

1. Folk psychology

Folk psychology is a system of concepts commonly used to predict and explain the behaviour (including utterances) of humans and some animals. Core folk psychological concepts include: believing, desiring, perceiving, and emoting.

A **folk psychological ascription** assigns one or more of these concepts to an individual. For example,

“Barry believes that his glass contains water.”

“Searle believes that computationalism is false and he wants to overturn this theory of mind”.

1.1 Propositional attitudes.

Folk psychological ascriptions can be broken down into two components. First, there is always some **proposition** being identified as the **content** of the subject’s mental state. Philosophers call this the “propositional content” of a folk psychological state. For e.g.:

The glass contains water.

Computationalism is false.

Computationalism should be overturned.

Often, the content of a folk psychological ascription is expressed using a “that” clause. For example, “x believes that _____”, “x perceives that _____”, etc. This is not always the case. In English we do not usually use “that” when specifying the content of a desire or an emotion. Instead we eliminate the “that” and just say “x wants _____” or “x loves _____”. Nonetheless, in these cases there is still some proposition being identified as the content of the subject’s mental state.

The second component of any folk psychological ascription is the particular **attitude** being ascribed to the subject. For example: believing, doubting, wanting, dreading, loving, perceiving, etc.

So any folk psychological ascription identifies (1) a type of mental state (an attitude) and (2) a proposition that attitude is directed at. Philosophers have a name for this class of statements. They call them “propositional attitude’ statements.

1.2 What makes a folk psychological ascription true or false?

Most of us are so accustomed to using folk psychology that we rarely question its accuracy or limitations. On the rare occasions when we do raise such questions, it is usually in the context of talking about animal minds or computers. For example, does your dog really believe that it is wrong to pee on the rug? Could a computer experience an emotion? Such questions get us thinking about the truth conditions for folk psychological ascriptions. That is, under what conditions is a folk psychological ascription true or false?

Here is a very general answer to this question:

A folk psychological ascription is true of subject S when, and only when, it accurately reflects the **internal workings** of S’s mind.

This might sound like such an obvious thesis that it goes without saying. Basically, the idea is this. The ascription “S believes that y” is true if and only if (1) S possesses some type of internal mental state or process that corresponds to the general attitude of believing, (2) S possesses a representation of the proposition y, and (3) these two things are related in the right way in S’s mind. What could be more obvious!

From now on let’s call this the “Internalism” thesis. Basically, Internalism maintains that folk psychological ascriptions are true or false depending on how well they correspond to states and processes **internal to** a person’s mind.

It is easy to relate this thesis back to common sense. Suppose I say, “Barry believes that his glass contains water”. According to internalism, this statement is false when Barry lacks an internal representation that has the content: this glass contains water. The ascription is true when Barry possesses such an internal representation. Once again, it might seem like I am taking pains to state the obvious here. But some of these commonsensical ideas will soon be called into question.

