This is a draft manuscript. Comments are most welcome, but please do not quote without permission.

Lecture 3
Mechanism and Phenomenal Experience: The Heuristic Identity Theory

Phenomenal experience, the experience we have when employing our senses, has long posed a problem for any materialist account of the mind. Why do stimuli of a given wave length, when seen under normal conditions, produce the experience we call *seeing red*? Why do sound waves with specific features produce the experience of *hearing Mozart's Eine Keline Nachtmusik*? Why do stimuli with a specific chemical structure produce the experience of *smelling fresh pastry*? Each of these experiences generates well-known phenomenal experiences, and it is far from clear how to explain these experiences in physicalist, let alone mechanistic, terms.

The problem has haunted the major versions of materialism advanced in the 20th century. Philosophical behaviorists, notoriously, tried to characterize mental experiences in behavioral terms. As problematic as this might be for intentional mental states such as believing and desiring, behaviorists maintained that such an account could be defended. But even a dedicated behaviorist such as U.T. Place rejected a behaviorist account when it came to phenomenal experience.[1] He proposed instead that "consciousness is a brain process," explicitly employing the *is* of identity. On first pass, as Place and the other *Australian Identity Theorists* recognized, such a claim identifying phenomenal states with brain states seems preposterous. There doesn't seem to be anything red or musical about brain stains, and they are odoriferous only in decay stays, and even then they do not exhibit the odors the experience of which they are supposed to be identical to. This objection, as Place was at pains to point out, resulted form misunderstanding the identity claim. It involved what he terms the *phenomenological fallacy*:

> the mistake of supposing that when the subject describes his experience, when he describes how things look, sound, smell, taste, or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the 'phenomenal field.'

To avoid the phenomenological fallacy and make the identity claim at least a plausible claim, one had to focus not on the content of experience, but on the experiencing of that content. The content of our phenomenological experience is identified with what we are experiencing. What is to be identified with a brain state according to Place is the activity of experiencing:

> We begin by learning to recognize the real properties of things in the environment. . . . [I]t is only after we have learned to describe the things in our environment that we learn to describe our consciousness of them. . . . [W]hen we after-image as green, we are not saying that there is something, the afterimage, which is green; we are saying that we are having the sort of experience which we

[1] Skinner, however, seems to have been unconvinced of the futility of treating experiencing as a kind of behavior, albeit it an inner one that could not legitimately be appealed to in explaining other behaviors.

> normally have when, and which we have learned to describe as, looking at a green patch of light.

This distinction between the phenomenological experience and its content lies behind the recent attempts by William Lycan and Michael Tye to construe phenomenal experience as representational. The account I will develop in this lecture will be in the spirit of these representational accounts of phenomenal experience as involving the representations created in the brain.

The identity claims of Place and the Australian materialists, however, have been the focus of sustained criticism in recent decades. Perhaps the most focused objections stem from Frank Jackson , who has maintained that the brain activity does not explain why phenomenal experiences have the phenomenal character that they do.[2] That is, the brain processes do not explain why we experience red or music as we do. Jackson develops his claim with the famous example of Mary, the renowned neuroscientist who, while unable to see colors, develops a complete account of the color processing system in human brains. From this, however, she cannot figure out what the experience is like—what red looks like. When she later comes to experience red, Jackson contends, she will not just have a new experience, but will learn something she did not know before. Thus, he maintains, the neural processing account does not explain phenomenal experience

> I think that there are certain features of the bodily sensations especially, but also of certain perceptual experiences, which no amount of purely physical information includes. Tell me everything physical there is to tell about what is going on in a living brain, the kind of states, their functional role, their relation to what goes on at other times and in other brains, and so on and so forth, and be I as clever as can be in fitting it all together, you won't have told me about the hurtfulness of pains, the itchiness of itches, pangs of jealousy, or about the characteristic experience of tasting a lemon, smelling a role, hearing a loud noise or seeing the sky (p. 469).

The conclusion Jackson draws is a strong one: "Physicalism is false." For the defender of a mechanistic model of explanation in cognitive neuroscience, the conclusion that neural mechanisms cannot explain an important element of mental life, phenomenal experience, would be devastating enough.[3]

The challenge for a mechanistic account is to show how phenomenal experience can be woven into the account of neural processing in such a manner that these objections cannot be raised. That is, I need to develop the mechanistic account in such a way that

---

[2] Another very common way of raising the concern about the inadequacy of mechanistic explanation appeals to the possibility of qualia inversion or absent qualia. Qualia, on these accounts, are the introspectively accessible qualitative characteristics of our phenomenal lives. Proposals of qualia inversion propose that in different individuals red and green qualia, for example, may be swapped so that other person has the same quale that the other person has when seeing a red object when they see a green turtle. But, ex hypothesi, the two individuals exhibit the same behavior. Absent qualia arguments propose that in fact one could exhibit the same behavior without having any qualia. The person would then be a zombie in Chalmer's sense of the term. I resist in this lecture using the term *qualia* since that reifies a feature of an experience, and the framework I develop will focus on the mental activity.

