5

How to Evaluate the Unity Thesis

Determining whether or not a creature has a unified consciousness is an empirical task that demands close and careful attention to its cognitive and behavioural profile. The structure of a creature's consciousness cannot be 'read off' from its behaviour in any direct fashion, but instead requires bridging principles that link consciousness to behavioural and cognitive capacities. Such principles will invariably be contested. The goal of this chapter is to examine some principles that might plausibly link phenomenal unity to cognitive and behavioural capacities, and in so doing develop a framework for the third-person assessment of the unity thesis that can be deployed in the following chapters. This chapter is very much a 'ground clearing exercise', and many of the ideas introduced here will be developed in following chapters.

Although it has multiple manifestations, when stripped to its core there is really only one kind of argument against the unity thesis. The argument is by counter-example, and has two 'moments': a *positive* moment and a *negative* moment. The positive moment involves establishing that the target creature has conscious states e_1 and e_2, and the negative moment involves establishing that e_1 and e_2 are not phenomenally unified. In short, the positive moment involves an argument for consciousness, whereas the negative moment involves an argument against conscious unity.

5.1 The positive moment

The positive moment faces us with the vexed problem of measures or criteria for the ascription of consciousness—the problem of other conscious minds. What would constitute adequate evidence that another creature is in conscious states of a certain kind?

The first point that must be made here is that there is no short and snappy answer to this question. The conditions under which we ought to ascribe conscious states to a creature will need to take into account the kind of creature that we are dealing with, the background state of consciousness that it would be in if it were conscious, and the kind of fine-grained conscious state in question. For example, the considerations that might be relevant to determining whether a linguistic creature is conscious are likely to differ in important ways from those that might be relevant to determining whether a non-linguistic creature is conscious. Similarly, the considerations that might be relevant to determining whether a creature enjoys (say) auditory experiences are likely to differ in important ways from those that might be relevant to determining whether it enjoys (say) affective experiences. However, in order to make the following discussion manageable I will leave these complexities largely to one side, and will assume that we can sensibly refer to measures or criteria for the ascription of conscious states in the abstract. However, readers are urged to keep this idealization firmly in mind in this (and following) chapters.

The received view within consciousness studies is that the only legitimate tool for measuring consciousness is introspective report. Frith and colleagues state that 'to discover what someone is conscious of we need them to give us some form of report about their subjective experience' (Frith et al. 1999: 107). According to Weiskrantz, 'we need an *off-line* commentary to know whether or not a behavioural capacity is accompanied by awareness' (Weiskrantz 1997: 84). Naccache tells us that 'consciousness is univocally probed in humans through the subject's reports of his or her own mental states' (Naccache 2006: 1396). Finally, Papineau claims that the 'canonical way of finding out what someone experiences is to take note of their *reports*' (Papineau 2002: 189), and he clearly has introspectively based reports in mind. Let us call this the *introspective criterion*.

There is certainly much to be said for treating introspective reports as a central guide to both the presence and absence of consciousness. Not only are they one of the most useful tools for studying consciousness, they are arguably also one of the most reliable. But there is reason to doubt whether introspection is the only—or even the most fundamental—means that we have for ascribing conscious states to creatures.

Note first that formulating a plausible version of the introspective criterion is far from straightforward. We clearly don't want to say that we are justified in ascribing a conscious state with content <p> to *S* only if *S* has *actually* produced an introspective report to the effect that she is in conscious state <p>. More plausible is the thought that we are justified in ascribing a conscious state with content <p> to *S* only if *S* has the *capacity* to produce introspective reports to the effect that she is in conscious state with content <p>. But what exactly is it to have the capacity to produce introspective reports? We don't want to demand that subjects can produce the relevant kind of introspective report right here and now, for surely we can be justified in ascribing conscious states to creatures who are unable to produce such reports due to distraction, inattention, or indeed the fact that they are introspecting some other aspect of their overall experiential state. Should we then require only that subjects have the capacity to produce introspective reports in general? That would surely be too weak, for the fact that a creature might have the capacity to produce a certain type of introspective report doesn't tell us much about its *current* states of consciousness. In short, nailing down a plausible version of the introspective criterion is not easily done (see also Schooler & Fiore 1997).

