

Why the question of animal consciousness might not matter very much

Peter Carruthers

ABSTRACT: According to higher-order thought accounts of phenomenal consciousness (e.g. Carruthers, 2000) it is unlikely that many non-human animals undergo phenomenally conscious experiences. Many people believe that this result would have deep and far-reaching consequences. More specifically, they believe that the absence of phenomenal consciousness from the rest of the animal kingdom must mark a radical and theoretically significant divide between ourselves and other animals, with important implications for comparative psychology. I shall argue that this belief is mistaken. Since phenomenal consciousness might be *almost* epiphenomenal in its functioning within human cognition, its absence in animals may signify only relatively trivial differences in cognitive architecture. Our temptation to think otherwise arises partly as a side-effect of imaginative identification with animal experiences, and partly from mistaken beliefs concerning the aspects of common-sense psychology that carry the main explanatory burden, whether applied to humans or to non-human animals.

1 Introduction

As is now familiar, a number of distinct kinds of consciousness can be distinguished (Rosenthal, 1986; Block, 1995; Lycan, 1996), and in many of these senses it isn't problematic that consciousness should be attributed to animals. Many animals are sometimes conscious (awake as opposed to asleep or comatose); and many animals are conscious *of* things in their environment. (Even insects perceive things and believe things, I would argue; see Tye, 1997; Carruthers, 2004c.) But our concern, here, is with *phenomenal* consciousness. Phenomenally conscious states are states that are *like something* to undergo; they are states with a distinctive subjective *feel* or phenomenology; and they are states that each of us can immediately recognize in ourselves, 'straight off', without having to engage in any kind of inference. And in addition, phenomenal consciousness is supposed to constitute the 'hard problem' for cognitive science to explain (Chalmers, 1996).

Not everyone agrees that the problem is all that hard, however. Representationalist theories, in particular, claim that phenomenal consciousness can be reductively explained in terms of some combination of intentional content and functional role. *First-order* theories of the sort defended by Dretske (1995) and Tye (1995, 2000) claim that phenomenal consciousness consists in a certain sort of intentional content (*analog*, or *non-conceptual*)¹ being available to the belief-forming and decision-making systems of the creature in question. Higher-order theories claim, in contrast, that phenomenal consciousness consists in the targeting of analog first-order perceptual states by *higher-order* states – either by higher-order thoughts (Rosenthal, 1993), or by higher-order percepts (Lycan, 1996).

In the version of higher-order thought account developed and defended by Carruthers (2000), perceptual states (and more generally, all states with analog content such as images and emotions) become phenomenally conscious when they are available to a faculty of higher-order thought (HOT), which is capable of framing beliefs about those very states. Because of such availability, and in virtue of the truth of some or other version of ‘consumer semantics’,² the states in question acquire a dual analog content. One and the same percept is *both* an analog representation of *green*, say, *and* an analog representation of *seeming green* or *experience of green*. And it is claimed that such an account can successfully explain all of the various supposedly-puzzling features of phenomenal consciousness.³

¹ *Analog* content is intentional content that is more fine-grained than any concepts we could possess (think of the fine shades of color that we can distinguish, but not remember), even if it is nevertheless imbued with, or shaped by, concepts. *Non-conceptual* content is intentional content that is not only fine-grained but independent of, and prior to, concept-deployment. See Carruthers (2000, chapter 5) for discussion.

² All forms of consumer semantics claim that the intentional content of a state depends, in part, on what the ‘down-stream’ systems that consume the state are capable of doing with it or inferring from it. Teleo-semantics is one form of consumer semantics (Millikan, 1984, 1989; Papineau, 1987, 1993). Functional or inferential role semantics is another (Loar, 1981; Block, 1986; McGinn, 1989; Peacocke, 1992).

³ *Inter alia* the account can explain: how phenomenally conscious experiences have a *subjective aspect* to them (their higher-order analog content); how they can be available to us for introspective recognition (their higher-order contents will present them to us in much the same sort of way that our first-order perceptions present to us a world of colors and shapes); how we can come to have *purely* recognitional concepts of our phenomenally conscious experiences (these will be recognitional concepts grounded in the higher-order analog contents of our experiences, in something like the way that first-order recognitional concepts of color, say, are grounded in the first-order analog content of perception, only without any of the surrounding beliefs about the mechanisms by

If some or other first-order account of phenomenal consciousness is correct, then such a form of consciousness will be widespread in the animal kingdom, and will perhaps even be present in insects (Tye, 1997). If some form of higher-order thought account is correct, in contrast, then it seems unlikely that many animals besides ourselves will count as undergoing phenomenally conscious experiences. It is still hotly debated whether chimpanzees are capable of higher-order thought, for example, and powerful arguments can be adduced to the contrary (Povinelli, 2000). And if chimpanzees aren't so capable, then probably *no* animals besides ourselves are phenomenally conscious, according to a higher-order theorist. I should stress, however, that I have no axe to grind here. If chimpanzees (or dolphins, or whatever) should turn out to have higher-order thoughts, then all well and good – they will turn out to be phenomenally conscious also. The point is just that for a higher-order thought theorist, the capacity for phenomenal consciousness is conditional on a capacity for higher-order thinking; and the latter capacity is unlikely to be widespread amongst non-human animals.

In what ways do these debates about phenomenal consciousness matter? Obviously they are of deep theoretical significance. For what it at issue is how best to solve the 'hard problem' of consciousness. But do they matter practically, for our treatment of non-human animals? And do they matter for psychology, having a bearing on the extent to which we can see continuities between the mental processes of ourselves and other animals? I claim not, in response to both of these questions. Although the issue of animal consciousness might be of deep importance for those of us trying to construct reductive theories of phenomenal consciousness, it has little bearing on anything else of significance.

Some people will resist the conclusion that most non-human animals lack phenomenally conscious states, on the grounds that this must then mean that the lives and sufferings of such animals are of no direct moral importance. Believing that the sufferings of a dog, or a cat, or a chicken are deserving of our sympathy and moral concern, such people will reject accounts of

means of which such recognitions are effected); why phenomenally conscious properties should seem especially *ineffable* (when we deploy purely recognitional concepts of them they cannot be further described, and – unlike public colors and textures – they cannot be exhibited to others); and why there should seem to be a pervasive *explanatory gap* between the higher-order account itself and the qualities of our phenomenal experiences (again, the blame falls squarely on our purely recognitional concepts). See Carruthers (2000, 2004a, 2004b) for extended discussion.

phenomenal consciousness that (they think) suggest otherwise. However, while it may *seem* inevitable that it can't be appropriate to feel sympathy for creatures that are incapable of phenomenal consciousness, this may just be because we have a difficult time imagining pains and disappointments that aren't phenomenally conscious ones. Indeed, it is arguable that the moral landscape can remain unchanged, no matter whether first-order or higher-order accounts of phenomenal consciousness should turn out to be correct.