[3] There are a host of philosophical counters to Jackson. For a taxonomy of various lines of response, see (van Gulick, 1993).

when it is filled in, it will not leave out phenomenal experience. This involves, in part, determining what kind of processing in the brain constitutes having phenomenal experience and how the features of phenomenal experience are generated from sensations and have effects on behavior. In essence, then, the response to Jackson's Mary example will be to show that she could not have a complete account of color processing in the brain without taking into account phenomenal experience.[4]

Mechanistic accounts of mental processes are committed to identity claims, and thus to identity claims relating phenomenal experiences to the activity of neural components. To see exactly what the identity claim comes to, recall that a mechanistic account of a system involves, as I have developed in the previous lectures, the decomposition of a system both functionally and structurally. The mechanisms in question when we are considering phenomenal experience are those for processing sensory inputs. The activities of the sensory processing systems are to provide the information needed to guide behavior either directly or indirectly by supporting higher cognitive operations that will eventually guide behavior. The activity of the processing system is not, per se, to generate phenomenal experience. Rather, phenomenal experience arises in the context of the mechanism carrying out its operations. Therefore, what we should expect is that some processes in the mechanism constitute having phenomenal experience. So, part of the task is to determine which processing activities carried out in the system constitute having phenomenal experience. So far we are identifying the having of phenomenal experiences with certain activities in the sensory processing system. But these processing activities are activities of particular physical components of the system. As I have employed the notion of localization, it involves identifying a component operation with a component part of the system. It is a component part that carries out the process. So there is a further identity claim that the component part (brain area) is what performs the component operation.

From Identity Claims to Heuristic Identity Claims

In advancing an identity claim, the mechanist seems to fall prey to an objection that bedeviled the Australian identity theorists: how can one ever show that two things are identical, not just correlated? All of the evidence advanced for an identity claim seems to be correlational. For example, one might show that whenever one has a certain phenomenal experience, a certain set of processes are transpiring in one's brain. But correlation does not prove identity and a dualist can accept a correlational claim that a particular mental experience is correlated with a particular brain processes since that is compatible with the mental process being something other than the brain process. Jackson, for example, does not rule out that Mary, once she has learned what it is to experience red, may learn in addition that whenever she experiences red, a specific set of

[4] Likewise, in response to those who pose the objection in terms of qualia, the challenge is to develop an account of neural processing that provides a central role to qualia and makes either qualia inversion or the absence of qualia impossible. Although I cannot develop it here, part of the key to this response will be to show that qualia are not simples. Rather, if we accept the reification of qualia, they turn out to be complex structures and the different features of those structures explain the different behaviors resulting from experiencing one quale rather than another.

brain processes are occurring in her. What she cannot do, is establish that those brain processes are the experiencing of red. After all, ex hypothesi, she knew everything there was to know about the brain processes themselves without knowing that they were correlated with seeing red.

Another way of raising the challenge to the physicalist/mechanist is to point to a gap between what we can say about brain processes and the character of phenomenal experience. Joe Levine (1983) refers to this as the *explanatory gap*. With ordinary physical identities, such as that pure water is the chemical substance with the molecular composition H_2O, one can explain the features of water, such as freezing at 0° Celsius in terms of the chemical structure with which it is identified. But in the case of having a phenomenal experience, for example, seeing red, nothing about the brain process explains why it has that qualitative character

As a strategy for answering these objections to an identity claim, I propose to consider in greater detail the role identity claims play in scientific research, especially in the attempts to understand a mechanism. Identity claims are not the end products of research. Rather, they are hypotheses made in the course of, often at the outset of, research toward understanding a mechanism. They serve to guide research. To emphasize this function, Robert McCauley and I call them *heuristic* and speak of the version of the identity theory which I will develop and defend here as *heuristic identity theory* (Bechtel & McCauley, 1999; McCauley & Bechtel, 2001).

