

CHAPTER 12

SUBJECTIVITY

JOHN PERRY

**Subjectivity** (Philosophy) . . . Of or pertaining to the thinking subject: proceeding from or placed within the individual consciousness or perception, originating in the mind, belonging to the conscious life.

*(Shorter Oxford English Dictionary)*

SUBJECTIVITY is to many philosophers what the frog is to many biologists: the object of fascination that first drew them to their discipline. That we have experiences; that we can't say for sure that they reflect an external world; that we might be alone in the world; that our mind might be the plaything of an evil demon; that perhaps *to be* is *to be perceived*; that will and idea might exhaust reality; that even the self might be an illusion—to certain teenagers, who may prefer books to ponds and introspection to vivisection, these are thrilling thoughts. As the adult biologist, awash in dissertations about DNA, NSF proposals, and university committee meetings, needs to return to the pond for a few weeks each summer to rekindle love for the subject, so the philosopher needs to return periodically to his or her own subjectivity, accept it for what it is, or at least seems to be, enjoy it, explore it, swim in it, and think about it. If your Chair or your Dean asks what you are doing, say 'phenomenology'.

In the first part of this chapter I discuss what we find there in our individual consciousness or subjectivity: experiences of various sorts, including thoughts and thoughts about thoughts and thoughts about other experiences. I also briefly consider what these things are doing there. In the second part I discuss whether Frank Jackson's 'knowledge argument' provides a good reason to doubt that all these things we find in our minds are events in and states of our brain.

## 12.1 WHAT WE FIND IN OUR MINDS

---

### 12.1.1 Experiences

Locke calls everything we find in our minds ‘ideas’. Hume calls them all ‘perceptions’. Neither of these seems felicitous as a general term. I’ll just stick with ‘experiences’, and use ‘perception’ with its ordinary meaning, as the recognition and interpretation of sensory stimuli, even when talking about Hume.

Hume divides experiences into impressions on the one hand and thoughts and ideas on the other. Impressions include sensations and passions—what we now usually call ‘emotions’. Impressions come unbidden; we are passive; we have them but do not *do* them; they are not the products of thought and will. Ideas and thoughts are less lively and vivacious than impressions, and we often bring them to mind at will. It’s easier to think about a roast-beef sandwich than to see or taste one; to do the latter you have to order one or make it yourself.

Once Hume gets to work, however, he finds things less simple than his dichotomy suggests. Impressions come intimately associated with ideas, based on previous experience; what we would ordinarily call a perception involves not just a passive component, but also the result of various activities on our part: we compare, we remember, we classify, we anticipate, we predict, all of this rolled up with the having of sensations. What we would ordinarily call seeing a chair involves not only the sensations caused by the colour and shape of the chair, but anticipations of visual sensations to be had by moving this way or that, of experiences we expect to have (being supported) and not have (falling) if we do try sit on it, and the like. Perceptions are sensations associated with ideas and thoughts. In Book II of his *Treatise* Hume also develops a theory of passions as combinations of impressions and thoughts.

This all suggests that what we need is not so much two disjoint categories of experiences, but rather two dimensions, along which experiences differ, each dimension being more or less important in different cases. I’ll call these the *feel* and the *content* of *experiences*. A third dimension is physical painfulness and pleasure. Intensely painful and intensely pleasant sensations are perhaps as close as we get to the limiting case: all feel, almost no content. But even these sensations have a bit of content: the sensations are taken as ‘located’ in various parts of our body, and may give us information about what to do or not do with those parts to eliminate the pain or sustain the pleasure. By the same token, thoughts are not all content and no feel. There is something it is like to think a thought, although we don’t think of this feel as definitive of the thought. Thoughts involve words, often unspoken, images, anticipations of other kinds of experiences, and inclinations to act in various ways, and are often intimately associated with emotions.

The most striking feature of a sensation, segregated, as best we can, from the anticipations and memories and attendant thoughts that flesh it out into a perception,<sup>1</sup> is *what it is like to have it*, its feel, the aspect philosophers have called the *qualia*, *subjective character*, *phenomenal aspect*, and *raw feel* (see Chalmers 1996; Nagel 1974; Block 2007; Feigl 1967). Having a sensation of a red fire hydrant is quite different from having a sensation of a green patch of grass, even if we bracket off the information we seem to be getting about hydrants and lawns. Having a sensation of green is quite different from having the sensation one has when one hears a high trumpet note; the pain of a toothache is different than the pain of a backache; such unpleasant sensations are quite different than pleasant ones, like tasting chocolate, or smelling a rose. Although we may be convinced that there is this aspect of raw feel in experiences, we are hardly able to categorize without employing the idea of *of-ness*, where the far side of the *of* relation is not something subjective, but something that is outside the mind—a colour, a sound, a back, a tooth, a piece of candy, a rose, a fire hydrant, a lawn—or would be outside if it were real—a unicorn or my (imaginary) red Porsche.

