

**Handout #2: JJC Smart, “Sensations and Brain Processes”**

**I. The Fregean Response to Descartes**

Frege’s Thesis I: The Distinction between *Sense* and *Reference*: Each expression has at least two meanings or semantic values. For example, a proper name has a **sense**—i.e. some concept it **expresses**—and a **referent**—i.e. some object it **denotes**. An expression’s sense *determines* its reference in the following respect: if two expressions have the same sense, they denote the same object or are true of the same range of phenomena. But two words can have different senses and the same referent. For example, one might argue that we associate different senses or concepts with ‘Samuel Clemens’ and ‘Mark Twain’ even though they denote one and the same person.

Frege’s Thesis II: The (First-Level) Reference-Shift Principle: When a word appears in a propositional clause (or ‘that’-clause) following a psychological verb (like ‘believes’, ‘doubts’, and ‘hopes’) the word does not refer to the thing to which it ordinarily refers, but instead refers to the sense (or concept) we ordinarily associate with it.

Consider now Descartes’ Argument for Object Dualism:

- (1) Descartes doubts that his brain exists.
- (2) Descartes does not doubt that his mind exists.
- (3) Descartes’ mind is not his brain.

If Frege’s theories are true this argument is not valid. Let P3= the proposition that Descartes’ brain exists. Let P4= the proposition that Descartes’ mind exists. P3 is composed of the customary *sense* of the expression ‘Descartes’ brain’ and the customary sense of the predicate ‘exists’. P4 is composed of the customary sense of the expression ‘Descartes’ mind’ and the predicate ‘exists’. Since the customary sense of ‘Descartes’ brain’ is not identical to the customary sense of ‘Descartes’ mind’ P3 is not identical to P4. Thus, even if Descartes’ mind is in fact identical to Descartes’ brain, and ‘Descartes’ mind’ *customarily* refers to the same thing as ‘Descartes’ brain’ we cannot infer the truth of ‘Descartes doubts that his mind exists’ from the truth of ‘Descartes doubts that his brain exists’ by substituting one expression: ‘Descartes’ mind’, for another ‘Descartes brain’ with which it is *customarily* co-referential. Though these expressions are customarily co-referential, in **psychological contexts** like those created by ‘doubts’, they are **not** co-referential. In psychological contexts these words switch their reference. Whereas they ordinarily refer to Descartes’ mind and brain respectively, when prefixed by ‘doubts’ they refer to their senses: i.e. the concepts we customarily associate with ‘Descartes’ mind’ and ‘Descartes brain’.

**II. Smart’s Thesis**

**The Token-Type Distinction:** How many words are written on the next line?

happy, happy

Answer: two tokens of one type. (On the most natural account, a word is a type.)

**Token-Token Identity Thesis:** Each particular mental state token is identical with a neurological state token.

Putative Examples: Suppose John has a searing headache at some point in time  $t$ . Consider the electrochemical state of John's central nervous system (or some part of it) at  $t$ . Call this particular (never to reoccur) state 'NT0'. The Token-Identity Theorist Claims that John's pain at  $t$ =neurological state token NT0.

Suppose that after looking at a bright light, John experiences a redish-orange after-image at  $t+1$ . Let 'NT1' name the electrochemical state of John's visual cortex (or some part of it) at  $t+1$ . The Token-Identity Theorist Claims that John's experience of a redish-orange after-image at  $t+1$ =NT1.

An ordinary model for token-token identity: Samuel Clemens=Mark Twain. (Or, for any time  $t$  at which Samuel Clemens exists: Samuel Clemens=Mark Twain.)

**Type-Type Identity Theory:** Each particular type (or kind) of mental state is identical with a type (or kind) of neurological state.

Putative Examples: Pain=C-Fibers Firing. Pain=Cortical Activation of type T to level L. Experiencing a Redish-Orange Afterimage=Activation of the Visual Cortex in pattern P to degree D. Wanting to help an injured animal=Activation of Prefrontal Cortex to intensity I in manner M.

Ordinary examples of type-type identity(?): Water=H<sub>2</sub>O. Lightning=Electrical discharge. Heat=Mean Molecular Energy. Light=A Photon Stream.

### III. Distinguishing Identity-Theory from Eliminativism

**Eliminativism:** there are no mental states.

The eliminativist typically allows that there are brains, neurons, and electrochemical events that involve neurons and take place in brains. But, the eliminativist claims, these events are *not mental*; they are *merely* physical (i.e. physical but not mental). The eliminativist claims that there are no things over and above these physical things, and no events "over and above" these physical events. Therefore, there are no mental things and there are no mental events.

Eliminative Expressivism: When a subject says "I am in pain" she does not say something true, and she does not say something false. What she says is not "evaluable" for truth and falsity. Her utterance is akin to an exclamation like "Ouch!" in this respect.