Heuristic identity claims are made long before there is sufficient evidence to assert even a strong correlation. They are made when evidence first suggests that what scientists might be dealing with are two characterizations of the same phenomenon. Very often, one characterization will be structural and the other functional. The hypothesized identity becomes the basis for further discovery by invoking Leibniz's law of the indiscernability of identicals, which holds that if two designations pick out the same entity, then everything that is true of the entity under one designation must be true of it under the other. Note that these are claims about the entity or operation in question. As the early Australian identity theorists recognized, sometimes one would be unlikely to make a certain attribution to the entity or operation while employing one designation, but since the attribution is about the entity or operation, the transfer goes through (Smart, 1959).

Often research that has been done on the entity or operation at the time the identity claim is advanced under the two different designations will attribute different things to it. But if the two designations are really about the same entity or operation, we should be able to discover that everything ascribable to the entity under one designation is also true of it under the other. If, however, the connection was just a brute correlation, there is no reason to expect such discoveries. As with any consequence of an hypothesis that would seem to be unlikely unless the hypothesis were true, the discovery that these ascriptions made on the basis of assuming identity turn out to be true provide powerful evidence for the correctness of the identity claim.

Consider a case, due originally to Lindley Darden but developed further by William Wimsatt. At the beginning of the 20th century cytologists had identified chromosomes as paired structures in the nucleus while the rediscovery of Mendel suggested paired factors accounting for heredity. Theodor Boveri (1902), in research on sea urchin eggs, discovered that if an egg was fertilized by two sperm, it both received an extra set of chromosomes and developed abnormally. Learning of Mendel's theory, he proposed a link between Mendian factors and chromosomes. During the same period, Walter Sutton (1903), working with grasshoppers, determined that different chromosomes made different contributions to development and likewise drew connections between chromosomes and Mendel's factors. At the time, only three similarities were identified under both characterizations, but three other features known to be true of chromosomes were inferred to be true of Mendel's factors, and two fetures of Mendel's factors were inferred to be true of chromosomes. These became the focus of investigation as did other features that were discovered to apply to either chromosomes or Mendelian factors on the basis of further research (Wimsatt, 1976). By the time the identity claim had supported numerous confirmed inferences about both chromosomes and genes, the claim that genes were units on chromosomes was no longer just a claim of correlation. There is no reason that a mere correlation would continue to match newly discovered properties. That it did so was powerful evidence that the identity claim was true. But by this stage, no one (except researchers who denied the materiality of genes) was questioning the identity claim.

In many cases, one of the two designations will pick out an entity in term functional terms while another will identify a structure. In this case, what one predicts form the identity claim is that the further decompositions of the structure and the function will continue to map unto one another. An example will make this clear. By the mid-1940s biochemists had developed a detailed account of the operations of oxidative phosphorylation, the basic process by which energy is liberated from foodstuffs and stored in the form of ATP until needed for such things as muscle contraction. During the same period the pioneers of the new field of cell biology developed techniques that located the critical enzymes of oxidative phosphorylation in the mitochondria of cells. These results led Albert Claude (1948) to propose that the mitochondrion was the power plant of the cell. This was an identity claim. This claim proved its mettle in what happened afterwards. Biochemists, noting that preparations that would carry out oxidative phosphorylation *in vitro* required membranes and that the agents responsible for the different reactions were not readily separated, proposed that the responsible system was a complex set of enzymes held in a specific arrangement which David Green (1951) termed a *cyclophorase system* (see also Lehninger, 1951). In the early 1950s George Palade (1952) discovered additional structure in the mitochondrion—the mitochondrion possessed a double membrane, the inner of which was repeated folded inwards in the cell, creating structures he labeled cristae mitochondrioles. He now proposed that it was the cristae that were the true locus of the enzyme system responsible for oxidative phosphorylation, and numerous biochemists tried to demonstrate how enzymes were localized in the cristae. One component that was isolated in this process by Efraim Racker was an enzyme that broke down ATP to ADP, which was accordingly identified as an ATPase. In 1962, when yet further structures, heads on stalks protruding from the cristae, were discovered by Humberto Fernández-Moran (1962), these were proposed to be the locus of the ATPase. But why an enzyme to break down ATP when the challenge was to explain its synthesis and what was the specific importance of the membrane? This became clear when Peter Mitchell (Mitchell, 1961, 1966) advanced his chemiosmotic hypothesis, which proposed that ATP synthesis was achieved through creation of a proton gradient over the cristae membrane, which then served to drive in reverse the activity of the ATPase localized in the stalks. In the two decades after Claude advanced the identity claim, the details of structure and function were progressively mapped onto each other, producing in the end a mechanism that accounted for the phenomenon. The initial identity claim had born rich fruit.