I have developed these points at some length elsewhere (Carruthers, 1999, forthcoming), and so will not pursue them here. The focus of the present paper will be entirely on the second of the questions raised above: does the thesis that non-human animals lack phenomenally conscious states have important implications for comparative psychology? I shall begin by considering the implications for common-sense psychology, turning to discuss the alleged implications for scientific psychology in later sections.

2 Consciousness and common sense

The thesis that few if any non-human animals are phenomenally conscious conflicts with common-sense belief, of course. Most of us believe that many of the species of animal around us enjoy phenomenally conscious experiences. But what, if anything, serves to ground this belief? It is hard to see what evidence we could have to support it. Indeed, Carruthers (2000) argues that the belief in question is merely the product of a kind of cognitive illusion. Rightly believing that non-human animals are subject to perceptual states, and wondering what those states are like, we naturally represent them to ourselves in imagination 'from the inside'. But in doing so we inevitably find ourselves representing a state that has all of the features of phenomenal consciousness, because *the images* that we form have those properties. Since we have no idea how to imagine, from the inside, a perceptual state that *isn't* phenomenally conscious, the fact that we find ourselves thinking that the mental states of non-human animals must be phenomenally conscious ones is little to be wondered at.

Our belief that non-human animals enjoy phenomenally conscious states may actually be quite groundless, then. It may merely be a product of a cognitive illusion. In which case the fact that higher-order thought theories of phenomenal consciousness are in conflict with this belief won't really count against the acceptability of those theories. And it will cost us little or nothing to alter our common-sense attitudes accordingly, if one of those theories should be accepted as

successful, explaining what needs to be explained about the character of our experience.

Lurz (2002) argues, however, that there is more to our attitudes towards animals than mere illusion, or the effects of over-extended imagination. He thinks that our real reason for believing that non-human animals enjoy phenomenally conscious experiences goes something like this. When we make use of our common-sense psychology, we deploy attributions of phenomenally conscious experiences to other people when explaining their behavior. And this doesn't just happen when we explain people's verbal behavior, either, or their higher-order beliefs about their own experiences. On the contrary, we also use such attributions when explaining the sorts of behaviors that we share with non-human animals. When I explain why someone picked out one tomato from the vegetable rack rather than another, for example, I might believe that this happened because she was subject to a phenomenally conscious experience of the luscious red color. And when I explain why someone froze in fear when meeting a grizzly bear in the woods, I might believe that this happened because she had a phenomenally conscious perception of the bear's ferocious aspect.

Similarly, then, when explaining why an animal picks out one object rather than another when foraging, or when explaining an animal's fear when confronting a predator – I should attribute phenomenally conscious experiences to the animal to do the explanatory work. And then I have much the same sort of reason to ascribe phenomenally conscious states to animals as I have for attributing them to other people. And a theory of phenomenal consciousness that required us to think otherwise would be deeply revisionary of our folk-psychological practice. Or so Lurz (2002) argues.

There is another kind of illusion at work here, however, which one might call 'the *in virtue of* illusion'. Granted, when someone picks out a tomato by its color, she will generally have a phenomenally conscious experience of red. But it is another matter to claim that she chooses as she does *in virtue of* the phenomenally conscious properties of her experience. While we cite experiences that are, *de facto*, phenomenally conscious ones when explaining other people's behavior, it is quite another matter to believe that phenomenal consciousness plays any essential role in those explanations. Indeed, this is just what a higher-order theorist should deny. And doing so doesn't force any deep revision of folk psychology. It just requires us to be somewhat clearer about the properties in virtue of which our explanations are successful.

It isn't just common-sense practice that is vulnerable to the *in virtue of* illusion, either.

The sorts of functional-role accounts of phenomenal consciousness constructed by psychologists can run into the same problem. Consider the ‘global workspace’ model of consciousness proposed by Baars (1988, 1997), for example. On this account, phenomenally conscious mental events serve the function of broadcasting their contents to a wide range of other cognitive systems. And it is indeed true that phenomenally conscious events in humans are *de facto* widely available. However, it is another matter to claim that those events are thus available *in virtue of* being phenomenally conscious; and a higher-order thought theorist will deny that this is so. On the contrary, animals lacking a capacity for higher-order thought may well have states that are globally broadcast. And such states only come to have the distinctive properties of phenomenal consciousness when embedded in an architecture that contains, in addition, a faculty of higher-order thought to which those states are available.

All of this requires some setting up to become convincing, however. In what follows I shall first present and discuss the ‘dual visual systems’ hypothesis of Milner and Goodale (1995). I shall then use this as a springboard from which to explain the *in virtue of* illusion, before discussing a couple of case-studies from the comparative psychology literature. I shall then close by considering the implications of two different accounts of our faculty of higher-order thought. The overall conclusion will be as advertised: very little of significance turns on the question whether or not non-human animals undergo phenomenally conscious experiences.

3 Dual visual systems and phenomenal consciousness

According to Milner and Goodale (1995), the human mind / brain contains two visual systems that are functionally and anatomically distinct; and indeed, there is now a wealth of evidence that this is so (Jacob and Jeannerod, 2003). The dorsal system is located in the parietal lobes and is concerned with the on-line detailed guidance of movement. The ventral system is located in the temporal lobes and serves to underpin conceptual thought and planning in relation to the perceived environment. Each receives its primary input from area V1 at the posterior of the cortex, although the dorsal system also receives significant projections from other sites. The dorsal system operates with a set of body-centered or limb-centered spatial co-ordinates, it is fast, and it has a memory window of just two seconds. The ventral system uses allocentric or object-centered spatial co-ordinates, it is slower, and it gives rise to both medium and long-term memories. Importantly for our purposes, the outputs of the dorsal system are unconscious, while

those of the ventral system are phenomenally conscious (in humans). Finally, homologous systems are widespread in the animal kingdom, being common to all mammals, at least.

On this account, the phenomenally conscious experiences that I enjoy when acting are *not* the percepts that guide the details of my movements on-line.⁴ Rather, the phenomenally conscious percepts produced by the ventral system are the ones that give rise to my beliefs about my immediate environment, that ground my desires for perceived items ('I want *that* one') and that figure in my plans in respect of my environment ('I'll go *that* way and pick up *that* one'). But my planning only guides my actions indirectly, by selecting from amongst a data-base of action schemata. The latter then directly cause my movements, with the detailed execution of those movements being guided by the percepts generated by the dorsal system. The basic two-systems arrangement is depicted in figure 1.

Insert figure 1 about here.