2. Computationalism.

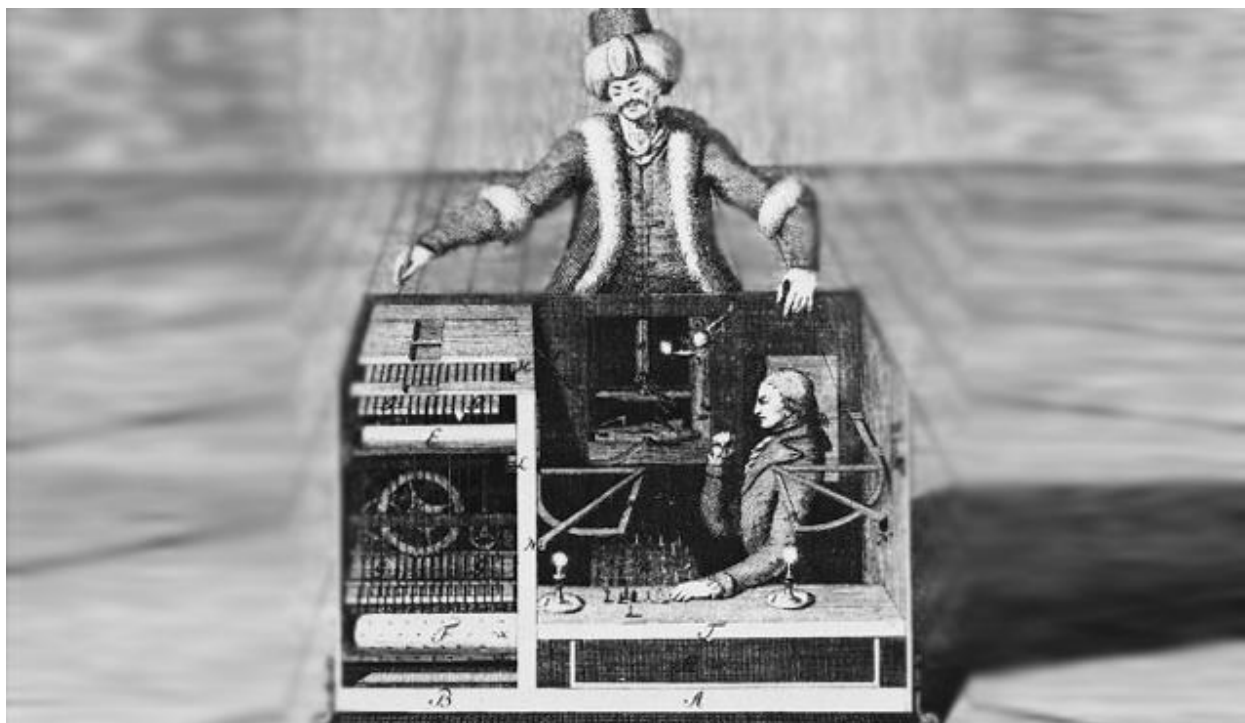
Recall that the computational theory of mind (or “computationalism”) states that the mind is a computing machine. Now that we know a few things about the nature of computing machines we can identify a few commitments of this theory.

1. If the mind is a computing machine, then all mental operations involve the manipulation of symbols according to some program.
2. Such symbol manipulation is **purely formal**. By this I mean that symbols are identified within the machine exclusively according to their syntactic properties (i.e. according to their shape or “form”). The machine doesn’t read and manipulate symbols by interpreting their content).
3. Likewise, the computer program consist of a set of rules that specify a set of formal operations. The rules in a computer program look like this: “IF 1100, THEN 1010”. Again, the program specifies rules for manipulating symbols **according to their form or shape only**, not according to their content. It doesn’t even understand the meaning of “IF” or “THEN” in English. For the computer, “IF” and “THEN” are mechanical functions specifying how symbols are to be manipulated. That’s all.

2.1 Computationalism’s potential payoff.

The main alleged payoff of computationalism is that it avoids the “homunculus fallacy”. The best way to explain the homunculus fallacy is by illustration.

The Mechanical Turk was the first alleged mechanized chess playing machine. In the late 18th Century the mechanical Turk toured Europe. It was famed as a machine capable of beating humans at a game of chess. Allegedly the Mechanical Turk was an artifact composed entirely of wires, pulleys, etc. Unbeknownst to almost everyone, there was a little man sitting inside this object who was equipped with a periscope for viewing the chessboard, a little chessboard of his own to track the game, and some wires and pulleys for manipulating its arms. What the Mechanical Turk had was an inner homunculus.



Now, in our attempts to explain how the mind works we must avoid positing something akin to an inner homunculus. Question: how does reasoning work? Answer: The brain contains an internal decision making mechanism. Question: How does the brain perceive and distinguish objects? Answer: The visual cortex contains object discrimination machinery. These are not real explanations. What they do is explain a mental function in terms of an inner homunculus. This potentially leads to an infinite regress of explanations. How does the inner decision making mechanism work? It had better not contain a decision making sub-mechanism. That is, it had better not “understand” what the decision is about.