An important thing to note about both of these examples is that by the time the details of the mappings were worked out, virtually no one questioned the identity claim. No one did so because the claims had provided the foundation stone to a history of productive research. This research was based on assuming the identity. Assuming correlation would not have been sufficient. If genes were only correlated with chromosomes, or oxidative phosphorylation were only correlated with mitochondria, one would not have had a basis for assuming that the details understood in terms of one designation would also map onto the details understood in terms of the other. But the identity hypothesis entailed, via Leibniz's Law, that these mappings would be found for everything known at the outset or discovered in the course of subsequent inquiry.

If identity claims in science typically function as indicated by the heuristic identity theory, then those who maintain that all we can do is look at the evidence that has been mustered at a time and then contend that it could never demonstrate more than mere correlations between phenomenal experience and brain processes are imposing an unrealistic demand on defenders of the identity of phenomenal experience and brain processes. The importance of the heuristic perspective is the recognition that what gives credibility to the identity claim is not that the correlations advanced at a given time, but the productivity of the claim in discovering new phenomenon that would only be expected on the basis of the identity. But that does not mean that the identity claim does not impose a tough demand. It does, but it is a demand not on evidence at a time but on the development of evidence in the course of research. What one should be seeking now is a promising identity claim between phenomenal experience and brain processes, one that points to future productive research. After advancing the identity claim, the demands imposed by Leibniz's Law apply. Future research must fill out the mapping by showing that everything known or learned about phenomenal experience maps onto neural processes, and vice versa. If this research is productive, then the identity claim will have proven its worth, and the charge of mere correlation will lose plausibility.

The idea of a two-way development in which researchers both use phenomenal experience to guide discovery about the neural mechanism and use features of phenomenal experience to guide discoveries about phenomenal experience may strike some as implausible. They might maintain that we already know all the features of phenomenal experience and the discovery process is completely in one direction. But in fact research regularly reveals new facets of phenomenal experience. The discoveries of new illusions demonstrate ways in which our phenomenal experience may be misled. The discovery of change blindness reveals that we are not nearly as aware of what is happening directly before our eyes as we generally think (Sinons & Levin, 1997).

Phenomena such as change blindness were discovered through behavioral experimentation, not based on features of the neural processing system. But there already are examples of discoveries about phenomenal experience that have resulted from considering features of the neural processing system. For example, based on the fact that the human color processing system employs, Paul Churchland (2002) has developed demonstrations in which, after fixating for a period on a red patch and then shifting to a grey patch, the grey patch appears to have a green tone. Drawing upon computational models based on what is known of the face-processing system in the brain, Churchland has developed similar demonstrations in which, after fixating on a hyperbolic male face and then looking at an androgynous face, appears female. Relying on the same technique, Churchland has developed demonstrations in which saturating on a color and then focusing on another color produces an experience outside the range of supposed possible color experiences. Our knowledge of neither phenomenal experience nor brain processes is static, and if by assuming identity claims and using discoveries about one we predict and confirm hypotheses about the other, then the result will be a productive research program of the sort envisioned by the heuristic identity theory.

There is a complicating feature in the case of phenomenal experience, though, which must be acknowledged. Although it may seem to philosophers consumed with phenomenal experience that if there is a responsible neural mechanism, then producing phenomenal experience must be the activity of this mechanism, it is now pretty clear that this is not the case. Phenomenal experience arises in the process of perception, and the activity of perception is to provide information needed to guide action, not to produce a Cartesian Theatre of visual experience (Dennett). Therefore, the outcome will not be a mechanism that generates phenomenal experience, as the outcome of research on mitochondria was a mechanism to achieve oxidative phosphorylation. Phenomenal experience must instead be viewed as a feature of the operation of the mechanism, not what the mechanism does.

By marking this distinction I am not suggesting that phenomenal experience is epiphenomenal. Although that is a logical possibility, it seems far more plausible that phenomenal experience plays an important role in the operation of neural processing systems. There are currently numerous hypotheses about what contribution phenomenal experience makes. At present many of these seem highly speculative. One of the payoffs of successful deployment of the heuristic identity theory would be an understanding not only of what are the parts of the mechanism whose operation constitutes having phenomenal experience but also how these processes make a difference in the activity of the larger mechanism.

Situating Visual Experience in the Visual Processing System

In the previous lecture I described how research, especially that in the second half of the 20th century, produced a detailed sketch of the visual processing system. In terms of this sketch we can explore where visual experience might arise. To do this we need to begin with an account of what our visual experience is. One thing to note is that our visual experience is perspectival. We see the world from a point of view. In addition, our visual experience is of a world of mostly middle-sized objects and events involving mostly middle-sized objects. These and other basic features provide some clues as to where to locate visual experience in the visual processing system.

