Phil 100D: Philosophy of Mind Professor Aaron Zimmerman

Handout #4: Analytic Functionalism

1. Some ontological categories: object, properties, propositions, states.

Objects: An object is the kind of thing that has properties. (Physical objects occupy portions of space for periods of time. Abstract objects do not.) An object is the denotation or referent of a singular term: e.g., a name, or a demonstrative expression like 'that'.

Properties: A property is the kind of thing that objects have. Properties are instantiated by objects. An object's properties account for its causal powers (i.e. its powers to affect and be affected by things or to enter into causal relations). Properties are the things expressed by predicates or predicative expressions (e.g. 'is red' or 'is spherical'), gerundive expressions ('being red', 'being spherical'), and they are the denotations of predicate-nominals (e.g. 'justice', 'sphericality').

Propositions: A proposition is something that can be true or false. It is the kind of thing we believe when we believe something. It is the kind of thing we entertain when we think of something. It is the kind of thing we assert when we assert something. A proposition is what is expressed by a grammatical declarative sentence.

States: A state is a complex non-truth-evaluable entity; it is the having of a property by an object for some duration. So, for example, solidity is a physical state consisting of an object's molecules being structured in a particular way. Solidity just is the having of certain structural properties for a period of time. When an object or substance goes from the solid state to the liquid state it changes its properties: it goes from being solid to being liquid.

> Being in pain, believing that 2+2=4, and feeling sad are all *mental properties*.
>
> The state of a person's mind when she is in pain is a particular *mental state* as are the different states her mind is in when she concludes that 2+2=4 or feels sad.

2. More Metaphysical Distinctions

Essential Property: (a) For any object x and property P: x has P necessarily if and only if x exists, x has P, and x cannot exist without having P; (b) The properties an object has necessarily are that object's ***essential*** properties.

If it is possible for x to exist without having P, having P is not necessary for being x and P is not one of x's essential properties.

Individuating Property: (a) For any object x and property P: P suffices for being x iff x has P and for all objects y: if y has P, then y=x; (b) Those properties that suffice for being a particular object are those object's ***individuating*** properties.

If an object y has P without being x, then even if x has P, having P does not suffice for being x and P is not one of x's individuating properties.

A sufficient condition for intrinsicality: For any object x and property P: P is an intrinsic property of x *if* for any possible object y: if y is an exact qualitative duplicate of x then y has P.

Thus, if x has a certain property Q that is had by all of the qualitative duplicates of x that are possible, then Q is an intrinsic property of x.

Note, however, that the condition above is sufficient but not obviously necessary because of arguably intrinsic properties of an object not shared by duplicates of it. A property G that is not possessed by all of x's possible intrinsic duplicates might still be an intrinsic property of x's. How? Well, this depends on the nature of properties but suppose we allow that G might be the property: being x. Intuitively, being x is intrinsic to x, (being you is intrinsic to you) but being x is not a property had by anything other than x. (Maybe we should say there are no properties of the sort here described as "being x." If so, how do we adjudicate different proposals concerning which gerundive expressions denote properties and which do not?)

Haecceity: For any object o: o's haecceitic properties include its being o and all those properties of its internal make-up that entail its being o (if there are any such properties).

Example: Origins? Being the product of a certain sperm-egg interaction?

Intrinsic Properties An object's intrinsic properties are its haecceitic properties and all of those non-haecceitic properties it shares with all of its qualitative duplicates.

Examples: Shape, size, surface color?

Dispositional Properties: For an object x and property P, P is a dispositional property of x if and only if x's having P *is identical to* x's being disposed to behave in ways $w_1..w_n$ given certain conditions $c_1 \ldots c_n$.

Examples: Solubility, fragility, toxicity.

Causal Powers: For any object x, intrinsic property P, and dispositional properties D: D are the *causal powers* bestowed on x by P if and only if x has D in virtue of having P.

Note that Armstrong countenances both "active" and "passive" powers:

Examples: (1) **Active Powers**: A poison will cause an organism to fall ill or die when it is ingested. The poison's disposition to impair or destroy life when ingested is a causal power of it: it has the power to kill in this way. (It is toxic.) Poisons have this causal power in virtue of their microstructures. Its micro-structural properties (i.e. its being composed of molecules that impair or destroy the essential metabolic processes of those lifeforms for which it is poisonous) are intrinsic properties of the poison that bestow this dispositional property/causal power (i.e. toxicity) upon it. The intrinsic properties of the poison explain why it will kill those animals that ingest it (or they "make it the case" that these animals will suffer or die should they ingest it). (2) **Passive Powers**: A fragile bowl will shatter when struck. Its disposition to shatter when struck is a causal power of it: it has the power to shatter. It has this causal power in virtue of its microstructure. Its micro-structural properties (i.e. its having molecules that are arranged and bonded to one another in a certain way) are intrinsic properties of the bowl that bestow this dispositional property/causal power upon it. They explain why it will shatter when struck or "make it the case" that it will shatter when struck.