Right now I have a visual field full of shapes and colours, which I take to be a computer in front of me, a book to my side, a cup full of hot coffee, a table underneath these things, my own hands perched on the keyboard of my computer, often waiting for inspiration, occasionally typing, a window, trees, and rooftops outside the house, and a lot more. I am listening to Johnny Cash sing about poor drunken Ira Hayes, so the sounds that comprise his words, and the sounds from his guitar, take up a lot of my auditory space. There are also, now and then, the sounds the keys on the computer make as I tap them, the sounds made by the cats pursuing various sort of mischief, occasionally the ring of a telephone, occasionally a request from my wife Frenchie. I feel the computer keys at the end of my fingers, and the surface on which my hands rest. I occasionally take a sip of coffee, so, and have that peculiar taste. While having all of these experiences, I think. I think about what to write, about whether to respond to Frenchie immediately or take a few seconds to finish what I am doing, whether to go investigate the cat noises or ignore them, whether coffee is unhealthy like the doctors said five or six years ago or actually not bad at all like they seem to say these days, and a zillion other things.

Thinking is often an active, purposeful activity. We intend to think about a certain subject matter: the weather, Calvin Coolidge, prime numbers, the commitments of the day, and then we manage to do it. But thoughts often come unbidden. While trying to remember the year of Calvin Coolidge's election, I find myself thinking about George Bush and Iraq. While trying to identify the first fifteen prime numbers in my

<sup>1</sup> I'll often ignore the 'success' or 'veridicality' or 'factive' implications of words like 'perception', and 'information'. This seems fair in an essay about subjectivity; by 'perception' I mean 'perception or would-be perception'.

mind, or follow the arguments of a student in my office, I find myself thinking about a roast-beef sandwich.

The most striking features of thoughts are their contents rather than their feels. We get at their contents by saying what things must be like to fit the thought, where the evaluative property and direction of fit are determined by what kind of thought it is.<sup>2</sup> The content of the thought ‘It’s sunny today’ is quite different from the content of the thought ‘Calvin Coolidge was a man of few words’. The direction of fit in the case of such *doxastic* or *belief-manifesting* thoughts is mind to world, the evaluative properties are truth and falsity. With wishes, like ‘Would that I receive a big raise’, the dimension of success is being granted or not; the direction of fit is world to mind; if my wish is not granted, the world, or at least the Dean’s office, is defective, not the thought. Thoughts are about things and their properties: days, weather conditions, presidents, loquaciousness, money, and such. The contents are conditions, involving these things—truth-conditions in the case of doxastic thoughts, conditions of being granted in the case of wishes.

But still it is like something to think, and it is like something to have thoughts simply occur to you: thoughts are experiences. A lot of thinking involves something like inner speech; something similar in feel to rehearsing what one is going to say, or anticipating saying it, or remembering saying it, or exhorting oneself. Indeed, people often lapse into talking to themselves when they are really into a piece of thinking. Thinking in English is different than thinking in German; it’s a thrill for the language student when the first thought formulated in the new language spontaneously makes its appearance. And thinking is not limited to words; all sorts of ideas, including images corresponding to various types of sensations are also involved.

The feel and content of experiences differ in their ontological status. The feel of an experience is an intrinsic quality of it. The sensation may be caused by an external object, and may be part of a contentful perception of that external thing. But the feel of the sensation is a fact about what is going on in my mind. A phrase like ‘the sensation of seeing red’ gets at the feel of the sensation in a roundabout way: it is (roughly) the type of visual sensation typically caused by seeing red things in favourable light.

The content of experiences, of perceptions and thoughts, is at least not entirely intrinsic to them. If I see Condoleezza Rice and think ‘That woman is the Secretary of State’, then my thought is *about* Rice, and is true because she is the Secretary of State. If it is not Rice that I see, but Angela Davis, then my thought is about Angela Davis, and is false because she is not Secretary of State, at least as I write this paragraph.<sup>3</sup> The truth-value and truth-conditions of my thought depend on whom my perception is of, and so are not intrinsic properties of it. The *feel* of the thought, however, seems intrinsic. I could have just those words running through my mind even if I didn’t perceive anyone. I’ll return to this issue below.

<sup>2</sup> For *direction of fit* see John Searle (1983).

<sup>3</sup> Angela Davis and Condoleezza Rice are the two most famous people I have ever actually met and talked to. Actually, it would be hard to mistake one for the other, especially if you are talking politics.

### 12.1.2 Cognitive States

Subjectivity does not exhaust what we ordinarily think of as ‘the mind’ or ‘the mental’. Beliefs and desires are mental, but they are not experiences in the way that sensations and thoughts are.<sup>4</sup> My beliefs and desires guide my thinking, and my thoughts and experiences in turn affect my beliefs and desires. Thoughts, like actions directed at the external world, manifest my beliefs and desires. If I believe that Sacramento is the capital of California, then when the question arises in my mind I will think ‘Sacramento is the capital of California’; if the question arises in conversation, that’s what I’ll say.

The word ‘concept’, as I shall use it, stands for cognitive structures involved in my beliefs and desires and other cognitive states. Using ‘concept’ in this way, concepts are not the same as ideas, considered as the bits of thinking that make up our thoughts.<sup>5</sup> Concepts, beliefs, and desires are not, like thoughts, transitory by nature; they are acquired at various times; they may last for years, or they may change after a few seconds. New experiences and new inferences lead to revisions in our beliefs; desires get changed by deliberation; sometimes they are satisfied and disappear; and we just forget things we once knew or at least believed or thought it was important to do or have.