As discussed in the previous lecture, current accounts of the visual processing system differentiate two streams of visual processing, the *what* versus *where* streams (Ungerleider & Mishkin, 1982) or the object identification versus action oriented streams (Milner & Goodale, 1995). A first localization question concerns whether visual experience arises only in one or the other of these two streams. There is, in fact, good reason to exclude visual experience from later stages in the *where* or action oriented stream. Much of the evidence for this stems from patients suffering from visual agnosia, the inability to recognize and report what one is seeing. Such patients are not blind. As we will see, they retain many visual abilities, including some visual experiences.

Visual agnosia results form damage primarily to the ventral or *what* stream. Patients with visual agnosia can perform a variety of visual tasks that only require the *where* or action-oriented stream. When asked to put their hand through a slot, the patients perform normally. When asked to match the orientation of a slot, however, the patients performance was approximately chance. The latter task seems to involve visual experience in a manner the former does not.

This suggests that in looking for the processing that constitutes visual experience we look in the ventral stream. But where in the ventral steam? In what follows I will discuss a proposal due to Jesse Prinz (2000; 2001), who argues that the processing in extrastriate cortical areas is the best candidate for the location of visual experience.[5] This is a controversial proposal. Many other theorists argue that later processing in inferotemporal cortex or in frontal areas should be part of the so-called *neural correlate of consciousness*. (Because I am advocating identity claims, not mere correlations, I will avoid as much as possible using the language of *neural correlate*.)

To make the case that it is activity in extrastriate cortex that constitutes phenomenal experience, Prinz first dismisses processing earlier in the visual pathway, including precortical areas and V1. The reason is that cells in these areas are not responsive to the sorts of stimuli of which we have phenomenal experience. For example, cells in V1 do not respond to illusory contours or to motion illusions and do not exhibit color constancy, all of which figure in our phenomenal experiences. On the other hand, we do not have phenomenal experience of features to which V1 cells are responsive, such as extremely high resolution bars of light.

---

[5] Prinz in turn credits the inspiration for his analysis to Ray Jackendoff (1987).

The situation, however, is very different when one turns to prestriate cortical areas. Cells in V2 are responsive to illusory contours. Cells in V4 exhibit color constancy in response to color stimuli. Research by William Newsome (Newsome, Britten, & Movshon, 1989) provides particularly striking evidence that cells in MT are responsive to experienced motion. Newsome trained monkeys to respond differentially to stimuli comprised of dots which either all moved in the same direction or randomly with respect to each other and identified cells in MT that responded to coherent motion in a specific direction. The critical test then came with stimuli exhibiting only partial correlation of dot movement. He showed that on stimuli for which the correlation was too small to produce a reliable response, activity of MT cells corresponded to the monkey's behavioral response as to whether it detected correlated motion. Apparently activity in these cells constituted the monkey's recognition of coherent patterns of movement in a given direction. Further confirmation of this came from additional studies in which he micro-stimulated these cells in the context of figures that had a low probability of detected correlated motion and show that he could thereby bias the monkey's behavioral response.

Nancy Kanwisher (2001) has extended these findings of a fit between activity in extrastriate cortex and visual experience to humans. Previous research had established different areas in extrastriate cortex that responded when a person viewed images of faces (an area in the fusiform gyrus she refers to as the fusiform face area or FFA) and another that responded when the subject viewed images of places (an area on ventral surface of the brain she refers to as the parahippocampal place area or PPA). She presented to human subjects an image of a face to one eye and of a place to the other eye. Using event relate fMRI she demonstrated that when subjects indicated a shift in their phenomenal experience, for example, from seeing the face to seeing the place, there was a corresponding change in blood flow from the FFA to the PPA areas.

Results such as these suggest that prestriate cortical cells figure in phenomenal experience. But are they the only brain processes that do? There is evidence, stemming initially from research of Nikos Logothetis (1998) which suggests that it is later parts of the ventral pathway that are the proper locus of visual experience. These experiments employed the paradigm of binocular rivalry in which different stimuli were presented to the two eyes. With such presentations, subjects do not experience a blend of the two stimuli, but successive experiences of each. Using a design similar to Newsome's, Logothetis trained monkeys to pull levels depending on which stimulus they experienced. Although the activity of some cells in prestriate areas (as well as some in V1) corresponded with the monkey's response, the correspondence was much greater in inferotemporal cortex.