The dual visual systems hypothesis provides the grounds for one of the main arguments against first-order accounts of phenomenal consciousness (of the sort defended by Dretske, 1995, and Tye, 1995), and in support of some sort of higher-order approach (Carruthers, 2000, 2004b). For looking just at the architecture depicted in figure 1, it is quite mysterious why the outputs of the dorsal system should be unconscious while those of the ventral system are conscious. Both generate analog intentional contents that represent the world as being a certain way to us. Admittedly, only the outputs of the ventral system are available for belief-formation and planning. But it remains unclear why this should make a difference. Why should such availability serve to confer on those (and only those) visual percepts a dimension of *subjectivity*?

A version of higher-order thought theory can provide the necessary explanation, in contrast (Carruthers, 2000, 2004b). Consciousness results because the ventral percepts are *also*

⁴ This is already enough to show us that common sense is vulnerable to the *in virtue of* illusion, since ordinary folk surely believe that the precise movements of my hand and fingers when I grasp a cup occur as they do in virtue of the phenomenally conscious experiences that I undergo at the time. But according to the dual visual systems hypothesis, this isn't so. Those movements are guided by the outputs of the dorsal system, which aren't conscious ones (although the *planning* and *monitoring* of the action is informed by the outputs of the ventral system, which *are* conscious).

available to a faculty of higher-order thought (HOT), as depicted in figure 2. By virtue of their availability to higher-order thought (and because of the truth of some or other version of consumer semantics), the perceptual outputs of the ventral system acquire a *dual* analog intentional content. At the same time as representing (in analog fashion) worldly redness, greenness, and so on, they *also* represent *seeming* redness, *experienced* greenness, and so forth (also analog), thus acquiring their distinctive subjective aspect.

Insert figure 2 about here.

For our purposes here, the point to stress is that a figure 1 architecture might be quite widespread in the animal kingdom. It may be common to all animals that are capable of thought, but that lack a faculty of higher-order thought (hence being incapable of phenomenal consciousness, for a higher-order theorist). In fact, the addition of a HOT faculty to the basic mammalian cognitive architecture might be the only relevant difference between ourselves and other animals. (Of course humans are also distinctive in possessing a language faculty, but this isn't relevant to issues concerned with phenomenal consciousness, I claim. See Carruthers, 2000, ch. 10.) In which case the correct explanation for all of the behaviors that we share with other animals will be exactly the same – these will be behaviors underpinned by first-order perceptual contents. The fact that those perceptual states *also* happen to be phenomenally conscious ones in the case of humans will be irrelevant to the explanation.

Consider the human who selects a tomato from the vegetable rack because of its luscious red color, then; and consider a monkey or any other fruit-eating color-sighted mammal making a similar choice. Although the experience of the human is phenomenally conscious, whereas the experience of the monkey probably isn't (according to a higher-order thought account), the explanations for their respective actions can be (and is) exactly the same. Each chooses as she does because she *sees* the luscious red color of the fruit, and judges in consequence that the fruit is ripe. The content of the *seeing* here, in so far as it is relevant to the explanation of the action, is entirely first-order in character. It is because the human sees the redness of the surface of the tomato (and hence enjoys a first-order analog content representing worldly redness), *not* because her percept also has the higher-order analog content *seeming redness*, that she chooses as she does. Likewise with the monkey.

What determines explanatory relevance, in general? And what makes it the case that the first-order analog content *red*, rather than the higher-order analog content *seeming red*, is what explains the human's behavior? The answer is that genuine explanations fall under laws, or nomic generalizations; and they should also generate true counterfactuals. The relevant psychological generalization, in this case, is surely something like this: *seeing a red fruit, and believing that redness is a sign of ripeness, and wanting ripe fruit, one will choose the red one*. The acts of both human and monkey fall under this first-order intentional law; and it is reasonable to believe that if the redness of the tomato had *not* been perceived, then neither human nor monkey would have acted as she did. In contrast, it seems reasonable to suppose that even if the human's percept had *lacked* the higher-order analog content *seeming red* (perhaps because the human had lacked a higher-order thought faculty), she would still have chosen and acted just as she did.⁵

Recall Lurz's (2002) argument. Since we explain the actions of humans by attributing phenomenally conscious experiences to them, and since non-human animals perform many of the same actions, he says that we have good reasons for believing that such animals, too, undergo phenomenally conscious experiences. The error in the argument is now plain to see. Granted, I explain the actions of a human by attributing to her experiences that are, *de facto*, phenomenally conscious ones. But on a higher-order thought account of phenomenal consciousness, their phenomenally conscious status will actually play no explanatory role (in most cases, at least; see below). Rather, the explanatory burden is carried by the purely first-order, not necessarily phenomenally conscious, character of those states. And these are just the sorts of perceptual

⁵ We don't have any theory-independent evidence for the truth of this counter-factual, of course. To the extent that it is warranted, it is supported by (a) the two visual systems hypothesis, together with (b) the reasons that there are for accepting a higher-order thought account of phenomenal consciousness. Note, however, that if both of these assumptions are granted, and it is also true that autism is characterized by an incapacity for higher-order thought (Baron-Cohen, 1995), then an autistic person selecting a fruit on the basis of its color will be a true instance of the counterfactual – this would be a case of someone lacking phenomenally conscious experiences, whose choices are nevertheless guided by color information, just as ours are. (But of course we can't verify that this is so by *asking* the autistic person whether or not they are phenomenally blind – for lacking a capacity for higher-order thought, they will be incapable of understanding the question. And note, too, that autism is a syndrome admitting of a wide spectrum of degrees and kinds of severity, so any claim about the absence of phenomenal consciousness in autism would in any case have to be suitably nuanced.)

states that non-human animals can share. So the cross-species explanations can be the very same, just as Lurz claims, but consistent with the truth of a higher-order theory that withholds phenomenal consciousness from most species of animal.

Similar points can now be made in respect of Baars' (1988, 1997) 'global work-space model' of phenomenal consciousness. Granted, the phenomenally conscious outputs of ventral vision, in figure 2, are available to a wide range of cognitive systems (belief-generating systems, memory systems, desire-generating systems, and practical reasoning). But on a higher-order thought account it isn't by virtue of being so available that those outputs are phenomenally conscious. On the contrary, the outputs of the ventral visual system in other mammals, too, will be almost as widely available (see figure 1). But because they aren't available to a higher-order thought faculty, they will lack a dimension of subjectivity, and so will fail to be phenomenally conscious.

Is phenomenal consciousness epiphenomenal, then? Do phenomenally conscious properties play *no* role in explaining human actions? Of course not. Phenomenal consciousness is intrinsically involved when we think *about* our experiences, as such, or when we *describe* our experiences to others. And phenomenal consciousness is implicated whenever we draw a distinction between the way things *are* and the way they *seem* or *appear*. So phenomenally conscious properties will have a genuine explanatory role whenever our actions manifest higher-order thought about our experiences, or whenever they depend upon the distinction between *is* and *seems*. But there is no argument from this to the phenomenal consciousness of non-human animals. For if such animals are incapable of higher-order thought, then by hypothesis they won't be engaging in the relevant behaviors either.

