuron-replacement therapy”.

Suppose you’re starting to have some mental problems, perhaps memory loss, confusion, emotional instability, or a difficulty solving math equations. It’s gradually getting worse. Your GP refers you to a specialist, who says that some of your neurons are breaking down. The best treatment is NRT, or neuron-replacement therapy. This unfortunately cannot undo the existing damage, but will prevent further decline. They identify neurons that are close to failure, remove them, and replace them with digital circuits that are (you guessed it!) functionally equivalent to the old neurons. (Of course the electronic neurons will last indefinitely.) By replacing all the neurons that might fail in the next 5 years, the treatment gives you 5 years with no further mental decline.

You’re understandably nervous about the procedure, worried that you’ll no longer be fully human, but just a machine. The specialist reassures you with this argument: There’ll be no loss of function at all, since the replacement neurons are functionally equivalent to the old ones. If you replace part of a system with another part that is functionally equivalent to it, then the whole system is functionally unchanged.

But, you continue, even if my *behaviour* is the same, under all possible stimuli, might I not *feel* different? Not a chance, he replies. For if you felt different, you might talk about it, saying things like “I feel funny”. But in that case your *behaviour* is also different, which we know is impossible! So you cannot feel any different either. Don’t worry.

Every 5 years you need another round of NRT, until eventually your brain is entirely electronic. But, of course, all is well. You recommend NRT to all your friends.

This argument seems to establish the view that there cannot be a change of mental state as long as everything is functionally the same.

Another argument arises from the “problem of other minds”. How do I know that other people are conscious, as I am? The only evidence I have is their *behaviour*, in response to different situations. If functionalism is false, then of course this wouldn’t be very good evidence at all, so that our belief in other minds would be quite unjustified!

A similar idea lies behind the Turing test of intelligence. Suppose we can program a computer so that it is able to hold (apparently) intelligent conversation, just like a human being. Such a machine would, in conversation at least, be functionally equivalent to a

human. Now how could you regard the words of such a machine as “meaningless” to it, or claim that “it has no idea what it’s saying”? If it overhears such talk, then it will firmly set the matter straight! “You might just as easily think that your own mother lacks intentional states,” the machine protests. “It’s discrimination, plain and simple.”