Logothetis' evidence suggests at least including processing in inferotemporal areas in the processing that comprises visual awareness. But Prinz rejects this proposal and restricts the processing involved in visual experience to extrastriate areas. One reason is that the processing in inferotemporal areas visual experience is always from a point of view (that is, it is egocentric) whereas processing in inferotemporal cortex is allocentric. Cells in IT also have very large receptive fields, rendering them relatively size- and position-invariant, whereas we experience things as being of a certain size or in a particular position. Moreover, the same cell will respond to presentations of an object from a variety of viewpoints.

Another line of evidence Prinz appeals to involves visual agnosia, which we already invoked to rule out processing in the dorsal stream as involving visual awareness. As discussed in lecture 2, the phenomenon of psychic blindness was first identified by Schäfer (1888), who reported that after large lesions to the temporal lobe, would seemingly fail to recognize objects with which they were quite familiar. Neither Schäfer nor Kluwer or Bucy (1938), who termed the deficit *psychic blindness*, were able to provide information as to what the animal experienced. But tests with human patients who suffer from this deficit provide some clues. These patients are able to see the object presented to them well enough to describe it, draw it, or distinguish it from another presented object. But they cannot *recognize* the object, and that deficit is not due to loss of the name for the object. They are typically able to name the object when presented in another modality, and can respond correctly to questions that use the name of the object (e.g., they can say what such an object can be used for).

Perhaps the most compelling evidence that patients with visual agnosia have visual experience of the object is that they can make drawings of the object. Rubens and Benson report on one patient:

> the patient could not identify common objects presented visually, and did not know what was on his plate until he tasted it. . . . He was never able to describe or demonstrate the use of an object if he could not name it. . . He could match identical objects, but not group objects by category (clothing, food) . . . He was unable to recognize members of his family, the hospital staff, or even his own face in a mirror. . . . Remarkably, he could make excellent copies of line drawings and still fail to name the subject (Rubens & Benson, 1971).

Although the line drawings their patient made were impressive, it is important to note that the drawing process was not normal—it was extremely slow and slavish. Nonetheless, since they carry out this active upon verbal instruction, and do not protest (as, for instance, do blindsight patients when asked to identify an object in the blind field), it seems highly plausible that these patients have visual experience.

Prinz also responds to the claim of Logothetis that the high correspondence of responses in inferotemporal cortex shows that these cells are the true site of visual experience. Prinz first develops an analogy in which extrastriate areas constitute an orchestra and inferotemporal cortex is the conductor—just as the sound comes from the orchestra and the conductor only insures that it performs the music. Then, in response to Logothetis, he expands the analogy:

> Instead of one orchestra, imagine that a group of orchestras are all playing at the same time in extrastriate regions. Each one is competing for the attention of IT. Imagine further that the loudest orchestra counts as the victor of this competition. If victory consists in being loudest, the victorious orchestra can count as victorious even if it is playing at the same time as the other orchestras. By parity of reasoning, some subset of firing cells in extrastriate cortex could correspond to the conscious percept, even though competing cells are active. The fact that only

> a limited percentage of cells in extrastriate areas change with changes in the percept, does not show that these areas are not correlated with experience. It may only suggest that we should identify conscious percepts with a subset of the active cells in these areas. Likewise, the fact that 90% of the cells in IT track the percept does not show that IT is a correlate of visual consciousness. It may only suggest that IT responds to just those cells that have won the battle consciousness (p. 281).

Prinz goes on to suggest that inferotemporal cortex serves in part to focus attention on selected activity in extrastriate areas. As with almost all forward pathways in the brain, there are recurrent or backwards projections from inferotemporal cortex to extrastriate areas. It is plausible to assume that as a result of backwards projections, once cells in inferotemporal areas register the identity of an object, backwards projections to the extrastriate areas will selectively enhance or inhibit representations in extrastriate cortex. This might explain, for example, the difference in drawing by a visual agnosia patient and a normal person—the visual agnosia patient lacks the feedback from inferotemporal areas that serves to accentuate the representations of features that go into a recognizeable object and inhibit the representations of the extraneous features.

Prinz's full claim is that it is only when attention results in increased processing in extrastriate areas does visual experience emerge. The attention mechanisms he envisions are not just those from inferotemporal cortex but, for example, from frontal areas. Early research by Corbetta (Corbetta, Miezin, Shulman, & Petersen, 1993) and collaborators has demonstrated that when subjects are directed to attention to specific features of visual input (for example, color rather than motion), there is increased blood flow in the relevant extrastriate areas. Moreover, as phenomena such as change blindness and others show, when attention is distracted or not focused on features, subjects are not aware of them. Prinz therefore refers to his theory as the AIR theory, where AIR stands for *attended intermediate representations*.