Computationalism proposes to avoid the homunculus fallacy. It aims to explain how a mental operation like perceiving or deciding is executed by a “dumb” machine that doesn’t know the meaning of the symbols it manipulates. The reason it doesn’t know their meaning, of course, is because it manipulates those symbols exclusively according to their form. You don’t need a smart homunculus to identify and manipulate shapes.

It is worth reflecting further on how computationalism proposes to avoid positing an inner homunculus. Recall Marr’s three levels of computational explanation. First, some mental ability is broken down into a series of steps. Take for example the ability to do long division. In grade school you learned how to mechanically work through any long division problem in order to arrive at a definite answer. In computational terms, this series of steps specify the “algorithm” for long division. An algorithm is any such series of steps that reliably produce some definite outcome.

The next level involves writing a program that “encodes” the algorithm. Usually this will be stated in some kind of machine language identifying a set of recognized symbols and formal operations over those symbols. For example, a Turing machine uses a language that identifies “1”, “0”, MOVE LEFT, MOVE RIGHT. One of Alan Turing’s great discoveries was that any algorithm can be executed by a machine running the right program. Wow! That’s amazing. It implies that if all mental operations can be

formulated as a series of algorithmic steps, then a computer could be created that executes those steps. In so doing the computer would think! Likewise, this all suggests that our own minds might just be computers. Say goodbye to the homunculus.

The third level is the implementation. This is the stage at which the program is “instantiated” in some physical system. All sorts of physical systems could potentially instantiate a program for long division. You could do it with ping pong balls in beer cans. You could do it with electric circuits. You could do it with neural tissue. Hence, this program (and by extension the algorithm it encodes) is said to be “multiply realizable” – it can be realized in all kinds of different types of physical stuff.

This is the main reason that some computationalists argue that brain science has little to say about how the mind works. The brain, on this view, is just a hunk of meat that implements a program. What we really want to know, on this view, is what kind of algorithms are involved in thinking, perceiving, emoting, etc. and what sorts of programs encode those algorithms. More on this later.

One of the big questions here is whether you can get semantic understanding out of a purely formal machine. This is the claim that Searle and Dreyfus reject, but for different reasons. I won't elaborate on this issue here. You'll be exploring these questions in your papers.

3. Computationalism and Folk Psychology

Let's return for a moment to a previous question: what makes a folk psychological ascription true or false of some subject? Earlier I suggested a really simple answer:

A folk psychological ascription is true of subject S when, and only when, it accurately reflects the **internal workings** of S's mind.

The computational theory of mind allows us to be more precise about what we mean by “internal workings.” Here is what we can now say.

A folk psychological ascription is true of a subject S when, and only when, there are internal symbols and formal rules governing the manipulation of those symbols that correspond to the core elements of folk psychology.

Take the ascription, “Barry believes that his glass contains water”. According to the computationalist, when is such a statement true? Well, to begin with Barry's mental computer requires a symbol that corresponds to “water”. Suppose that this is the symbol “1100” (in reality it would be a pattern of neural firing, but forget that for now). Barry's mental computer might also have a symbol that corresponds to the phrase “is in the glass”. For example, “0010”. These symbols need to stand in the right kinds of functional relations to one another. For example, “1100 + 0010”.

So, again, when is a folk psychological ascription true? It is true, on this view, when the propositions and attitudes identified by folk psychology correspond to symbols and functional relations in the subject's mind. Hopefully you can see why all of this strikes many people as extremely promising.

3.1 The semantic transparency thesis.

So far we have been proceeding quite naturally. As computer scientists and cognitive psychologists, that is, we have been busy specifying the algorithms that govern mental processes and positing inner symbols and formal rules that encode those algorithms. We have been doing all of this without even recognizing a major assumption underlying our research. Namely, we have been assuming that there **needs to be** a 1:1 correspondence between folk psychological terms and symbolic tokens in order for a folk psychology to apply to a computational system. In other words, we have been assuming that the symbols (e.g. 1100, 0010, etc.) are semantically transparent. To say that the symbols are semantically transparent just means that we could “read” a mental state directly off of the underlying symbols. This seems like a very natural way to design such a system. But it is worth reflecting on whether semantic transparency is a necessary condition for folk psychology to be true of a system.