Now recall our definition of causal powers: For any object x, intrinsic property P, and dispositional properties D: D are the causal powers bestowed on x by P if and only if x has D in virtue of having P.

The Realization of Powers or Causal Roles: When x has certain causal powers or dispositional properties D1…Dn in virtue of having a set of intrinsic properties P1…Pn, we can say that X's intrinsic state at t (i.e. x's having P1…Pn at t) **realizes** its causal powers D1…Dn at t.

We can also speak of the intrinsic state of o in question as the **realizer** of the causal powers in question.

3. Functional Accounts of Mental States

A functional property P is a property for which there exists a dispositional profile (or set of dispositions) D such that for any x: x's having D is necessary and sufficient for x's having P.

A functional state is the state of some object's having some functional property.

Being a carburetor is a functional property. Nothing *can* be a carburetor and yet fail to be so disposed that it feeds an engine an explosive mixture of vaporized fuel and air when properly activated. Moreover, anything that is disposed to feed an engine an explosive mixture of vaporized fuel and air when properly activated is a carburetor.

Importantly, the matter out of which an object is constituted doesn't directly impact whether or not that object is a carburetor. Being a carburetor consists in functioning in a certain way, not in being made of a certain kind of stuff. You classify being a heart as a similarly functional property so long as you think an artificial heart is a heart.

Functionalist conceptions of a heart say that whatever circulates blood in a certain way is a heart.

Type-Functionalism: **Types of mental states are identical with types of functional states**.

Example schema 1: being pain=being disposed to produce behaviors of type B1…Bn and mental states of types M1…Mn when presented with stimuli of types of T1…Tn given that one is in mental states of types S1…Sn.

Example schema 2: Believing that 2+2=4 =being disposed to produce behaviors of type B1…Bn and mental states of types M1…Mn when presented with stimuli of types of T1…Tn given that one is in mental states of types S1…Sn.

> *Question*: How might we fill these schemata out?

Token-functionalism: Mental state tokens are identical with token functional states.

> *For example*: John's pain at t=the state of John's being disposed at t to produce behaviors of type B1…Bn and mental states of types M1…Mn given that he has been presented with stimuli of types of T1…Tn and is in mental states of S1…Sn.

4. Kinds of Functionalism

In "Psychophysical and Theoretical Identifications," David Lewis connects causal roles or dispositional profiles with the significance of **theoretical terms**. According to Lewis, psychological terms like "belief," "desire," "pain," and so on are used to advance theoretical conjectures which are defined in terms of the stories we tell when explaining our actions and reactions to each other.

> Think of common-sense psychology as term-introducing scientific theory, though one invented long before there was any such institution as professional science (Lewis, in Chalmers (ed.) 2002, 92).

To help us get a handle on the structure of this "theory," Lewis compares our familiar psychological expressions with variables introduced to label three perpetrators of a crime who are initially identified or described through the details of the crime, where their existence is supposed as the best explanation of what is known to have happened. The detective is fairly confident that three people committed the crime and that each played a certain role in its commission but he is unsure (or not willing to share) which three people accomplished the deed.

On Lewis' view the detective's theoretical terms pick out the **causal roles** played by the three characters in the scenario and the three perpetrators themselves realize these causal roles. Similarly, Lewis says that the folk psychological language we use when explaining ourselves to one another pick out the causal roles played by states of our mind/brain (i.e.

our mental states) and certain intrinsic states of our minds/brains (i.e. our neural states) realize these causal roles. When you know you're in pain, you are aware that there is something connected to a given region in your body causing you to flinch, shout or nurse the area. "Pain" is defined as "that sort of thing, whatever it is, that is playing this causal role." When you learn that (say) c-fibers are firing in your nervous system whenever you're in a state of mind that is defined in this way via its folk psychological causal role, you learn that c-fibers firing **realize** these pains.

> When we learn what sort of states occupy those causal roles definitive of the mental states, we will learn what states the mental states are—exactly as we found out who X was when we found out that Plum was the main who occupied a certain role, and exactly as we found out what light was when we found that electromagnetic radiation was the phenomenon that occupied a certain role (Lewis, in Chalmers (ed.) 2002, 92).