Employing the Identity as a Heuristic

The evidence advanced in the previous section for Prinz's AIR theory was not intended to prove the identity claim, but only to render it plausible. According to the heuristic identity theory, the point of such claims is to guide further research. If it is correct, then additional research ought to further explicate the relation between visual experience and processing in inferotemporal cortex. More specifically, in accord with Leibniz's Law, as research reveals features of processing in extrastriate cortex, this ought to guide development of our understanding of visual experience. And if research produces additional research about visual experience, research ought to map that onto neural processing in extrastriate cortex. And if the understanding of the processing in these areas expands, that ought to allow us to identify either new aspects of visual experience or to at least control these features in ways we cannot now anticipate.

One approach to further research on the role of extrastriate areas on visual awareness is to investigate either whether there can be activity in extrastriate areas that does not involve visual experience or activity processes outside of extrastriate cortex which seem to be necessary for visual experience. There seems to be a compelling case for the latter. Damage to the inferior parietal cortex produces a phenomenon of visual neglect in which a patient fails to notice a face on the left side of a visual display. Nonetheless, as (Rees et al., 2000) showed with imaging, there is activation in the face area in the right hemisphere that does not appear to be different than that generated in normal subjects who have visual experiences. (Luck, Vogel, & Shapiro, 1996) produced similar results using the attentional blink paradigm in which, when subjects are presented two successive masked stimuli, but are required to respond to the first, they do not report seeing the second stimulus, although there is evidence in the form of ERP responses, that it is processed normally.

One way to respond to evidence such as this is to view them as the proposed identity. But if one takes the identity claim as a heuristic, a different attitude is called for. One searches for differences in processing in extrastriate cortex that separates cases where it gives rise to visual experience from cases where it does not. This may involve expanding the neural processes that we take to be involved in visual awareness. Kanwisher offers such a proposal—what is lacking it these experiences but is needed for visual awareness is "the binding of activated perceptual attributes with a representation that specifies the time and place that the word appeared (i.e. a 'token')" (p. 108), a process that she proposes involves activity in the parietal lobe. But there may turn out to be resources within the extrastriate areas themselves to explain the difference. The measures indicating normal processing in these areas are relatively crude (electrical patterns recorded on the skull surface or blood flow). More refined measures may reveal differences in the extrastriate processing. And yet another possibility is to maintain that the subject has normal visual experience but simply cannot report it because of some inability to access the visual experience.

Since what I am advocating is for putting forward a relation between visual experience and brain processing as a heuristic identity claim, it is not incumbent on me to show how challenges to that claim in terms of apparent differences between visual experience and what is processed in extrastriate cortical areas will eventually be resolved. This will require ensuing research. The existence of such cases, moreover, is not a problem. Apparent failures to find a perfect mapping are what get the research program advocated here off the ground. Whether the proposed identity claim, or a revised version of it, will be vindicated is a question for the future.

What we can say now is what we should expect if the proposed identity claim gives rise to a productive research tradition in which new mappings between visual experience and processing in extrastriate cortex. At that stage, the identity will have proven its mettle and most researchers will simply take it for granted. The claim that it all we have is merely correlation will have long since been dismissed. Moreover, as we noted, there will be good reason for dismissing these worries—if all that we were dealing with were correlations, we would not expect the further mappings to have proven successful. That they have supports the claim of identity.

At this stage we can also deal with one of the factors motivating the mere correlation worry—the explanatory gap. When looked at as just a correlation between visual experience and neural processing, there does seem to be a gap. Why should this processing constitute one sensory experience rather than some other. But notice that from the heuristic perspective, the gap has been partly bridged. As we enrich our understanding of visual experience on the basis of our understanding of the neural mechanism, we are moving beyond brute correlation. The feature of visual experience will be tied to and explained by the features of neural processing.

Conclusion

Sensory experience poses a challenge for a mechanistic account of the brain, but not the insuperable obstacle some philosophers contend. What is required is to identify sensory experience with sensory processing. Such an identity claim, though, must be understood as a hypothesis to be supported by further research. Looking just at the amassed evidence, currently or in some completed science, there is no reason to maintain identity rather than a mere correlation between the sensory processing and sensory experience. But when the identity claimed is viewed an a hypothesis that generates, via Leibniz's Law, the implying the potential false claim that what is known or learned about sensory experience but not already mapped onto sensory processing will map in the future, and vice versa, then it is subjecting itself to empirical evaluation like any other hypothesis. If in future research the requisite mappings are developed, then the identity claim will have proven successful and will be accepted into the framework of science in the manner of identity claims made in developing mechanistic explanations elsewhere in science.