We then must distinguish between that in our minds that we experience, the parts of the stream of consciousness, and the beliefs and desires that are more like rocks and fallen trees below the surface of the stream that direct its flow. These we are not directly aware of but can, in at least a wide range of cases, easily determine, and often alter.

Let me return now to the question of whether *any* of the intentional or semantic properties of beliefs and the thoughts to which they give rise are intrinsic. This question is somewhat vexed, as it is connected with Big Issues like anti-individualism and narrow content, and I won’t discuss it at great length here. But I will say a bit. There are two mistakes that are made in discussions on this topic I do want to mention.

The first, common in discussions of anti-individualism, is failing to distinguish between cognitive properties and cognitive states. In general, we distinguish between the properties systems have in virtue of the states of the system—the intrinsic properties of their parts and the relations the parts have to one another that are relevant to the way the system works—and the further properties that the system has in virtue of the way it is embedded into the wider world.<sup>6</sup> An engine may run at 4800 rpm while in a car, connected to the transmission, driveshaft, and wheels, or unconnected, while being tested in the factory. In the former circumstance it will also have the property of moving the wheels at some number of rpm, depending on the state of the transmission. The formula we use to think about such things, which has proven too

<sup>4</sup> This is a point I did not adequately grasp when I wrote *Knowledge, Possibility and Consciousness* (see Bach 1981).

<sup>5</sup> This is a departure from how I have used these terms in the past.

<sup>6</sup> For a helpful discussion see Hall (1993).

useful to be discarded by philosophers just when we get to truly interesting things like subjectivity, is that the states of a system plus its external relations yield its properties.

There is no doubt that having the property of *perceiving that Condoleezza Rice is reviewing the troops* is not something I have simply on account of what is going on in my head, the states of my brain, and so too with *believing that Condoleezza Rice did review the troops last week*. Those properties require that what goes on in my head be related in certain ways to Condoleezza Rice, and these relationships require things of the external world, not just things inside my head. It does not follow from this that the state I am in, in virtue of which, together with the external situation I find myself in, I have those properties is not simply a matter of what goes on in my head. It seems, for example, that there might be a state such that when a person is in it, in a fairly wide range of circumstances, that person perceives that the person they see is reviewing the troops. Well, maybe this is too optimistic, for how about the fact that what Rice was reviewing were *troops*? We can retreat further, to the state one is in such that, in a wide variety of circumstances, one who is in that state perceives that the person they are seeing is examining in an authoritative manner the group of uniformed persons that person is looking at. Well, one could have further qualms. That is, one might have to work hard to squeeze the commitments about the external world out of the property, so that the intentional description that is left, stripped of these commitments, only constrains the internal properties of the state. But the avenue for getting fewer and fewer commitments is fairly clear. We start with some intentional characterization of a minded being, such as:

Bush believes that Condoleezza Rice is reviewing the troops.

and then we reformulate things in a way that makes the intentional contributions of internal and external factors clear, relative to some account of the structure of internal states, which for purposes of illustration can be rather clunky and naive:

Bush is in a belief state with the structure  $C(n, g)$ , where  $C$  is a concept of the activity of reviewing,  $n$  is a notion of Condoleezza Rice and  $g$  is a notion of some troops, that is true only if the person  $n$  is of is performing the activity  $C$  is of upon the group  $g$  is of.

and then we abstract over the specific external factors we can identify:

Bush has a belief  $b$  with the structure  $C(n, g)$ , such that if there is an individual  $X$ , and activity  $A$  and a group  $G$ , and  $C$  is of  $A$ ,  $n$  is of  $X$ , and  $g$  is of  $G$ , belief  $b$  is true if and only if  $X$  performs  $A$  upon  $g$ .

The second mistake infected the once common and often quite worthwhile discussions of narrow content. This was the supposition that the content that we end up with, when we have completed this stripping enterprise, will be a proposition  $P$  such that the agent has the property *believing*  $P$ . The model was the transition from *de re* belief to a supposed underlying *de dicto* belief. This supposition is quite groundless. The whole idea of ‘intentionality’ is to describe what goes on inside of us in terms of what is outside of us; it is things outside of us that we by and large have beliefs about. But the conditions our stripping operation will leave us with will be existentially quantified conditions on things inside of us, our perceptions, concepts, notions,

and what have you. Bush may believe that Condoleezza Rice is reviewing the troops because he is in some state  $C(n, g)$  that meets the conditions above. That does not mean Bush believes the proposition *that there is a person  $X$  a relation  $R$  and a group  $G$  and my ideas  $C$ ,  $n$  and  $g$  are of respectively  $R$ ,  $X$ , and  $G$  and  $X$  is  $R$ -ing  $G$* . What he *believes* will be that Condoleezza Rice is reviewing the NATO troops. I doubt very much that Bush has many beliefs about his own concepts and notions. The narrow content we assign to what goes on inside our heads, as a part of an account of what we believe, perceive, know, conjecture, and the like, can play its role without itself being what we believe, perceive, conjecture, and the like (see Perry and Israel 1981).