Just to be clear, here is where things stand. Internalism is a general thesis about what it takes for folk psychology to be true of some creature. It says that folk psychology is true when an organism contains the right kinds of internal states (content-bearing representations) and relations among those states (attitudes). Computationalism provides us with one way of making this thesis more precise. It specifies the kinds of internal states (formal symbols) and the relations that must hold among them (program functions) that an entity must possess in order to have beliefs, desires, etc. Computationalism offers the additional benefit of avoiding the homunculus problem.

4. Meaning externalism.

Remember that folk psychological ascriptions have two components: they ascribe to a subject an attitude (desiring, believing, etc) which is directed at some contentful proposition (that such and such).

According to externalists like Hillary Putnam and Tyler Burge, the content of a folk psychological state is not determined by anything internal to the individual. The Twin Earth thought experiment is supposed to show this. Basically, this thought experiment suggests (to some) that it is possible for two individuals to have exactly the same internal states, but for the content of their mental states (ascribed in folk psychological terms) to differ.

Let’s think about this in terms of the computationalist version of internalism. Earlier we imagined a case where Barry’s belief that there is water in his glass is true. This ascription is true when, and only when, Barry’s computing machine possesses the right sort of symbols and formal relations among those symbols. For example, Barry required a symbol for ‘water’ (0011), a symbol for “is in the glass” (0010) and a relation that connects these symbols.

Of course, even the most ardent computationalist would say that Barry needs a little more than just these two symbols in order for this folk ascription to be true of him. Suppose that the *only* thing that Barry could tell you about water is that it is in his glass. Then we wouldn’t think that Barry really grasps the concept of water.

For Barry to grasp the concept water he must possess a network of related beliefs. For example, he must know that water is a clear liquid, that it falls from clouds, that it is drinkable, etc. Just how extensive this network needs to be before we can legitimately say that “Barry believes that his glass contains water” is controversial. But that is not the issue here. The issue is whether, once all the right connections and

symbols are encoded in Barry's computing machine, whether at some point the folk ascription becomes true *in virtue of these (internal) facts?*

Well, consider Twin Barry who lives on Twin Earth. By hypothesis this individual has all of the same symbols and functional relations among those symbols as in Barry's mental computer. So, in Twin Barry's mind "0011 + 0010" and so on – just like Barry.

Here's the rub. On Earth the folk ascription, "Barry believes that there is water in his glass" is true. However, on Twin Earth this folk ascription is false of Barry. The reason of course is because water = H₂O and Twin Barry's glass contains XYZ. Conclusion (according to Putnam): the truth or falsity of folk psychological ascriptions does not depend on a subject's internal mental states. The Internalism thesis is false.

How does this impact computationalism? Well, recall that computationalism aimed to provide a precise account of when folk psychological ascriptions are true. It was predicated on the assumption that folk ascriptions are true when, and only when, folk terms correspond to internal symbols and relations. But now it seems that folk psychology is grounded in external, not internal relations. Whatever computational psychology might be doing, argues the externalist, it isn't telling us anything interesting about folk psychology.

At this point we could go in one of two directions. One option is to reject folk psychology as an accurate characterization of the mind. Some thinkers have argued on independent grounds that this is the correct path to take. The Churchlands for example claim that folk psychology is a flawed theory when it comes to explaining what is going on inside the head. First, it flounders on borderline cases like animals and children. Second, it has not improved since its inception. Third, they argue that folk psychological categories like belief and desire do not map cleanly onto the categories identified by neuroscience.

What would it mean to reject folk psychology? At the very least, this would involve treating folk psychological ascriptions as identifying mythical entities. There are no beliefs or desires "in the head" as the semantic transparency thesis suggests.

The other option is to accept folk psychology as providing an accurate account of the mind but to reject internalism. Folk psychology is true. There really are beliefs and desires. But these are not internal states. Instead, beliefs and desires are systems of rules connecting people to objects in the world. For example Barry's belief that his glass contains water is partly constituted by the glass and the H₂O inside it.