The upshot of this section is that very little of significance for comparative psychology need follow from the fact that phenomenal consciousness is denied to many non-human animals by higher-order thought theories of consciousness. All types of first-order experience, first-order cognitive state, and first-order cognitive process can be shared across (many) species, and can be common to humans and other animals. The only predicted differences would concern thoughts and behaviors that in one way or another depend upon higher-order thought processes. Certainly there is no radical Cartesian divide here, between genuinely minded humans and mere mindless automatons, or anything of that sort. On the contrary, non-human animals can be full subjects of (first-order) mindedness.

I shall now reinforce and elaborate upon these points by discussing two case-studies: one concerning blindsight in monkeys, and the other concerning alleged meta-cognitive behaviors in other species besides ourselves.

4 Case-study (1): blindsight in monkeys

One of the things that can be explained within the framework of the two visual systems hypothesis is the striking phenomenon of *blindsight*. As is now familiar to most, these are cases involving people who have had a region of the primary visual projection area, V1, damaged, and who consequently become blind in a portion of their visual field (at least as regards conscious experience). While they are aware of seeing nothing within that region, however, they prove remarkably good at *guessing* the orientation of an illuminated grating, at tracing the movement of a light across a screen within their blind field, or at reaching out to grasp an object on their blind side (Weiskrantz, 1986, 1997; Marcel, 1998).

The explanation of these findings is as follows. Besides the main projection from the retina to V1 (via the lateral geniculate nucleus in the thalamus), there is also a substantial, and phylogenetically older, projection to the superior colliculus in the mid-brain. (In fact this sub-cortical pathway alone is as large as the whole of the auditory nerve.) And from the superior colliculus there are substantial projections, not only to motor and pre-motor neurons in the brainstem and spinal cord, but also to the parietal cortex, which is the hypothesized site of the dorsal pathway (charged with the control of movement). Yet there are no projections to the temporal cortex, which is the hypothesized site of the ventral system (and which is intimately connected with conscious experience). So even with cortical area V1 completely removed, substantial quantities of visual information should be available in parietal cortex for the control of movement.⁶

Now, it has been known for some time that monkeys who have had area V1 of visual cortex surgically removed are nevertheless capable of a range of visual discriminations, and a range of visually-guided actions. But what Cowie and Stoerig have shown is that the results are

⁶ How, then, does such information enable patients, not only to point and grasp appropriately, but also to answer questions concerning the orientation of a grating? Milner and Goodale (1995) suggest that such patients may be influenced by subtle behavioral and motor-programming cues of which they lack conscious awareness, e.g. to begin tracing out the line of orientation of the grating with a hand or with the eyes.

analogous to human blindsight, at least in the sense that stimuli in the blind field are judged by the monkey not to be present (Cowie and Stoerig, 1995; Stoerig and Cowie, 1997). A monkey with one half of area V1 missing can be trained, first, to press a button following a warning tone if a light does *not* illuminate on a screen presented to her sighted field. When a light or a bright moving stimulus is then presented to her blind field, she presses for ‘not seen’, thus indicating that she lacks awareness of the stimulus – and this despite it being one that she can be shown to discriminate in other conditions.

Does the discovery of blindsight in monkeys show that monkey visual experiences are, normally, phenomenally conscious? Many people seem to assume so. For, like Dretske (1995), they use evidence of the residual capacities of blindsighted monkeys when speculating about the function that phenomenal consciousness may have. They assume that blindsighted monkeys must have lost whatever blindsighted humans have lost, and therefore that normally sighted monkeys must be phenomenally conscious, given that blindsighted humans have lost their capacity for phenomenally conscious visual experience. But this is a mistake. Despite the common presence of (at least) two functionally-distinct visual pathways in monkeys and humans, there are other differences between the species that may be sufficient to make a difference in phenomenal consciousness. In particular, humans exemplify the figure 2 architecture, whereas monkeys (probably) only exemplify the figure 1 architecture.⁷

The point is that when monkeys press the key for ‘not seen’ they are, more strictly, only signaling that *a light* isn’t present in front of them. There isn’t really any reason for claiming that they are making a higher-order comment on their lack of *awareness* of a light. And conversely, when a monkey omits to press a key because a light *is* seen, it is strictly only expressing a judgment that a light is present, not that it is currently *experiencing* a light. So there is nothing in the data that can’t be explained in purely first-order terms, by deploying the figure 1 architecture. And so from the perspective of a higher-order thought account of phenomenal consciousness, too, there is nothing in the data to suggest that monkeys are, normally, phenomenally conscious.

5 Case-study (2): meta-cognitive processes in animals?

One of the morals of our discussion so far has been that the denial of consciousness to animals

⁷ Let me reiterate that I have no axe to grind here. I am happy to let the facts concerning higher-order thought in monkeys and apes fall where they may.

needn't imply any sort of Cartesian divide between humans and the rest of the animal kingdom; and nor, therefore, need it have significant consequences for comparative psychology. In the present section I shall argue that the cognitive divide between humans and other animals might be even smaller than might appear thus far. For even processes in humans that we are strongly disposed to describe in meta-cognitive terms (that is, in such a way as to involve higher-order thought, and hence phenomenal consciousness), may actually be so described erroneously. The underlying processes may really be first-order ones, shared with other animals. They may merely be given a meta-cognitive gloss by the human self-interpretation system (our higher-order thought faculty).

Consider some of the evidence of uncertainty monitoring in other species of animal, recently discussed at length by Smith *et al.* (2003). The authors show how some animals in a state of uncertainty in a forced-choice situation (the pattern is dense / the pattern isn't dense; the tone is a high one / the tone isn't high) will choose adaptively in favor of some third alternative when given the chance (e.g. opting to move on to another trial without a timeout). They show how the response profiles of monkeys and dolphins in such circumstances closely parallel those of humans (even displaying similar degrees of individual variation in strategy). And they note that in these sorts of situations humans will explain their own behaviors as resulting from their awareness of their own state of uncertainty (i.e. humans will couch their self-explanations in meta-cognitive terms). Since uncertainty behaviors in humans display both higher-order thought and phenomenal consciousness, Smith *et al.* (2003) argue, it is likely that similar states are present in monkeys and dolphins in such circumstances.

Uncertainty monitoring can be explained without resort to attributions of meta-cognitive processes, however. Suppose, first, that beliefs come in *degrees*. Just as desires can be more or less strong or powerful, so something can be believed with greater or lesser confidence. This assumption is routinely made by philosophers, at least.⁸ So both animals and humans in the

⁸ Some people also claim, however, that there is a distinct kind of ungraded 'flat-out' belief, which may be dependent upon language, and which consists in an explicit *commitment* to assent to a sentence / proposition and to think and reason as if that proposition were true. These states are variously called 'opinions' or 'acceptances' and belong within so-called 'dual-process' models of human cognition (Dennett, 1979; Cohen, 1993; Evans and Over, 1996; Frankish, 1998, 2004). Note that dual-process models imply that a capacity for *opinion* is dependent upon a capacity for higher-order thought, which is also what is responsible for phenomenal consciousness, according to higher-order theories of the latter. So creatures incapable of higher-order thought will not only lack phenomenal

described experimental situations will have a particular *degree* of belief that a tone is high, or that a pattern is dense, say.