### 12.1.3 The Epistemology of Subjectivity

While writing some of the paragraphs above I was not only having experiences and thinking, I was aware of my experiences; I attended to them, thought about them, and indeed wrote about them. It seems to me that having experiences, and becoming attuned to the information they carry, is something I share with all sorts of animals. All of these animals we might ordinarily call ‘conscious’, during those periods when they haven’t been knocked unconscious. But for this use of ‘conscious’ I’ll use ‘sensitive’.

The way ‘conscious’ is used in the OED definition of ‘subjectivity’ quoted at the beginning of this paper is different. We are conscious of things. We use the term in this way quite broadly. I am conscious of the dangers posed by Bush’s overspending, I am conscious of the racket made by the leaf-blower across the street, and I am conscious of the sensations I have as I type. Not everything we are conscious of is subjective, but it is consciousness of the subjective that we now turn to. It seems clear that we not only *have* experiences and think thoughts, we can become aware of, focus our attention on, compare, classify, ponder, admire, detest, describe, and in all sorts of ways *think about* our experiences and thoughts.

There is also the matter of remembering, imagining, and anticipating experiences, particularly sensations, perceptions, and emotions, which is a bit like thinking about them, but not quite. As I sit here typing, I begin to think about lunch, and the sandwich I can make from the fresh supply of roast beef from the deli. I anticipate eating and tasting it. Or maybe I just imagine doing so, for I know that when the time comes I’ll opt for some healthier alternative. Remembering, imagining, and anticipating experiences is a very special way of thinking about them, which I suppose we share with animals that cannot think about their experiences in the more robust way that we can.

So in order to be aware of my experiences I have to have them, but having them is not sufficient. Right now I am having an experience of seeing a computer. It is like something to have this experience. The experience plays a role in my life; it is involved in my perception of the computer, and my perception of it is involved in the interactions I am having with it: supporting it on my lap, typing on the keys, watching letters appear on the screen, and the like. None of this requires me to be aware of

the experience; having the experience is part of my being aware of the computer. I am attuned to the fact that when I have this sort of experience, in relatively normal conditions, there is a computer on my lap, and I can make letters appear on the screen by typing the keys. I am able, however, not only to have the experience but to attend to it, to form a concept of it, to apply concepts to it, and the like. It is a visual experience; it is neither pleasant nor unpleasant; I didn't directly pay money to have the experience, although I paid good money for the computer. I can also form a concept of the type of experience I am having, and of course I can classify the experience in various ways. It is an experience of a computer; of this particular computer; a visual experience; an experience of a square white region embedded in an off-white region, with letters that form words on the white region and letters on keys in the off-white region.

When we attend to the experiences involved in perception, the path of least resistance is to classify them in terms of what they are *of*. There are two ways the phrase 'of' can be taken here, and often it is not necessary to mark the difference. I see my computer; my experience is *of* a computer; that is, a computer causes it. But it is also *of* a computer in that I see what I am seeing *as* a computer, and not, say, a mere computer façade or a television. I could mistake my television for a computer; then my experience would be of a television, in the first sense, but of a computer, in the second.

To see the object I am looking at as a computer involves much more than mere sensation. I take various discontinuities in my visual field as corresponding to the edges of the screen, for example; I see it as three-dimensional, which involves expectations of what would happen if I were to move my head a bit, or stand up and approach the computer, or reach out and touch it. I expect the pattern in my visual field to remain relatively unchanged as long as I keep my eyes open and stay still; because I see what I see as a computer, I don't expect it to suddenly walk away. I actually can see two computers, the laptop on which I am typing, and, in the distance, a desktop computer. The two experiences seem similar, but the similarity in the experiences mostly amounts to their being of the same sorts of things: computers with visible screens, keyboards, and the like.

When we turn from perceptions whose importance is the information they carry about the external world to those whose dimension of painfulness and pleasure dominates our interest in them, we still have difficulty squeezing all of the intentionality out of our way of thinking of them. I have a slight pain in my right wrist. Locating the pain in this way is to describe it in terms of the information it gives me about the part of my body that is the source of the pain. I know the pain is in an important sense not located in my wrist but in my head. I could have a phantom pain like this, without even having a wrist, were the nerves between wrist and head properly stimulated. The pain is in my wrist in that it gives me information about my wrist, and it is my wrist I will move in order to deal with the pain.

#### 12.1.4 What's it All About?

Why do we have experiences? In particular, why are we sentient? Why are some of our brain states *like something* to be in? There are two questions here. Perhaps the

first is what David Chalmers calls the ‘hard question’ of consciousness (1996). It is pretty much the same as Heidegger’s question, ‘Why is there something rather than nothing?’, at least on one interpretation of it. That is, the question is not why there is any universe at all, rather than nothing, but why there is experience at all, rather than a ‘dark’ universe.