A further problem with the introspective criterion is that it does not sit easily with the thought that the introspective judgements that subjects make are often wrong. Inattentional and change blindness—to name just two of the many phenomena that might be mentioned in this connection—suggest that introspection may mislead us about what particular conscious states we are enjoying at any one time. Indeed, the debate about the existence and nature of cognitive phenomenology suggests that introspection might even mislead us about the *kinds* of phenomenal states that we possess (Bayne & Montague 2011). These claims are of course controversial, but leaving to one side the debate about these particular examples, there is good reason to think that introspection might not furnish us with quite the direct and immediate access to consciousness that it is often thought to.

A third challenge for the introspective criterion concerns its inability to accommodate our pre-theoretical assumptions concerning creatures—such as young children and non-linguistic animals—that lack introspective capacities at all. Taken in its unrestricted form, the introspective criterion entails that the question of consciousness in such creatures is a purely theoretical one—a possibility that must be forever beyond our ken. That, I think, would be a highly unpalatable view. Of course, one could restrict the scope of the introspective approach, and hold that it applies only to creatures with the capacity to produce introspective reports. But although restricting the introspective criterion in this way might accommodate our intuitions about non-linguistic creatures, the very act of restricting the scope of the account raises questions of its own. Once we have recognized that there are non-introspective means of detecting consciousness in non-linguistic creatures, why shouldn't we allow that those same means can also be applied to linguistic creatures? Introspective reports may play a core role in the ascription of consciousness—especially when it comes to creatures who possess introspective abilities—but they are *not* the sole basis on which conscious states can be legitimately ascribed.

A final criticism of the introspective criterion is that it doesn't reflect what those who study consciousness actually *do*. Although the introspective criterion represents something akin to the official self-understanding of consciousness science, most consciousness scientists actually demand only environmental reports from their subjects. Typically, subjects will be asked to report when (say) the light is red, rather than when they have an experience (as) of a red light.

Of course, one might argue that environmental reports can be substituted for introspective reports without loss. Subjects who are able to report (say) that the light in front of them is red will typically also be able to report that they had an experience (as) of a red light. In this way, perhaps, a suitably constrained formulation of the introspective criterion might be able to account for experimental practice. But although this proposal might be able to account for experiments involving self-conscious subjects who appreciate the relations between introspective and environmental reports, the science of consciousness also concerns itself with creatures who are not able to substitute introspective reports for environmental reports. Indeed, the science of consciousness draws on data from creatures whose ability to produce any kind of reports is questionable.

In an influential set of experiments designed to identify the neural correlates of visual consciousness, Logothetis and colleagues examined the neural responses of rhesus monkeys to binocular rivalry (Sheinberg & Logothetis 1997; Logothetis et al. 2003). The monkeys were first trained to press bars in response to various images—horizontal and vertical gratings, for example—and then presented with rivalrous stimuli. As expected, their responses closely modelled those of human observers to the same stimuli. The question that concerns us here is not what this research tells us about the neural correlates of visual experience, but what we should say about the monkeys' button-presses. Logothetis and colleagues describe the monkeys as *reporting* their mental states, but I would want to resist this interpretation. It seems to me that there is little reason to suppose that the monkeys were producing reports of any kind let alone introspective reports. Arguably, to report that such-and-such is the case one has to conceive of the one's behaviour as likely to bring about a particular belief in the mind of one's audience—indeed, as likely to bring this belief about in virtue of the fact that one's audience appreciates that one's behavior carries the relevant informational content—and I know of no good reason to believe that the monkeys conceived of their button-presses in these terms.