Nothing meta-cognitive is required for degrees of belief and desire. Having a strong desire doesn't mean that the animal believes *of* itself that it has a desire with a certain strength. Rather, it has a desire that is apt to beat out other desires in the competition to control behavior, and which is also apt to have further cognitive and physiological effects of a distinctive sort (e.g. increased heart-rate). Equally, having a strong belief doesn't mean that the animal believes itself to have a belief with a certain high degree of warrant. Rather, the animal has a belief on which it is more likely to act, all else being equal (and especially given equal strengths of competing desires). And degrees of belief might be realized in different levels of activation in the appropriate neural network, for example.

Suppose, second, that in circumstances in which an animal has conflicting desires of roughly equal strength (to get a food pellet now; to avoid *not* getting a food pellet now – i.e. to avoid a timeout), and where it also has conflicting beliefs of equal strength (the pattern is dense; the pattern isn't dense), the animal will (a) be in a state of some anxiety, which will be experienced as aversive, and (b) will be disposed to engage in behaviors of a sort that tend to eradicate such conflicts of belief (e.g. by attending more closely, by moving to change its angle of view by shifting its head from side to side in the way that the dolphins in the experiments reported by Smith *et al.* did, and so on). Call such a state 'an uncertainty state'.

Again, nothing meta-cognitive need be involved. A state of uncertainty is a state that is caused when the animal is equally disposed to act in two contrary ways, and which has a variety of characteristic further effects of cognitive, affective, physiological, and behavioral sorts (e.g. engaging in information-seeking behaviors). And that an animal engages in behaviors designed (by evolution or by learning) to elicit new information doesn't mean that the animal represents itself as lacking a sufficient degree of belief. It just means that these behaviors are ones that have been sculpted to issue in changed degrees of belief (changes that are apt, in turn, to diminish an aversive state of anxiety).

What is an animal to do when in a state of conflict, and when information-seeking

consciousness, but will also be incapable of opinions. And then to the extent that dual-process models are necessary and important for describing human behavior, to that extent there will be significant differences between the psychologies of humans and other animals.

behaviors are failing to resolve that conflict? Plainly it needs some heuristic to enable it to reach a decision and move on, or (like Buridan's ass) it will remain perpetually frozen. The simplest such heuristic is: *when in a state of that sort, choose at random*. It seems that this is a heuristic that many animals (and some humans) employ. A more complex heuristic is: *when in a state of that sort, opt for a less-favored third alternative if you can*. This seems to be the heuristic adopted by some monkeys, dolphins, and humans. Why is this heuristic more complex? Because it requires the animal to represent and to factor into its decision-making, not just the two alternatives between which it faces the initial choice, but also a third option. But what would explain the individual differences in the use of this heuristic? Perhaps differing degrees of risk-aversion.⁹

Once again, nothing meta-cognitive is required in order for these heuristics to operate. Granted, the animal needs to have some way of telling when it is in a state of the required sort. (That is, it needs a way of telling when it is in the uncertainty state that is the trigger for the heuristic to apply, picked out by the antecedent clauses in the conditionals above.) But this doesn't mean that the animal has to conceptualize the state *as* a state of uncertainty, or as a state in which a conflict of belief remains unresolved. Rather, it just has to have some way of reliably picking out a state that *is* a state of uncertainty. And here its continuing state of anxiety and/or its awareness of its own distinctive information-seeking behaviors (e.g. its side-to-side head movements) would be sufficient.

But how is it, then, that humans will describe these situations in meta-cognitive terms, if nothing meta-cognitive is involved in the decision-making process? The answer is that humans are chronically self-interpretative creatures, by virtue of possessing a faculty of higher-order thought. Humans will correctly conceptualize the state that they are in as one of uncertainty – i.e. as one in which they lack sufficient information to adjudicate between two conflicting beliefs. And they generally interpret themselves to be choosing as they do *because* they are *aware* that they are in a state of uncertainty. But from the fact that their explanations of their feelings and behaviors are meta-cognitive, it doesn't follow that the decision-making process itself is a meta-cognitive one. And indeed, there is good reason to think that much of the self-interpretative cognition of humans is *merely* interpretative, not reflective of an underlying meta-cognitive

⁹ Amongst human subjects, males are more likely to choose randomly when uncertain, females are more likely to opt for a new trial without any penalty (or reward). See Smith *et al.*, 2003.

decision-making process. (For recent reviews of the evidence, see Gazzaniga, 1998; Wilson, 2002.)

A quarter century of research, beginning with Nisbett and Wilson's seminal article (1977), has uncovered a wealth of evidence that many of our beliefs about the causes of our own behaviors are actually *confabulated*. Far from having direct access to our own reasoning and decision-making processes, what we actually do in many cases is interpret our own behaviors, ascribing mental states to ourselves in much the same sort of way that we might ascribe them to another person. And where the true causes of behavior are obscure to common-sense psychology, such self-attributions will frequently be false. But we aren't aware that what we are doing is confabulating. We tend to operate with a simplified model of the mind, according to which our own thought processes are transparent to us; and so we tend to think, too, that the higher-order awareness that we report is intrinsic to the process reported on.

Consider one of the examples of confabulation described by Gazzaniga (1998). A split-brain patient gets up and begins to walk across the room because the instruction, 'Walk!' is flashed on a screen in his left visual field (hence being available only to the right hemisphere). But when asked what he is doing, the left hemisphere responds, 'I want to get a drink from the fridge.' (In most people the left hemisphere houses both the language centers, which answer the question, and the main elements of the theory of mind faculty, which generates the self-interpretation.) Although the explanation is plainly confabulated, the patient isn't aware that this is so. On the contrary, he takes himself to be acting as he does *because* he is *aware* of a desire for a drink (i.e. as a result of a meta-cognitive state). Yet the causes of his action were actually first-order ones: he got up to walk because he comprehended a written instruction to do so, delivered by a person of some authority (the experimenter).

I don't mean to be claiming, of course, that our meta-cognitive processes are entirely epiphenomenal, and have no effect on our behavior whatsoever. On the contrary, the higher-order interpretative narratives that we weave for ourselves can constrain us and guide us in various ways.¹⁰ One constraint is provided by the desire to act consistently with the self-image that we have constructed. I would be prepared to bet, for example, that having formulated his confabulated explanation, Gazzaniga's patient did indeed go to the fridge to get a drink. Since he

¹⁰ This is one of the insights that is developed by 'dual process' models of human cognitive processes, discussed briefly in footnote 8 above.

thinks that this is what he wants, this is what he would probably do; and the explanation becomes self-fulfilling. My point is just that much *less* of our behavior is actually caused by conscious thought-processes than we are intuitively inclined to believe.