References

Bechtel, W., & McCauley, R. N. (1999). Heuristic Identity Theory (or Back to the Future): The Mind-Body Problem Against the Background of Research Strategies in Cognitive Neuroscience. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st Meeting of the Cognitive Science Society* (pp. 67-72): Mahwah, NJ: Lawrence Erlbaum Associates.

Boveri, T. (1902). On multipolar mitosis as a means of analysis of the cell nucleus.

Churchland, P. M. (2002). Out space and inner space: The new epistemology. *Proceedings and Addresses of the American Philosophical Association, 72*(2), 25-48.

Claude, A. (1948). Studies on cells: morphology, chemical constitution, and distribution of biochemical functions. *Harvey Lectures, 43*, 121-164.

Corbetta, M., Miezin, F. M., Shulman, G. L., & Petersen, S. E. (1993). A PET study of visuospatial attention. *The Journal of Neuroscience, 13*(3), 1202-1226.

Fernández-Morán, H. (1962). New Approaches in the Study of Biological Ultrastructure by High-Resolution Electron Microscopy. In R. J. C. Harris (Ed.), *The Interpretation of Ultrastructure* (Vol. 1, pp. 411-427). New York: Academic Press.

Green, D. E. (1951). The cyclophorase system of enzymes. *Biological Reviews, 26*, 410-455.

Jackendoff, R. (1987). *Consciousness and the computational mind*. Cambridge, MA: MIT Press.

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition, 79*, 89-113.

Klüver, H., & Bucy, P. (1938). An analysis of certain effects of bilateral temporal lobectomy in the rhesus monkey, with special reference to "psychic blindness". *The Journal of Psychology, 5*(1938), 33-54.

Lehninger, A. L. (1951). The organized respiratory activity of isolated rat-liver mitochondria. In J. Edsall (Ed.), *Enzymes and Emzyme Systems* (pp. 1-14). Cambridge, MA: Harvard University Press.

Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly, 64*, 356-361.

Logothetis, N. K. (1998). Single units and conscious vision. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 353*, 1801-1818.

Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meansing can be accessed but not reported during the attentional blink. *Nature, 383*, 616-618.

McCauley, R. N., & Bechtel, W. (2001). Explanatory pluralism and heuristic identity theory. *Theory and Psychology, 11*(6), 736-760.

Milner, A. D., & Goodale, M. G. (1995). *The visual brain in action*. Oxford: Oxford University Press.

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature, 191*, 144-148.

Mitchell, P. (1966). *Chemiosmotic coupling in oxidative and photosynthetic phosphorylation.*: Bodmin., Glynn Research Ltd.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature, 341*, 52-54.

Palade, G. E. (1952). A study of fixation for electron microscopy. *Journal of Experimental Medicine, 95*, 285-297.
Prinz, J. (2000). A neurofunctional theory of visual consciousness. *Consciousness and Cognition, 9*, 243-259.
Prinz, J. (2001). Functionalism, dualism, and the neurocorrelates of consciousness. In W. Bechtel & P. Mandik & J. Mundale & R. M. Stufflebeam (Eds.), *Philosophy and the neurosciences: A reader* (pp. 278-294). Oxford: Blackwell.
Rees, G., Wojciulik, E., Clarke, E., Husain, M., Frith, C. D., & Driver, J. (2000). Unconsious activation of visual cortex in the damaged right hemisphere of a parietal patient with extinction. *Brain, 123*, 1624-1633.
Rubens, A. B., & Benson, D. F. (1971). Associative visual agnosia. *Archives of Neurology, 24*, 305-316.
Schäfer, E. A. (1888). On the functions of the temporal and occipital lobes: A reply to Dr. Ferrier. *Brain, 11*, 145-166.
Sinons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science, 1*(261-267).
Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review, 68*, 141-156.
Sutton, W. S. (1903). The chromosomes in heredity. *Biological Bulletin, 4*, 231-251.
Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle & M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586). Cambridge, MA: MIT Press.
van Gulick, R. (1993). Understanding the phenomenal mind: Are we all just armadillos? In M. Davies & G. Humphreys (Eds.), *Consciousness: Psychological and philosophical essays*. Oxford: Blackwell.
Wimsatt, W. C. (1976). Reductionism, levels of organization, and the mind-body problem. In G. Globus & G. Maxwell & I. Savodnik (Eds.), *Consciousness and the Brain: A Scientific and Philosophical Inquiry* (pp. 202-267). New York: Plenum Press.