The only way I can imagine of answering this question is by identifying the physical characteristics that differentiate states it is like something to be in from those that it is not like something to be in, an enterprise that will doubtless lead to a more complex set of distinctions than we have adumbrated here, as scientists mine the interplay between the three ways we have of knowing of such things: examining brains, studying behaviour, and attending to the states in which we find ourselves. As this knowledge develops, the question why a given brain-science-identified state is like *this<sub>i</sub>* (using ‘this<sub>i</sub>’ as an inner demonstrative for experiences and the types they exemplify) may have a clear answer, against a background of what it is like to be in other brain-science-identified states. At that point, when we can answer each such question against the background of answers to many others, there will not be, as far as I can see, any ‘explanatory gaps’ left; at least, none of the sort that ever get closed. Perhaps at some point there will be a moment of conceptual clarification, where the mysteries of the more general question no longer grip us. I hope so, for then philosophers may play a role in the Great Day when the answer to this secret is laid bare.

The second question is this. Given that there are brain states it is like something to be in, experiences, what purpose does it serve to be in such brain states? What are they for? What is mother nature trying to do with them?

In his *Dialogues on Natural Religion* Hume complains that an omnipotent, omniscient, and indulgent deity could have come up with something better than pain to motivate us to get out of dangerous situations. One can imagine a sort of permanent semi-erotic pleasure that normally suffuses all parts of our body. An injury to the foot or a sprained elbow or a decayed tooth wouldn’t cause the sorts of pain it actually does, but instead merely a diminution of pleasure in foot or elbow or tooth.

But we’re not here to complain. Working within the metaphor of a purposeful mother nature, this mother is clearly either not omnipotent or not omniscient or not indulgent. She makes do with what she has, to work her bizarre purpose, which is basically to get plants and animals to reproduce so that there will be plenty of nutritious stuff, decomposing, crawling and walking around for other plants and animals to absorb one way or another so they can last long enough to reproduce and get absorbed in turn. Given these purposes, and the fact that she has brain states it is like something to be in her repertoire, what can she do with them? What are experiences for?

Behind Hume’s complaint is the idea that our sensations give us information and motivate us to act in ways that makes sense given that information. Experiences that are dramatically painful or pleasant provide both motivation and information. When we step on a tack we are motivated to do something to bring the experience of pain to an end, and we know, innately or based on relatively little experience, what to do about it.

Less dramatic experiences, with no pain or pleasure intrinsic to them, provide more delicate kinds of information, that can be exploited through mechanisms built up through evolution, experience, accretions of culture, and the memories and thoughts of a particular person.

So our experiences are there to give us information about our bodies and the broader environment, in order to enable us to act in ways that increase the probability of successful reproduction, and to give us motivation for so acting. The basic way we use the information that is made available to us is through habit or, more generally, *attunement*. As Hume points out, these matters are much too important for nature to leave mainly to understanding and reason. Attunement means basically that (i) being in some state *S* carries the information that we are in some situation *E*; (ii) being in *S* causes us to act in some way that makes sense, given that we are in *E*. If I step on something sharp, I move my foot. I may not have the concept of danger, or injury, or survival, or of pain, or of a foot, or of me. Still, moving the foot makes sense given that not moving it will cause injury. I am attuned to the regularity or *constraint* that stepping on sharp things causes injury.

Most of our visual and auditory sensations aren't unpleasant or pleasant enough or informative enough on their own to motivate us to do much of anything. But we can learn. Pavlov's dog learned that the bell meant food; he became attuned to this constraint in his environment, and started to salivate when he heard the bell. My goats learn that when they hear the sound I make by pounding the side of an old five-gallon paint bucket there will be some alfalfa and sweet cob for them to eat if they walk in the direction of the sound. Dogs and goats need to discriminate and have the capacity to learn new habits—to become attuned to new constraints.

An *information game* is a pair of episodes in which a being gathers the information that *P* at some time, and some being, perhaps the same being, perhaps not, uses that information at some time, perhaps the same time, perhaps not, to do something that makes sense given the fact that *P*. If the beings and the times are the same, we have the 'straight-through' information game. Given the pain in my foot, I move it. Given the bell, the dog salivates. Given the sound of stick on bucket, the goats start to move in the direction of the sound.

We, and a number of other animals, have the ability to store information for later use. We detach some of the information from the perception that carries it, and reapply it later, when we recognize the same object. I call this 'the detach and recognize' information game. This is where beliefs and concepts come in. Perhaps a goat isn't hungry at the time it hears me pound on the bucket. It wanders off to do some other business for a while. Then it moves towards the place where the sound came from. The goat picked up the information that there was food in such-and-such a direction at one time, and then later—but not so much later that the food wouldn't still be there—it does something that makes sense given that information. To do this, the goat has to reorient itself, to reidentify the direction from which it heard the sounds of food, probably by recognizing local landmarks. It

won't suffice to be attuned to general patterns in its sensory field; it will have to have the conceptual apparatus to separate out trees and people and other salient objects, and reidentify them.

The process of identification and reidentification requires a sort of concept I call a 'buffer'. A buffer is a temporary structure in which we store information about a perceived object, prior to forming a full-blooded detached notion, or identifying the object as something we already have a notion of. Such buffers, and the whole phenomenon of recognition, lead to problems for certain received doctrines about the structure of knowledge and belief.