The moral of the discussion is as follows. Even if a higher-order thought theory of phenomenal consciousness is accepted, and in consequence phenomenally conscious experiences and thought processes are denied to most species of animal, it can still be true that there are vast swathes of behavior common to humans and other animals that can receive identical psychological explanations. The fact that humans are strongly inclined to explain their own behaviors by self-ascribing *awareness* of their own perceptions or *awareness* of their own decision making (and so in terms that implicate phenomenal consciousness) should not be allowed to mislead us. For a great many of these behaviors will have cognitive explanations that are purely first-order, of the sort that we can share with other animals. And once again, the overall moral is that a higher-order thought theory of phenomenal consciousness turns out to have many fewer implications for comparative psychology than you might initially have been inclined to think.

6 Two models of the HOT faculty

It might be objected that we have moved too hastily from the evidence that humans *sometimes* confabulate when explaining their own behavior, to the claim that human reasoning and decision-making processes will *generally* be inaccessible to, and uninfluenced by, higher-order thought.¹¹ And indeed, such a move is too hasty. There are two different models of the way in which our higher-order thought faculty is embedded into the architecture of our cognition, and connected with other systems, which carry different implications on just this point. It is worth briefly sketching these models, and articulating some reasons for preferring one to the other of them. For on this question will turn the extent of the cognitive differences that we should expect to find between ourselves and those animals that lack a capacity for higher-order thought.

On one model – which has been implicitly assumed up to now, and is the one represented in figure 2 – the higher-order thought faculty only has direct access to perceptual (and to quasi-perceptual, e.g. to imagistic or emotional) states. The explanation is the standard one given in the

¹¹ I am grateful to an anonymous referee for this point, which prompted the discussion that is contained in the present section.

cognitive science literature, namely that the faculty in question evolved for purposes of predicting and explaining (and hence also manipulating) the behavior of other people (Byrne and Whiten, 1988, 1998). This required it to have access to perceptual input. The directing of this faculty upon oneself to explain one's own behavior, in contrast, is a secondary phenomenon, not under independent selection. Hence, while the higher-order thought faculty might have access to stored information (e.g. about previous actions) when constructing its representations, it has no access to the inferential *processes* that generate that information. Call this model, 'the mind-reading model', since it conceives the function of the higher-order thought faculty to be that of attributing mental states to others (and derivatively to oneself).

The mind-reading model needn't claim that the higher-order thought faculty has *no* access to the inferential processes that take place within the subject, of course. But that access will be limited to processes that in one way or another implicate sensory representations. For example, if some of our thinking is conducted consciously in imaged natural-language sentences, in 'inner speech', as Carruthers (1996) claims, then those processes of thinking will be both phenomenally conscious and available to higher-order thought. Indeed, the mind-reading model can mesh successfully with 'dual process' theories of human cognition (already mentioned in footnote 8), and for accounts like that of Dennett (1991), according to which there is a level of language-dependent mentality (the 'Joycean machine') which is imbued with higher-order thinking. For these processes, involving states that are quasi-perceptual in character, can be made available to higher-order thought in just the same way that perceptual contents are.

The alternative model – that might as well be called 'the self-monitoring model' – claims that at least one of the functions of the higher-order thought faculty is to monitor, to intervene in, and to correct where necessary, the subject's own reasoning processes. This then requires, of course, that the faculty should have *access to* those processes, as depicted in figure 3. The model might be framed in terms of one of two distinct evolutionary explanations. On the one hand it might be claimed that self-monitoring is the *primary* evolutionary function of a higher-order thought faculty – in which case we could expect to find higher-order thought in creatures that as yet lack any capacity to attribute mental states to others (this seems to be the position of Smith *et al.*, 2003). Or it might only be claimed, in contrast, that self-monitoring is an *additional* function of the higher-order thought faculty, that either evolved on the back of a prior mind-reading ability, or that co-evolved simultaneously with it (this seems to be the position of Shallice, 1988).

Insert figure 3 about here.

The self-monitoring model needn't claim that the higher-order thought faculty has complete and infallible access to the inferential processes of the subject, of course. So the model is consistent with the confabulation evidence reviewed in section 5. (Although we are still owed a story about *which* reasoning processes are available to the self-monitoring system and which aren't.) But it will claim that *much more* of human cognition will implicate higher-order thought than would be allowed by the mind-reading model. And then one would expect to find, in consequence, *much greater* cognitive differences between humans and those animals that lack a capacity for higher-order thought than the mind-reading model would predict.

These are large issues for cognitive science to address, of course, and they cannot be discussed adequately here. But it is worth articulating briefly why the mind-reading model is much the more plausible of the two. For if this model is correct, then we should expect the basic psychological differences between ourselves and other animals to be fairly small.¹² But even if it is wrong (and some version of the self-monitoring model is correct), those differences will still be far less than one might intuitively have been inclined to think. Certainly the absence of higher-order thought (and hence phenomenal consciousness, given a higher-order account of the latter) from non-human animals need not represent any sort of radical 'Cartesian divide' between our minds and theirs.

The fundamental point is this: in the case of mind-reading, we have a highly developed competence with obvious adaptive significance. Human beings are quite remarkably good at attributing mental states to others (and to themselves), and hence at engaging in the sorts of manipulative and/or co-operative behaviors whose success depends upon those attributions being accurate. And everyone agrees that this competence is part of the natural endowment of any

¹² Note that this isn't to claim that there will only be minor cognitive differences between humans and other animals *overall*. On the contrary, if some version of 'dual process' theory or Dennett's 'Joycean machine' is correct, then the overall differences might be quite large. But on these views, the higher-level processes will be *realized in* the operations of the lower-level systems that we mostly share with other animals (Frankish, 2004). The two main differences, that make the higher-level reasoning system possible, would consist in our possession of a higher-order thought faculty together with a language faculty.

normally-developing human being. In contrast, it is far from clear that we possess any very useful competence in the domain of self-monitoring. Of course we are good at *attributing* mental states to ourselves (absent the confabulation cases), but for a mind-reading theorist this is just a matter of our turning our mind-reading abilities upon ourselves. What is in question is whether we have the sort of highly-developed capacity to monitor, trouble-shoot, intervene in, and improve upon our own reasoning processes on-line, in the way that the self-monitoring model requires.