The subject does not express her belief that she is in pain (just as a sudden instinctual exclamation of “Ouch!” needn’t express one’s belief that one is in pain). And the subject does not directly express her pain (since there are no such things). Instead, the subject in question manifests a certain *behavioral tendency*. A tendency to say “I am in pain” when her body is injured and/or her nervous system is stimulated in a certain way.

Eliminative Nihilism: When a subject says “I am in pain,” what she says is true if and only if she is in a certain mental state: namely, pain. Since there are no pains (only neurons and their activity) what a subject says when she says “I am in pain,” is always something false.

If the Identity Theory is true then: “John has a pain” is true iff John’s c-fibers are firing. John’s c-fibers sometimes do fire, therefore there are occasions on which “John has a pain” is true. If the Eliminativist is right “John has a pain” is never true because there are no pains. Thus, the Identity Theory is not identical with Eliminativism.

**Sub-points:** (1) Causation is not Identity. (2) Correlation is not Identity. Smart says, “You cannot correlate something with itself. You correlate footprints with burglars, but not Bill Sikes the burglar with Bill Sikes the burglar.”

#### **IV. Arguments for Identity Theory**

Smart says, “That everything should be explicable in terms of physics (together of course with descriptions of the ways in which the parts are put together—roughly; biology is to physics as radio-engineering is to electro-magnetism) except the occurrence of sensations seems to me frankly unbelievable. . . I cannot believe that the ultimate laws of nature could relate simple constituents to configurations consisting of billions of neurons (and goodness knows how many billions, billions of ultimate particles) all put together for all the world as though their main purpose in life was to be a negative feedback mechanism of a complicated sort. Such ultimate laws would be like nothing so far known to science. They have a queer “smell” to them.” Smart then admits, “The above is largely a confession of faith.” But, anyway, his project in this essay is more modest. He’s not trying to (convincingly) argue for physicalism or the identity theses. He’s just trying to defend these identity theses from various objections.

#### **V. Some of the Objections**

##### Objection 1. A Cartesian Argument

I can know that I am in pain without knowing that my brain is in neural state N. So my pain is not identical with neural state N.

Smart gives Frege’s Answer: ‘This pain=NT0’ may be true without being analytically true. Just as ‘Samuel Clemens=Mark Twain’ is true but informative to some people. There are non-trivial identities that are hard to know. The same

goes for 'Experiencing a Redish-Orange Afterimage=Activation of the Visual Cortex in pattern P to degree D'.

Note that Smart says the identities are *contingent*. But the fact that these sentences are not analytic, and the fact that the propositions they express are not a priori knowable, doesn't immediately entail that these propositions are contingent truths. Nevertheless, if a simple descriptivist version of Frege's theory is right, the apparent identity statements in question will be contingent. Suppose, for example, 'My current pain=NT0' is synonymous with "There is a unique x, such that x feels like this to me now, and there is a unique y, such that y is the current electrochemical state of my central nervous system, and x=y." The identity *part* of that statement is necessarily true if true at all. But the statement as a whole is contingent as it may be a contingent matter that the y in question is the current electrochemical state of my central nervous system. One might say that there is a "possible world" in which my central nervous system is in a state right now (e.g. z) that differs from its actual state (i.e. y). After all, I could have eaten something different for breakfast and then my brain/mind would be in a slightly different state than it (actually) is in right now.

### Objection 3: An Argument for Property Dualism

(4) This glass of water=this glass of H<sub>2</sub>O.

How can this be true and yet informative given that

(5) This glass of water=this glass of water

is trivial? After all, given that water just is H<sub>2</sub>O, (4) is true in exactly those (actual and possible) circumstances in which (5) is true. ***The Descriptivist Fregean Answer:*** (4) means the same as (6), and (5) means the same as (7).

(6) The glass of clear, drinkable, exceedingly common liquid that I see before me=the glass of stuff composed of molecules of two Hydrogen atoms and one Oxygen atom that I see before me.

(7) The glass of clear, drinkable, exceedingly common liquid that I see before me=the glass of clear, drinkable, exceedingly common liquid that I see before me.

But now consider:

(8) This pain=NT0.

The sense of 'NT0' is given by a complicated description couched in the vocabulary of cognitive neuroscience (something like "these c-fibers firing"). The sense of 'this pain' should be given by a description too. But, the worry many philosophers have, is that the sense by which we pick out pain-tokens will have to express or correspond to a *purely mental* property.

## Smart's Reply to the Property Dualism Argument

The meaning of the type identity statement identifying pain as a type of thing with some neurological kind would look something like this:

(9) The sensations animals have when their bodies are damaged and they are awake and attending to their bodies = NT0.

Where, again, "NT0" is shorthand for "Nervous state Token 0," which is standing in for a sophisticated neurochemical description of a neurological event of the sort neuroscientists will identify with pain sensations (according to Smart). (Something like "c-fibers firing.")