Does it follow that we have no grounds for thinking that the monkeys were experiencing binocular rivalry? Not at all; in fact, I think the monkeys' button-presses qualify as very good evidence for the claim that they had rivalrous experiences. However, their button-presses constitute evidence of consciousness not because they were reports of any kind but because they were *intentional actions*. Intentional agency, I suggest, functions as a legitimate ground for the ascription of consciousness. Indeed, it is utterly commonplace to suppose that the non-verbal behaviour of an organism can give us evidence about its experiential life. We take the lioness to be conscious of the gazelle as she tracks its movements across the savannah, we take the bloodhound to be conscious of the scent that it follows through the woods, and we take the infant to be conscious of his mother's voice as he orients towards her. Let us call the claim that intentional agency can underwrite ascriptions of consciousness the *agentive criterion* of consciousness. Note that unlike the introspective criterion, the agentive criterion does not place necessary conditions on the ascription of consciousness, but only sufficient conditions.[1]

The notion of agency is of course a very broad one, and includes both environmental and introspective reports within its ambit. Because of this, adopting the agentive criterion in no way implies that we cannot use a creature's reports to determine its states of consciousness. Indeed, there will be many situations in which what a creature 'says'—that is, reports—may be the most useful way of getting the measure of its conscious states. But the important point is that from a theoretical perspective the agentive criterion treats reports as merely one tool among many when it comes to identifying what a creature might or might not be conscious of. Reports—even introspective reports—have no privileged status, and they are most certainly *not* the 'gold standard' of consciousness research.

Although it has its advocates, the agentive criterion does not command widespread support. I suspect that this is because it is widely thought that cognitive science has shown that the bulk of cognitive and behavioural control is under the control of so-called 'zombie systems'—systems that operate on the basis of unconscious representations (Koch & Crick 2001; Koch 2004). And if consciousness has little role in guiding thought and action, then—so the thought goes—we can hardly use agency as a legitimate marker of consciousness.

How forceful is this argument? Let us start by examining the evidence for zombie systems. Much of the contemporary enthusiasm for zombie systems began with blindsight, a condition that involves damage to the primary visual cortex that produces a scotoma (Weiskrantz 1986/2009). Although patients with blindsight deny that they are conscious of stimuli that are presented within the scotoma (the 'blind field'), they can reliably discriminate them when prompted to guess, at least when those stimuli are reasonably elementary.

[1] For discussion of the ways in which agency might function as a marker of consciousness see Clark (2001, 2009); Dretske (2006); Flanagan (1992); Morsella (2005); and van Gulick (1994).

Indeed, in some cases the acuity threshold of the patient's blind field can exceed that of his or her sighted field (Weiskrantz et al. 1974; Trevethan et al. 2007).[2]

Another influential source of evidence for the existence of zombie systems comes from work on the two visual systems—the dorsal stream and the ventral stream (Goodale & Milner 2003; Milner & Goodale 2006, 2008). Much of our knowledge of the two visual systems derives from the study of D.F., a woman who at the age of 35 experienced anoxia due to carbon monoxide poisoning. The anoxia damaged D.F.'s ventral stream, leading to impairments in her ability to see the shape, orientation, and location of objects, but it left intact those dorsal stream processes that are responsible for fine-grained online motor control. In one particularly striking study, D.F. was able to 'post' a letter through a slot despite being unable to report the slot's orientation or say whether it matched the orientation of another slot. D.F., we might say, has 'blindsight for orientation'. Numerous studies have shown that normal subjects also draw on dorsal representations that lie outside of consciousness for fine-grained visuomotor control. Much of this research is summarized—somewhat misleadingly, as we shall see—by suggesting that dorsal stream processing is 'for action' whereas ventral stream processing is 'for perception'.[3]