There is little reason to think that we possess any such natural competence. Indeed, naïve subjects are quite remarkably poor at distinguishing good sequences of reasoning from bad ones, or at fixing up the latter to make them better. This isn't to say that naïve subjects are bad *at reasoning*, of course. For each of our first-order information-generating systems will have been under selection pressure for speed and reliability. And moreover, some of the heuristic reasoning processes that people employ turn out to be quite remarkably successful (Gigerenzer *et al.*, 1999). Rather, it is to say that naïve subjects are bad at reasoning *about reasoning* – at identifying mistakes in reasoning, at theorizing about standards of good reasoning, and at improving their own and others' reasoning. Yet this is precisely the competence that the self-monitoring model predicts we should have.

One of the defining features of human civilizations, in fact, is that they contain socially-transmitted bodies of belief about the ways in which one *should* reason. And these bodies of belief (such as the canons of good scientific method, developed piecemeal over the last five centuries or so) have to be laboriously acquired through processes of formal education. Insofar as we have any competence in evaluating and improving reasoning, therefore, this isn't a *natural* competence, but a socially transmitted one. Hence we have no reason to think that the architecture of our cognition is as the self-monitoring model claims, or as figure 3 depicts.

7 Conclusion

I have been concerned to argue that the question of animal consciousness probably doesn't matter very much, except for those of us who are interested in the strengths and weaknesses of competing theories of consciousness. I have argued elsewhere that whether or not non-human animals are phenomenally conscious, sympathy and concern for animal suffering can still be just as appropriate (Carruthers, 1999, forthcoming). And I have argued in the present paper that

whether or not non-human animals are phenomenally conscious, the behaviors that humans and animals share can (and should) receive just the same kinds of (first-order) psychological explanation. So the question of animal consciousness matters little for psychology. As folk-psychologists, we merely need to adjust our beliefs about the properties of our mental states that carry the true explanatory burden. And scientific / comparative psychology, too, should continue to expect that many of the mental processes that are responsible for animal behaviors are also responsible for our own.

Acknowledgements

A earlier version of this paper was delivered in May 2003 to the Association for the Scientific Study of Consciousness meeting #7, held in Memphis Tennessee. I am grateful to all those at the conference with whom I discussed this material, and especially to Daniel Dennett and Robert Lurz. I am also grateful to an anonymous referee for this journal for critical comments and helpful advice.

References

- BAARS, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.
- BAARS, B. 1997. *In the Theatre of Consciousness*. Oxford: Oxford University Press.
- BARON-COHEN, S. 1995. *Mindblindness*. Cambridge, MA: MIT Press.
- BLOCK, N. 1986. Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10, 615-678.
- BLOCK, N. 1995. A confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18, 227-247.
- BYRNE, R. and WHITEN, A. (eds.) 1988. *Machiavellian Intelligence*. Oxford: Oxford University Press.
- BYRNE, R. and WHITEN, A. (eds.) 1998. *Machiavellian Intelligence II: Evaluations and extensions*. Cambridge: Cambridge University Press.
- CARRUTHERS, P. 1996. *Language, Thought and Consciousness: an essay in philosophical psychology*. Cambridge: Cambridge University Press.
- CARRUTHERS, P. 1999. Sympathy and subjectivity. *Australasian Journal of Philosophy*, 77, 465-482.

- CARRUTHERS, P. 2000. *Phenomenal Consciousness: a naturalistic theory*. Cambridge: Cambridge University Press.
- CARRUTHERS, P. 2004a. Phenomenal concepts and higher-order experiences. *Philosophy and Phenomenological Research*, 68.
- CARRUTHERS, P. 2004b. HOP over FOR, HOT theory. In R.Gennaro (ed.), *Higher Order Theories of Consciousness*, Philadelphia: John Benjamins, 115-135.
- CARRUTHERS, P. 2004c. On being simple minded. *American Philosophical Quarterly*, 41.
- CARRUTHERS, P. forthcoming. Suffering without subjectivity. *Philosophical Studies*.
- CHALMERS, D. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- COHEN, L.J. 1993. *An Essay on Belief and Acceptance*. Oxford: Oxford University Press.
- COWIE, A. and STOERIG, P. 1995. Blind-sight in monkeys. *Nature*, 373, 247-249.
- DENNETT, D. 1979. How to change your mind. In his *Brainstorms*, London: Harvester Press.
- DENNETT, D. 1991. *Consciousness Explained*. London: Allen Lane.
- DRETSKE, F. 1995. *Naturalizing the Mind*. Cambridge, MA: MIT Press.
- EVANS, J. and OVER, D. 1996. *Rationality and Reasoning*. London: Psychology Press.
- FRANKISH, K. 1998. A matter of opinion. *Philosophical Psychology*, 11, 423-442.
- FRANKISH, K. 2004. *Mind and Supermind*. Cambridge: Cambridge University Press.
- GAZZANIGA, M. 1998. *The Mind's Past*. Los Angeles: California University Press.
- GIGERENZER, G., TODD, P., and the ABC Research Group. 1999. *Simple Heuristics that Make Us Smart*. Oxford: Oxford University Press.
- JACOB, P. and JEANNEROD, M. 2003. *Ways of Seeing*. Oxford: Oxford University Press.
- LOAR, B. 1981. *Mind and Meaning*. Cambridge: Cambridge University Press.
- LURZ, R. 2002. Reducing consciousness by making it hot. *Psyche*, 8.
<http://psyche.cs.monash.edu.au/v8/>
- LYCAN, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- MARCEL, A. 1998. Blindsight and shape perception: deficit of visual consciousness or of visual function? *Brain*, 121, 1565-1588.
- MCGINN, C. 1989. *Mental Content*. Oxford: Blackwell.
- MILLIKAN, R. 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- MILLIKAN, R. 1989. Biosemantics. *Journal of Philosophy*, 86, 281-297.

- MILNER, D. and GOODALE, M. 1995. *The Visual Brain in Action*. Oxford: Oxford University Press.
- NISBETT, R. and WILSON, T. 1977: Telling more than we can know. *Psychological Review*, 84, 231-295.
- PAPINEAU, D. 1987. *Reality and Representation*. Oxford: Blackwell.
- PAPINEAU, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- PEACOCKE, C. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.
- POVINELLI, D. 2000. *Folk Physics for Apes*. Oxford: Oxford University Press.
- ROSENTHAL, D. 1986. Two concepts of consciousness. *Philosophical Studies*, 49, 329-359.
- ROSENTHAL, D. 1993. Thinking that one thinks. In Davies and Humphreys (eds.), *Consciousness*, Oxford: Blackwell.
- SHALLICE, T. 1988. *From Neuropsychology to Mental Structure*. Cambridge: Cambridge University Press.
- SMITH, J., SHIELDS, W. and WASHBURN, D. 2003. The comparative psychology of uncertainty monitoring and meta-cognition. *Behavioral and Brain Sciences*, 26.
- STOERIG, P. and COWIE, A. 1997. Blind-sight in man and monkey. *Brain*, 120, 535-559.
- TYE, M. 1995. *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
- TYE, M. 1997. The problem of simple minds. *Philosophical Studies*, 88, 289-317.
- TYE, M. 2000. *Color, Content and Consciousness*. Cambridge, MA: MIT Press.
- WEISKRANTZ, L. 1986. *Blindsight*. Oxford: Oxford University Press.
- WEISKRANTZ, L. 1997. *Consciousness Lost and Found*. Oxford: Oxford University Press.
- WILSON, T. 2002. *Strangers to Ourselves*. Cambridge, MA: Harvard University Press.