Suppose, for example, that I see Condoleezza Rice at a Stanford party. Because of the stress of her years in the Bush administration, and her diplomatic wardrobe, she doesn't look quite the same as she did while a professor at Stanford, and at first I don't recognize her. For a while I have two notions of Rice, my long-standing notion, associated with such concepts as being brilliant, strong-willed, and having left Stanford to join the Bush administration, and a temporary buffer, associated only with the concepts my perceptions deliver. What do I learn when I recognize her? It isn't hard to say how my doxastic states change. The two notions have to be of the same person, if my belief, 'Oh, that person is Condoleezza Rice', is to be true. But the change from before to after the recognition can't be captured by a proposition involving only Rice, without bringing in my buffer. It can't, that is, be captured by a proposition whose constituents are confined to the *subject matter* of my thoughts, the external objects my notions are *of*. To suppose that all knowledge can be so captured is what I call 'the subject-matter fallacy'. Such recognitional knowledge is fleeting, for our buffers are usually simply absorbed into the more permanent notions.<sup>7</sup>

It is this strategy, developed in many different ways, that has come to dominate human cognition. We each possess a rich set of concepts, or *notions* of individual things. We have two uses for the information associated with each of these notions. We use it to help us to recognize the objects of which we have notions, and we use it to help us do something that makes sense with those objects (or, more generally, something that makes sense given what those objects are like and their relation to us), once we have found them. So, for example, I have a notion of Michael Bratman that provides me with the ability to recognize him in favourable circumstances: when I can see him, when I hear his voice on the phone. But my notion also provides me with facts about him that are useful in interacting with him. For one thing, I know that his name is 'Michael Bratman', so I can address him with that name. I know he thinks a lot about the philosophy of action, so I can ask him questions about his work, as a way of being friendly, and as a way of learning about the topic.

Communication is an extension of the detach and recognize information game. A large part of what we are interested in communicating—and this includes putting it

<sup>7</sup> For more on the subject-matter fallacy see Perry (2001) and Perry (2003).

in books to be read by others as well as conveying it in face-to-face conversation—is *incremental* information about objects; that is, information that is only useful, in guiding action, once we have recognized the object in question and gotten ourselves in a position where we can act on it, or at least do things whose results depend on its properties. It is this sort of information that our language is designed to convey. We deal with detached information, in print, and in our minds, when we think and reason about things we are not perceiving.

Suppose, for example, that my goats have language and cellphones and somewhat more altruistic relations with one another than they actually do. One goat might call a friend, out of earshot of the signs of food, and tell him that there is food near tree X. This would be much simpler than providing the goat friend with *all* the information the latter needs to get to the food. The responsibility for *reattaching* the detached information is left to the goat friend; he can wander around until he sees tree X, then he can apply the incremental information, that X has food near it.

It seems impossible to envisage the development of communication without a pretty rich structure of concepts and belief-like states already in place. But, in turn, it seems impossible to understand thinking, of the sort that humans do, apart from abilities to conceptualize and symbolize information in the way required for communication. Our thinking, both theoretical and practical, is typically detached from the exact relationships to the objects thought about that will be necessary to act fruitfully on them. We deliberate, imagine, rehearse, and conjecture, all with detached notions of things and aided by the symbols of communication. And of course many concepts and thoughts are totally enmeshed with words, agreements, practices, and complex institutions made possible by language. As Norman Malcolm once remarked, the thought ‘I need to put the bottles out because the milkman comes tomorrow’ isn’t one that arose in anyone’s consciousness before there were milk bottles and milkmen (1970). Come to think of it, many younger readers may have no idea what milk bottles and milkmen are, or were.

Being able to attend to our experiences, classify them, note whether we like them or not, think about their causes and how to avoid having them or increase the chances that we will have them, adds power to our deliberations. We can plan not only to bring about certain results in the external world, but also to bring about or avoid certain results in our own subjectivity. Whether this itself serves any of mother nature’s purposes I rather doubt. It is certainly useful in subverting her plans. We invent ways to have the internal pleasures of the procreative act without procreating, for example, or ways to link the tastes we like to substances of no nutritional value. Most likely it is something she didn’t plan on. Human thought, language, and culture have taken the detach and recognize information game and run amok. That, at any rate, is my picture of human science and culture, but I’ll spare the reader details.<sup>8</sup>

<sup>8</sup> For more on the topics of this section see the last part of Perry (2002).

## 12.2 COULD EXPERIENCES BE BRAIN EVENTS?

### 12.2.1 Mary

In Frank Jackson's classic statement (1986) the 'knowledge argument' has one character and three steps. Mary is a brilliant woman who, for one reason or another, is raised in a room where she never sees colours, only black and white and, I suppose, shades of grey in between. Mary has *new* knowledge when she finally steps out of the black and white room and sees a red fire hydrant. But while in the black and white room she could well have known *all* the physical facts relevant to colour vision. Conclusion: her new knowledge is of a non-physical fact. We need to look closely at Mary.