Questions: Is folk psychology a theory? Do you assume the truth of a theory when describing yourself as experiencing pain or wanting something to drink? If the concepts we use to describe each other in these ways do constitute a theory, how does this theory resemble and how does it differ from theories introduced by scientists or academics? Lewis admits that common sense psychological did not originate in a theory of any kind, but he calls the idea that folk psychological discourse is theoretic a "good myth" (p. 92). Is it a good myth?

Analytic truth: a sentence s is analytically true if and only if s expresses a truth in virtue of the meanings, concepts or intensions of its constitutive expressions alone and their mode of combination in s. If s is analytically true it is impossible for s to express a falsehood so long as it retains its customary meaning.

> Putative Examples: 'If there is a strongest human, then she is stronger than every other human', 'Vixens are foxes', 'Bachelors are unmarried'.

Where Armstrong writes of analyses of psychological terms, Lewis writes of platitudes like "Toothache is a kind of pain" and writes of definitions of mental terms couched in terms of these platitudes.

Synthetic truth: (Negative characterization): A sentence s is a synthetic truth if and only if s expresses a truth but s is not analytically true. (Positive characterization): A synthetic sentence is one in which the concept associated with the predicate is added to the concept that delimits the sentence's subject. If s is synthetically true it is possible for s to express a falsehood while retaining its current customary meaning.

> Examples: 'I am teaching', 'Grass is green', 'Biden is president'.

A priori knowledge: A proposition p is known a priori by a subject S if and only if S knows p and S's reasons for believing p do not include facts about S's sense experience.

Examples: A mathematician's knowledge that 2+2=4. Your knowledge that all bachelors are unmarried.

A posteriori knowledge: A proposition p is known a posteriori by a subject S if and only if S knows p and S's reasons for believing p do include facts about S's sense experience.

Examples: Your knowledge that Biden is president. Your knowledge of the weather.

Questions: Is 'Samuel Clemens=Samuel Clemens' analytic or synthetic? 'Is Samuel Clemens=Mark Twain' analytic or synthetic? (Surely, the former is a "truism" while the other is or can be an interesting biographical fact.) Is John's knowledge that (if he exists) Samuel Clemens is Samuel Clemens a priori or a posteriori? Is John's knowledge that (if he exists) Samuel Clemens is Mark Twain a priori or a posteriori?

Analytic or A Priori Functionalism: Mental concepts are functional concepts. The concept a person has in mind in virtue of understanding 'belief', 'desire', 'pain', and 'anger' are best "unpacked" or analyzed as concepts given or expressed by expressions of the form: 'being disposed to produce behaviors of type B1…Bn and mental states of types M1…Mn when presented with stimuli of types of T1…Tn given that one is in mental states of types S1…Sn.'

(*) Believing that it is raining outside just is being disposed to take an umbrella when it looks to one as though it is raining outside or someone tells you that it is raining outside. . .and one wants to stay dry or one wants to help someone else stay dry . . . ..'

(**) Wanting to stay dry just is being disposed to take an umbrella when it looks to one as though it is raining outside or someone tells you that it is raining outside. . .and one believes that things are as they seem or believes that things are as one has been told they are, and believes that umbrellas help one stay dry and . . .

The analytic functionalist thinks equations like (*) and (**) are true and analytic: i.e. true in virtue of the meanings of the terms involved. They also think that (*) and (**) are a priori knowable—they can be shown to be true by philosophers doing conceptual analysis. The construction of experiments and the observation of the results of those experiments is not necessary. In this sense (*) and (**) are definitional truths we can use to interpret the results of experiments rather than hypotheses confirmed through experimentation.

Synthetic or A Posteriori Functionalism: There are true equations identifying mental properties with functional properties, but they cannot be known a priori. (They are not like the truisms (*) and (**) described above.) Instead, we must rely on theories and our observations of the results of experiments performed by cognitive scientists to discover these truths.

Armstrong and Lewis are analytic functionalists. See Armstrong's attempts to analyze the concept of a "purpose" and the integrated analyses of "perception" and "belief" that he provides to accomplish this task. (Lewis' analyses are even more partial or schematic.)

5. The Functionalist Response to the Property Dualism Argument.

(a) Water=H20.

The identity theorist (like Smart) claims that (a) is true but that it is not analytic and the proposition it expresses is only knowable a posteriori.

How can this be? Frege offers an answer: The meaning of 'H_2O' and the meaning of 'water' do not simply consist in the substances these expressions pick out.