A third field of research that has provided evidence for the existence of zombie systems involves masked stimuli.[4] Such stimuli have been shown to trigger not only simple movements requiring only one muscle group, but also more complex movements 'requiring simultaneous activity in at least two sets of muscles on opposite sides of the body' (Taylor & McClosky 1990: 445). A number of other studies have demonstrated that stimuli of which the subject reports no conscious awareness are able to activate motor programmes, and thus influence reaction times and error rates (see e.g. Klotz & Neumann 1999). As Wilhelm Wundt's student Hugo Münsterberg wrote in the late nineteenth century:

> When we apperceive [consciously perceive] the stimulus, we have usually already started responding to it; our motor apparatus does not wait for consciousness, but does restlessly its duty, and our consciousness watches it and is not entitled to order it about. (quoted in Neumann and Klotz 1994: 143)

[2] See Kolb & Braun (1995) and Lau & Passingham (2006) for evidence of blindsight-like phenomena in normal subjects.

[3] See e.g. Aglioti et al. (1995); Bridgeman et al. (1979); Bridgeman et al. (1981); Castiello et al. (1991); Fourneret & Jeannerod (1998); McIntosh et al. (2004a); Slachevsky et al. (2001); Schenk & McIntosh (2010).

[4] See also Ansorge et al. (1998); Lau & Passingham (2007); Leuthold & Kopp (1998); Miller et al. (1992); Schlaghecken & Eimer (1997); Smid et al. (1990); Taylor & McClosky (1996).

So much for zombie systems—how much pressure do they really put on the agentive criterion? Not a great deal, in my view. The study of zombie systems may give us reason to *reconfigure* our intuitive conception of the relationship between consciousness and agency (see e.g. Clark 2001), but by no means does it show that 'in order to discover what someone is conscious of we need them to give us some form of report about their subjective experience' (Frith et al. 1999: 107).

The first point to note is that even when online visuomotor activity is not grounded in visual experience it is still likely to be *correlated* with the existence of such states, at least in normal human beings. Although one's grip on a cup might be guided by unconscious dorsal stream representations, such representations are likely to be accompanied by ventral stream representations of the cup, and those representations will be conscious. And as far as the ascription of consciousness is concerned, all we need is the claim that agency is reliably correlated with the presence of consciousness. Secondly, very little behaviour is under the *exclusive* control of zombie systems. Zombie systems are not homunculi—'mini-me' who form and execute their own plans. Instead, their operations are very much under the control and guidance of the contents of perceptual and intentional consciousness. Milner and Goodale liken the dorsal stream to the robotic component of a tele-assistance system: the ventral stream selects the goal object from the visual array and the dorsal stream carries out the computations required for the assigned action (2006: 232). Left to its own devices the dorsal stream is capable of little. We can see this by considering just how impoverished D.F.'s behavioural capacities are when she is unable to draw on conscious information. She can pick up a screwdriver, but she fails to pick it up from the appropriate end for she is unable to recognize it *as* a screwdriver. She can post a 'letter' through a slot without error, but she has difficulty posting T-shaped objects (Goodale et al. 1994) or grasping X-shaped objects (Carey et al. 1996; McIntosh et al. 2004b). To contrast 'vision for action' with 'vision for (conscious) perception' is somewhat misleading, for there is very little action without conscious perception, and the vast majority of D.F.'s actions involve a great deal of conscious perception.

Close inspection of the blindsight data reveals the same story. Not only are blindsight subjects unable to discriminate complex shapes in their blind field, even those properties that they can unconsciously identify are not spontaneously integrated into their practical and theoretical reasoning. We might say that the availability of unconscious perceptual content is not 'manifest' to the patient, and as a result it is not really available to *them* as such. Blindsight subjects take themselves to be 'merely guessing', a fact that is reflected in their reluctance to incorporate blind field content into their spontaneous behaviour.