And the meaning of the token identity statement (8) above is therefore equivalent to the meaning of something like (10) below:

(10) This sensation is like what is going on when I am awake, attending to my body, and my bodily tissues are damaged and it (i.e. that sensation) just is NTO.

Smart's idea is that our introspective awareness of pain can be enabled by what he calls "topic neutral" properties of pain. The topic neutral property here is the role pain plays in making an animal aware of bodily damage (where this is assumed to be the main functional role of bodily pain). Neuroscientists use the neurological and electrochemical properties of pain to identify it and treat it pharmacologically. The person who is experiencing the pain does not use these neurological properties of pain to describe how she is feeling. But, according to Smart, she needn't use purely psychological or non-physical, qualitative properties—**so called "qualia"**—for this purpose. Instead, Smart proposes that the subjective identification and description of pain is enabled by the functional properties of pain – properties which can be characterized in "topic-neutral" language: language which is neither neurological nor psychological in character.

But Smart admits that there is something a little weird about this. "This [proposal] does not mean that sensations do not have properties, for if they are brain-processes they certainly have properties. It only means that in speaking of them as being like or unlike one another we need not know or mention these properties...The strength of my reply depends on the possibility of our being able to report that one thing is like another without being able to state the respect in which it is like. I am not sure whether this is so or not..." (Smart, 64).

Can we judge that a pain (which we experience without bodily damage as with phantom limb pain) *is like* the sensation we experience when our bodies really are damaged, without being aware of some property (i.e. the way each feels qualitatively) in virtue of which these experiences are like one another? If this makes no sense, how can Smart effectively reply to the property dualist?

## A Common Road to Qualia:

**Thesis A: Introspective Resemblances between Perception and Mere Sensation:** If the sensory component of a perception A and a mere sensation B are the same type of sensation (e.g. both of them are bodily pains or both of them are visual experiences of reddish-orange), then A must resemble B in some introspectable respect R.

So if the visual sensory component of your perception of a real reddish-orange chair is the same type of thing as the visual sensory component of your experience of a reddish-orange afterimage, the two experiences or sensations must resemble each other in some respect you can appreciate on the basis of introspection. If an amputee's phantom limb pain and the actual limb pain of a legged person are the same type of sensation (i.e. pains or "leg" pains) then they must resemble each other in some similarly introspectable respect.

**Thesis B: From Introspective Resemblance to Qualia:** If the sensory component of a perception A and a mere sensation B resemble each other in some introspectable respect R, then A must have some qualitative character Q, and B must have some qualitative character Q\* where Q and Q\* are instances of the same kind or type.

Introspection is being likened to perception here. When I see a fire hydrant and I see a stop sign and I judge that they share a color (i.e. they are both red), I am noticing a similarity between the visible character of the one object and the visible character of the other. I judge that they are both red on the basis of this similarity in their visible surfaces. The idea behind the introduction of qualia is that they are a sort of "mental paint." If I have a reddish orange afterimage and judge that it is the same kind of sensory experience as the visual sensory component of my perception of a reddish orange chair, this is because I have introspected the character of the sensation and introspected the character of the visual sensory component of the perception of the reddish orange chair and noticed the similarity between these two states of consciousness. *Of course, the sensations or experiences can't be the same color; they aren't colored at all (e.g. light isn't bouncing off the sensations as they're either in my nervous system (if Smart is right) or not in space at all (if Descartes is right)).* **The chair** is reddish orange, **my sensation of the chair is not**. So I judge that the visual experience of the chair and the afterimage experience are both experiences **of** or **as of** reddish-orange. When I do this, I am assigning each of them a **qualitative type**.

When two sensations share a qualitative type, we say that both sensations *instantiate* or *realize* a given **quale**. Though we have just discussed pains, perceptions and afterimages, the sensory component of tasting something or hearing something or feeling a surface are also supposed to be qualia.

Smart rejects qualia when he says we can report that afterimages are like perceptions of colored surfaces but we cannot "state the respect in which" they are alike. To posit qualia is to state the respect in which these two things are alike: their introspectable or qualitative similarity.

Questions: Why doesn't Smart posit the existence of qualia? Is the identity theory compatible with the existence of qualia? Do those who believe in qualia so understood have to embrace dualism of some sort?

In either event, to move from mere belief in the existence of qualia so conceived to embrace a more substantive form of property dualism, a theorist would have to embrace a further claim about qualia.

**Qualia are Irreducible:** Qualia are irreducible just in case they are not identical to or analyzable into the functional properties of sensations nor identical to the biological properties of sensations, nor identical to the chemical properties of sensations, etc.

Question: Is the qualitative character of a sensation or experience reducible to its functional or biological or electrochemical properties? We'll discuss the prospects of a positive answer further when we read Armstrong's functionalist response to the argument for property dualism.