So what is the meaning of (a)? The descriptivist has an answer. It's something like the proposition expressed by the following sentence:

(b) The clear, drinkable, exceedingly common liquid that fills the lakes, and streams around here=the stuff composed of molecules of two Hydrogen atoms and one Oxygen atom.

So we have ***two*** (higher-order) properties: (*) *being a clear, drinkable, exceedingly common liquid filling the lakes and streams around here*, and (**) *being composed of molecules of two Hydrogen atoms and one Oxygen atom*, had by ***one*** substance: water/H_2O.

In this case, the two different properties provide ***two different ways of knowing*** the one substance. The proposition is synthetic because these two *different* properties are the meanings of the relevant expressions (or are determined by the meanings of the relevant expressions). The proposition is only knowable a posteriori because experiments are needed to provide one with a reason to believe that these are two properties of one and the same substance.

But now consider:

(c) Pain=Neural Events of type T.

Clearly, this is not analytic and the proposition it expresses cannot be known a priori. How can this be? To answer this question to the satisfaction of those tempted to reject physicalism, the Fregean descriptivist needs to provide an identity that captures the meaning of (c). The appropriate statement would look something like this:

(d) This type of experience=the activation to level L of neural nets of type N in anatomic structures A in the human brain and central nervous system.

(where 'this type of experience' is used by a person in pain to pick out the sensation she is experiencing while in pain).

So what are the two properties by which we know of pain? *Being activation to level L of neural nets of type N in anatomic structures A* is clearly a physical property the realization of which can only be known via observation of the human nervous system by those who know the relevant theories of cognitive neuroscience. But ***being this type of experience*** [when this phrase is uttered by someone in pain focusing on or attending to the pain they are experiencing] looks to be a mental property of the pain in question. Thus, while Smart's Fregean view seems to allow him to avoid **object dualism** it seems to force him into **property dualism**.

Armstrong's response: Replace (d) with (e):

(e) being disposed to produce behaviors of type B1…Bn and mental states of types M1…Mn when presented with stimuli of types of T1…Tn given that one is in mental states of types S1…Sn= the activation to level L of neural nets of type N in anatomic structures A in the brain and central nervous system.

And so long as our ordinary concepts of M1…Mn and our ordinary concepts of S1…Sn can also be given functional analyses (a big "if" to be sure) we can avoid property dualism. For instance, *being disposed to produce behaviors of type B1…Bn and mental states of types M1…Mn when presented with stimuli of types of T1…Tn given that one is in mental states of types S1…Sn* is not a mental property (so long as the other Ms (i.e. metal properties) mentioned in the analysis can be analyzed in similar terms). Instead, this descriptive gerundive phrase picks out a functional or ***topic neutral property***. In this way, Armstrong and Lewis try to escape the property dualism argument.

So this is Armstrong's view: When suitably filled out, 'pain= being disposed to produce behaviors of type B1…Bn and mental states of types M1…Mn when presented with stimuli of types of T1…Tn given that one is in mental states of types S1…Sn' is ***analytic*** or conceptual or truistic.[1]

Our knowledge that pain=being disposed to produce behaviors of type B1…Bn and mental states of types M1…Mn when presented with stimuli of types of T1…Tn given that one is in mental states of types S1…Sn is ***a priori***.

'Pain=the activation to level L of neural nets of type N in anatomic structures A in the brain and central nervous system' is ***synthetic*** or substantive or nontrivial.

---

[1] Though Armstrong thinks the statement is analytic or knowable through conceptual analyses, he admits that the two expressions in question are not synonymous as they clearly differ in meaning along some dimensions. What relation must hold between two non-synonymous expressions A and B if A is to provide an adequate analysis of B? Armstrong says he is tempted to an account that distinguishes concepts from word meanings and identifies them with the components of thoughts instead, where language and thought are less closely connected than many suppose. But beyond that (and a nod to the notion of "deep structure") he just says, "I wish I were able to take this topic further" (Chalmers (ed.), 2002, 85).

Our knowledge that pain=the activation to level L of neural nets of type N in anatomic structures A in the brain and central nervous system is ***a posteriori***.

Question: How plausible is Armstrong's claim that 'pain= being disposed to produce behaviors of type B1…Bn and mental states of types M1…Mn when presented with stimuli of types of T1…Tn given that one is in mental states of types S1…Sn' is analytic and the proposition it expresses a priori knowable? Does Putnam effectively refute this variety of analytic functionalism with his Super Spartans counter-example? Are Super Spartans a coherent idea?