What would we say if zombie systems did start to act 'under their own steam', that is, independently of the control and guidance of conscious states? Suppose that certain blindsighters began to develop 'superblindsight' (Block 1995), and incorporated the contents of their blind field into their action plans. To the best of my knowledge, human blindsight patients do not exhibit 'superblindsight', but perhaps Humphrey's monkey Helen did (Humphrey 1972, 1974).

> Helen, several years after the removal of visual cortex, developed a virtually normal capacity for ambient spatial vision, such that she could move around under visual guidance just like any other monkey. This was certainly unprompted, and in that respect 'super' blindsight. (Humphrey 1995: 257)

Was Helen visually conscious of her environment? One might argue that she wasn't, on the grounds that her visual cortex had been removed. But one might equally argue that her spontaneous activity—the fact that she could move around under visual guidance 'just like any other monkey'—suggests that she was visually conscious. Far from undermining the agentive criterion, blindsight actually reveals the depth of intuitive support that it enjoys.

At this point a critic might allow that although intentional agency can ground the ascription of consciousness in the absence of introspective reports, introspective reports ought always to be privileged over any other form of agency. Warming to this theme, the theorist might point out that Helen showed spontaneous behavioural guidance in the absence of introspective report, and we don't know what introspective judgements she might have produced had we been able to ask her about the nature of her experiential states. Had she been able to produce introspective reports, the critic might continue, we ought surely to have trusted her reports ('I don't have any awareness in this area of my visual field') over her spontaneous behaviour.

I'm inclined to agree that introspective reports ought in general to be accorded *some* kind of privilege over other forms of behaviour as far as the ascription of consciousness is concerned. But it is, I think, a rather delicate matter about just what to say when introspective reports dissociate from non-introspective forms of behavioural control.

As an aid to reflection, let us consider some actual cases of 'agentive fragmentation'—instances in which the agent's overall behavioural profile provides conflicting cues about their state(s) of consciousness. In a series of unpublished masking experiments Cummings presented a row of five letters to subjects in rapid succession. Subjects were instructed to press one key if the letter 'J' (for example) was present in the display and to press another key if no 'J' was present.

> When urged to respond as fast as possible, even at the cost of making a good many errors, subjects now tended to respond to the occurrence of a target letter in the 'blanked' positions with a fast (and *correct*) press of the 'target present' key, and then, a moment later, to apologize for having made an error. (Allport 1988: 174–5)

On the face of things this study involves a dissociation between environmental reports and introspective reports. The subject (correctly) reports that the display contained (say) a 'J', but then (incorrectly?) reports that she did not consciously perceive a 'J'. Which of these two reports provides us with a better measure of the subject's experience: the initial action, which appears to be an environmental report, or its subsequent introspectively based retraction?

Here is one argument—a bad argument, as it turns out—for privileging introspective reports over any other form of behaviour. Distinguish two kinds of errors that might be produced by a method for detecting consciousness: false positives and false negatives. A method produces a false positive when it says of a creature who is not in conscious state K that it is in conscious state K, and it produces a false negative when it says of a creature who is in conscious state K that it is not in conscious state K. Both errors are to be avoided, but which type of error is worse? If one thought that false positives were worse than false negatives, then one might be inclined to argue that introspection ought to be privileged over other possible measures of consciousness, for only by so doing could we be sure of avoiding false positives.

Here are two reasons why the argument is a bad one. First, it is far from clear that the introspective criterion *is* guaranteed to be more conservative—that is, to generate the fewest false positives—than its rivals. Arguably, a maximally conservative approach to the ascription of consciousness would insist that we should ascribe conscious states to a creature only when each of the various possible markers of consciousness point in the same direction. Secondly, there is no reason to think that we should be looking for a (maximally) conservative marker of consciousness in the first place. The concern to avoid false positives must be balanced against the concern to avoid false negatives, and as far as I can see there is no reason to regard one of these two errors as more serious than the other.