Mary emerges from the black and white room, sees a fire hydrant, and has her first colour experience, call it *E*, of the type  $\text{quale}_{\text{RED}}$ . She thinks: 'This<sub>i</sub> experience is the type I have when I see, in these conditions, the colour of that fire hydrant'. She has a certain relation to the experience: she *has* it. A less inquisitive person might have left it at that, but she also *attends to* the experience. *E* is the referent of her thought 'that<sub>i</sub> experience' because of the relations it has to her: it is the one she is having and attending to. She is having an experience of the colour of the grass beside the fire hydrant too, but 'this<sub>i</sub> experience' doesn't refer to that experience, because it is not the one to which she attends.

Mary forms a concept of the *type* of colour experience *E* exemplifies. She notes that *E* is similar to the colour experience she has of the fire engine parked nearby, and not similar to the colour experience she is having of the grass next to the fire hydrant. She can introduce a term, 'quale<sub>?</sub>' and ask:

Is  $\text{quale}_{?} = \text{quale}_{\text{RED}}$ ?

Mary's concept  $\text{quale}_{?}$  seems like a good candidate for a *phenomenal concept*, for it is tied to her current experiences of the fire engine and the fire hydrant; it is the type of colour experience of which those two colour experiences are instances. Many philosophers put great weight on such phenomenal concepts, and they are useful in considering the fine structure of Mary's cognitive states at the moment of liberation. But such concepts are by their nature temporary, like the perceptual buffers involved in recognizing external objects.

Since  $\text{quale}_{?}$  is  $\text{quale}_{\text{RED}}$ , Mary is thinking about the same type of experience in two different ways when she uses the two terms in thought or language. The referential relations are quite different; she is related to the quale in two quite different ways. On the one hand, it is the quale that two of her current experiences exemplify. On the other, it is the quale that her textbooks referred to, and identified as the type of experience normal people have in favourable light when they see red objects. Her conceptions of the two are different. She believes that  $\text{quale}_{\text{RED}}$  is the one people

with normal vision have when they see red things in favourable light, and she also believes—since my version of Mary is a physicalist—that  $\text{quale}_{\text{RED}}$  is a type of brain state:  $B_{52}$  to be precise; that to experience  $\text{quale}_{\text{RED}}$  is to be in brain state  $B_{52}$ . She believes  $\text{quale}_?$  is the type of two of her present colour experiences, and is the quale caused in her in her present conditions by the colours of the surfaces of the fire hydrant and the fire engine. She also believes that *if* her vision is normal, and the present lighting is favourable, and this fire hydrant is, as most of them are, painted red, *then*  $\text{quale}_?$  is  $\text{quale}_{\text{RED}}$  and having  $\text{quale}_{\text{RED}}$  is being in  $B_{52}$ . Once she is confident that her vision is normal and conditions are favourable and no one has repainted the local fire hydrants and fire engines to fool her, she will believe that  $\text{quale}_? = \text{quale}_{\text{RED}}$ .

Once she draws this inference, she knows something new about  $\text{quale}_{\text{RED}}$ , that she didn't know in the black and white room; namely, that it is the type of colour experience exemplified by her current experiences of the colours of the fire hydrant and the fire engine. This is a relational fact about particular experiences that hadn't occurred when she was in the black and white room; it is not knowledge of some new fundamental property of  $\text{quale}_{\text{RED}}$  that was of necessity missed by her physicalist texts.

Perhaps Mary is so brilliant that she was able to predict that upon leaving the black and white room she would see a fire hydrant, grass, and a fire engine, all with their normal colours, in favourable light. She can introduce terms for the predicted experiences, say  $E_H$  and  $E_E$ . So she had a way of referring to and thinking about the experiences she is now having and their common quale,  $\text{quale}_{\text{RED}}$ , before having them. But the referential relations, in virtue of which she was able to refer to these things and talk about them, are quite different than the referential relations that enable her to think and talk about them as 'this<sub>i</sub> colour experience' (attending to the hydrant), 'this<sub>i</sub> colour experience' (attending to the engine), and ' $\text{quale}_?$ '. She can predict, while in the black and white room, that she will have a phenomenal concept, but she cannot yet think of it as one she is having. She can ask herself whether this<sub>i</sub> colour experience is  $E_H$ , this<sub>i</sub> colour experience is  $E_E$ , and whether  $\text{quale}_?$  is the one she predicted,  $\text{quale}_{\text{RED}}$ . She can figure out that they are. So she still has new knowledge that she didn't have before.

Does anything in all of this give Mary a reason to abandon her physicalist view, that  $\text{quale}_{\text{RED}}$  is  $B_{52}$ ? I cannot see that it does. If her physicalist views are correct, and she is wearing a new-fangled autocerebroscope, that produces visual images of the goings-on in her brain as she had her experiences, she would have to grant that she was seeing the very experiences she was having. She might think, 'Goodness, *seeing* an experience of  $\text{quale}_{\text{RED}}$  through an autocerebroscope (which I've never done before) is certainly nothing like *having* an experience of type  $\text{quale}_{\text{RED}}$  (which I've never had before). But then, why should it be?'