Figure 1: A mind with dual visual systems (but lacking a HOT faculty)

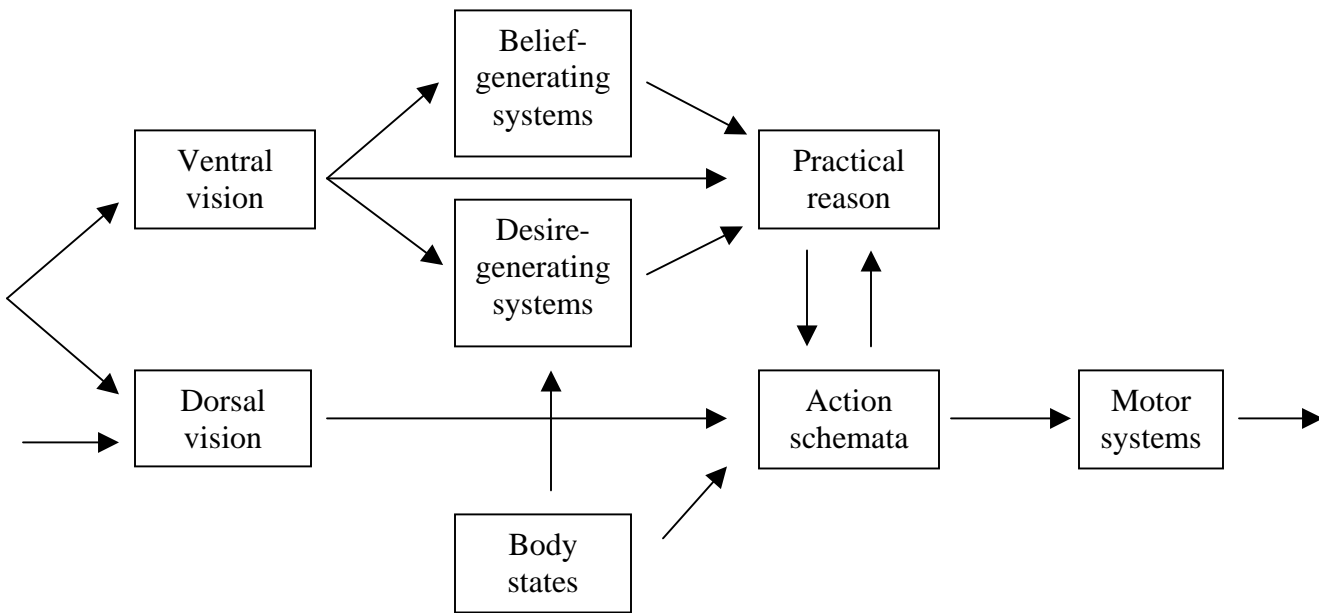


Figure 2: A mind with dual visual systems and possessing a HOT faculty

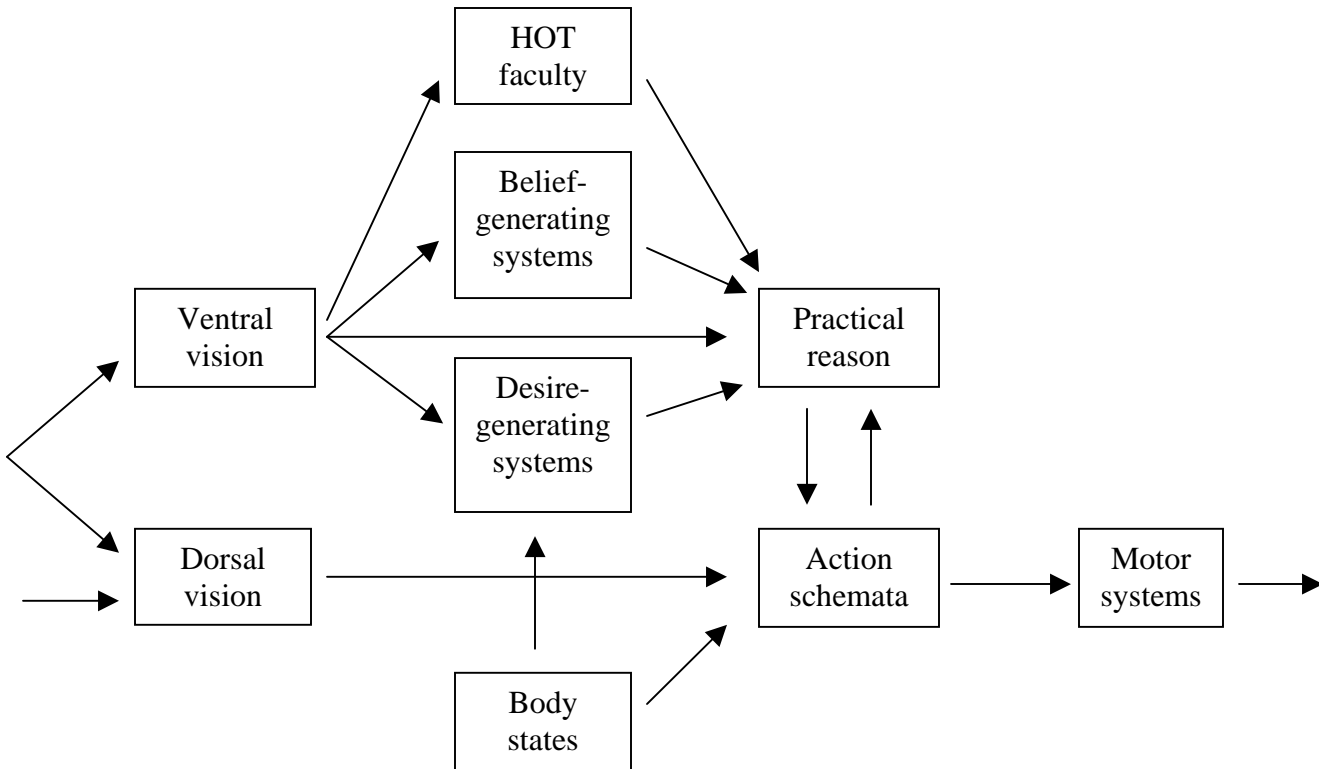


Figure 3: A mind with a self-monitoring HOT faculty