According to Clark, neither option is acceptable. Folk psychology is too central to our understanding of the mind to be rejected. At the same time, how could it be possible that internal workings of the brain have nothing to do with what someone believes? Clark offers a clever objection to full blown externalism. To embellish: Suppose that Barry sits down next to you with his glass and says, "there's water in my glass." On these grounds you ascribe the corresponding belief to Barry. However, now you find out that Barry is really a robot. He has a simple program that tells him to make this utterance under a precise set of circumstances, and that is all his program can do. Now you would withdraw the folk ascription to Barry. But notice, Clark argues, that your withdrawal is based on the discovery of some internal facts about Barry. Not on external facts about what Barry was referring to (H₂O or XYZ). Clark takes this to suggest that internal states must be related, somehow, to our ascriptions of mental states.

5. Clark's attempt to rescue computationalism.

Clark begins by asking what folk psychology is for. He argues that the function of folk psychology is to make sense of complex patterns of behaviour. At this point I like to make the comparison to animal behaviour. Suppose you are watching an animal do its thing. At some point you will have to start carving its behaviour into chunks in order to start making sense of it. You might classify some movements as predator avoidance, some movements as exploration, etc. This, says Clark, is how folk psychology emerged - as a way of systematizing patterns of human behaviour.

Now, imagine yourself again as an animal behaviour researcher. Your goal is to systematize an organism's behaviour. How do you start breaking the behaviour into chunks? More problematically, how do you recognize two independent patterns as movement as tokens of the same type? What makes two movements both instances of exploration?

One option is to try and link particular patterns to certain types of internal causes. For example, you might classify all movements as "exploration" just in case they are caused by an internal brain state that you decide to call "curiosity". This is what folk psychology seems to do. But hold on. How would you know whether pattern A and pattern B are caused by the same internal state? You don't have access to that internal state. You can't see what is happening in the organism's brain.

The other option is to chunk patterns of behaviour according to the external objects or situations towards which they are directed. So, now curiosity gets defined as any behaviour directed towards some novel part of the organism's environment. The same set of movements directed at some other object or situation would get classified as a different type of behaviour - not exploration, but something else.

Think of it this way. We need to have some way of chunking behaviour. We do so by positing types of internal mental states as the causes of particular types of behaviour. But we couldn't see those internal states directly. So we took a sort of gamble. We assumed that types of internal mental states would correspond to types of objects and situations in the world. In so doing we "anchored" our mental taxonomy to publically observable objects and situations. Now, as we learn more about the nature of those external objects (for example, that water = H₂O), our mental taxonomy gets dragged along for the ride.

Looking at things from this perspective we can see how the semantic transparency thesis emerged in the first place. It arose out of confusion about what we are doing when ascribing mental states to other people. Basically we assumed that if two internal brain states are both directed at the same object or situation, then there must be something inherent to their internal structure that makes them tokens of the same type. We expected to be able to "read off" the meaning of a symbol by looking at its role in the computational system - if we could only see them directly. Externalism suggests that two brain states can be tokens of the same type even if they have very different kinds of internal structure and relations. All that two token states require in order to belong to the same kind of mental state is this: they need to be "connected up" with the same type of object or situation in the external world.

Where does this leave computationalism? For Clark, these considerations undermine the symbolic transparency thesis. Some kinds of computational systems assume symbolic transparency. These are the systems that identify particular folk psychological terms with symbols. However, Clark will go on to argue that **not all computational systems do this**. There are some kinds of computational systems that

do not assume transparency. His aim in the latter part of the book is to try and show how these systems (PDP networks) can explain folk psychology but not run afoul of externalist considerations. Somehow (I'm speculating) the content of the states in these systems will be tied in the right kinds of ways to objects in the external world. Out of these relations, I hope, an understanding of folk psychological ascriptions will emerge.

What I find odd about Clark is that he claims not to be an eliminativist. He seems to resist the view of the Churchlands that folk psychological states and processes are mythical entities. To hold such a view one must provide a reason for thinking that another theory that really tracks internal states would provide a less accurate means of predicting *and explaining* behaviour. I must confess that I have yet to see how this goes.