### 12.2.2 Locating Mary's Knowledge

Variations on the Mary story have been proposed, in order to produce the insights that will lead those of us unconvinced by the original story to dualism. But, actually, most of the more interesting and dramatic aspects of the various Mary stories

seem irrelevant. The basic point is well made by Feigl's original autocerebroscope fantasy. Here is a slight variation on it that may prove helpful. Suppose Mary is never imprisoned and has experienced a normal range of colours by the time she makes it to graduate school to work on colour vision and brain states. She has good colour vision, good colour memory, and a good command of the standard names of dozens of colours and shades. Then she learns as much as you please about the various brain states involved in colour vision. Suppose she has clearly before her mind what it's like to experience six colours: red, yellow, blue, orange, brown, and mauve. And she has, on her computer, complete physical descriptions of the six brain states that correspond to the state one is in when one perceives those colours. This is a sophisticated hyper-linked set-up, so she can zoom in on pictures, x-rays, sonograms, or whatever else you want. She can do in a virtual way what Leibniz imagined doing; she can enlarge the relevant parts of the brain and walk in and look around them. So far so good. However, the final interface between this sophisticated program and natural language, the identification of the scientifically described brain states in terms of ordinary colour words, is not in Mary's native Australian but in Cantonese, which she doesn't know. Now, can she match up the colours, or, more precisely, the quale that she associates with seeing the colours, with the physical states that she is in when she perceives those colours? I can't see how she could. When her Cantonese roommate returns home, and deciphers the names, Mary will learn something.

Here is the situation as I see it. Let's take mauve as our example. Mary has a memory of seeing mauve, she knows what it is like to see mauve; and, if you like, she can generate an actual experience of seeing mauve by pulling her mauve scarf from the drawer. So she has a concept, as phenomenal a concept as could be wished for, of the state one is in when one sees mauve. And she has a concept of the brain state people are in when they see mauve things in ordinary conditions, whose scientific name is  $BS_{\mu}$ . Between her and her computer program the whole of brain science is at her fingertips; she can look at actual pictures of  $BS_{\mu}$ ; she can look at helpful diagrams of it; push a button and the chemical composition of the various stuff involved in such states will pop up annoyingly; and so forth. These two concepts are both concepts of the same physical state, according to the physicalist; and, in addition, if she got her scarf from the drawer it is the state she is in. Does this mean that the physicalist should expect her to figure out, before her room-mate comes home to decipher the Cantonese, that the two concepts are concepts of the same state?

There is no reason the physicalist should suppose this. What magic would drive a mental identity sign between two such different concepts as these, even for someone as brilliant as we suppose Mary to be? Neither concept is defined in terms of the other; neither is introduced in terms of the other; neither makes reference to the other; they have no common parts. At this point the knowledge argument says: *but then, what does she learn when her room-mate comes home?* But the answer is clear. It's just the difference between having two concepts that are, in fact, of the same thing, and two concepts that are required by that internal identity sign to be of the same thing. She learns that the type of experience she is having, and so the type of experience her phenomenal concept is of, is the type of experience that her other,

brain-science-based concept is of; namely, the mauve quale. The contents of her doxastic states change, in that the truth of her total doxastic state, abstracted from the referential relations of her concepts, requires that the two concepts are concepts of the same type of experience; that is,  $BS_{\mu}$ ; that is, the experience one has when one sees mauve. That is how her beliefs change, and how her knowledge changes.

## REFERENCES

- Bach, K. (1981), 'An Analysis of Self-deception', *Philosophy and Phenomenological Research*, 41: 351–70.
- Block, N. (2007), 'Max Black's Objection to Mind-Body Identity', in T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism* (Oxford: Oxford University Press), 249–306.
- Chalmers, D. (1996), *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press).
- Feigl, H. (1967), *The 'Mental' and the 'Physical': The Essay and a Postscript* (Minneapolis, Minn.: University of Minnesota Press).
- Hall, L. (1993), 'Individualism, Mental Content and Cognitive Science', Ph.D. diss. (Stanford University).
- Jackson, F. (1986), 'What Mary Didn't Know', *Journal of Philosophy*, 85: 291–5.
- Malcolm, N. (1970), 'Scientific Materialism and the Identity Theory', in C. V. Borst (ed.), *The Mind/Brain Identity Theory* (New York: Macmillan), 171–80.
- Nagel, T. (1974), 'What is it Like to Be a Bat?', *Philosophical Review*, 83: 435–50.
- Perry, J. (2001), *Knowledge, Possibility and Consciousness* (Cambridge, Mass.: MIT Press).
- (2002), *Identity, Personal Identity, and the Self* (Indianapolis, Ind.: Hackett).
- (2003), 'The Subject Matter Fallacy', *Journal of Applied Logic*, 1: 93–105.
- Perry, J., and Israel, D. (1981), 'Fodor and Psychological Explanations', in B. Loewer and G. Rey (eds.), *Meaning in Mind* (Oxford: Blackwell), 165–80; repr. in Perry, *The Problem of the Essential Indexical and Other Essays*, expanded edn. (Stanford, Calif.: CSLI, 2000), 301–22.
- Searle, J. (1983), *Intentionality: An Essay in the Philosophy of Mind* (Cambridge: Cambridge University Press).