So, what *should* we do when introspective reports dissociate from other forms of behaviour? I'm not sure that there is any simple answer to this question. Dissociations of this kind—and we will come up against many such dissociations in future chapters—will need to be evaluated on a case-by-case basis. There are few rules to guide us here, and we will need to proceed largely by instinct. That being said, let me sketch some considerations that might be of some use to us as we attempt to find a clear path through the undergrowth.

Perhaps the most important point to keep in mind is that the agentive criterion grounds the ascription of consciousness in the exercise of *personal-level* agency. This isn't to say that our evidence for the existence of consciousness in a creature is *restricted* to intentional agency, but it is to say that our evidence for ascribing consciousness to a creature will be strongest when we are most sure that we are responding to something that *the agent* has done. By contrast, our evidence for consciousness will be correspondingly weaker to the extent that we are unsure about whether we are dealing with something that the agent has done as opposed to something that some sub-personal component of the agent—some 'mechanism inside them'—has done. Consider here the control of eye-movements. Some eye-movements are controlled by low-level mechanisms that are located within the superior colliculus; others are directed by high-level goal representations (Kirchner & Thorpe 2006; de'Sperati & Baud-Bovy 2008). Arguably, the former are more properly ascribed to sub-personal mechanisms within the agent, whilst the latter are better thought of as things that the agent does—as instances of 'looking'.

In drawing this distinction between the personal agency and sub-personal motor control I am *not* suggesting that we think of 'the agent' as some kind of homunculus, directing the creature's behaviour from its director's box within the Cartesian theatre. Nor am I suggesting that we should identify personal-level agency with 'willed', 'deliberate', or 'endogeneous' agency and exclude from its domain 'stimulus-driven', 'automatic', and 'exogenous' action. That would also be a mistake. As agents we are not 'prime movers' but creatures that behave in either reflexive or reflective modes depending on the dictates of the environment. The agent isn't to be identified with the self of rational reflection or pure spontaneity, but is to be found by looking at how the organism copes with its environment. Although there is evidence that the circuits involved in stimulus-driven agency are distinct from those that subserve self-generated agency (Jahanshahi & Frith 1998; Lengfelder & Gollwitzer 2001), few of our actions are either purely self-generated or purely stimulus-driven. Instead, the vast majority of what we do involves a complex interplay between our goals and environmental affordances (Prochazka et al. 2000; Haggard 2008). (Think of a typical conversation, where what one says is guided by both the behaviour of one's interlocutor and one's own goals and intentions.) Consciousness might manifest itself most obviously when deliberative, reflective control is required, but many stimulus-driven, automatic, and exogenous actions are also guided by conscious representations, zombie systems notwithstanding. Absent-mindedly making myself a cup of coffee whilst chatting to a colleague counts as a stimulus-driven automatic action on any reasonable construal of the notion, but there is little reason to deny that I draw on my perceptual experience of the world—and, indeed, my own body—in performing such a task.

A better approach to the contrast between personal and sub-personal control, I suspect, invokes the notion of cognitive integration. What it is for an action to be assigned to the agent herself rather than one of her components is for it to be suitably integrated into her cognitive economy. Where we have behaviour that is not suitably integrated into the agent's wider mental life, there is some temptation to think that it shouldn't be assigned to the agent, at least not without reservation.

Consider in this respect the following pair of studies. In one study, three blindsight patients were required to report when they saw a light that had been flashed into their blind field (Zihl & von Cramon 1980). The patients were instructed to produce eye-blink reports on some trials, key-press reports on other trials, and verbal reports (saying 'yes') on a third set of trials. Although the subjects were able to perform well by blinking and key-pressing (after practice), their verbal responses remained at chance. In the second study cognitively unimpaired subjects were required to report the onset of a light (illuminated for 200 milliseconds) in three ways at once: by blinking, by pressing a button, and by saying 'yes' (Marcel 1993, 1994). Surprisingly, the reports that subjects gave via one report modality often failed to cohere with those they produced via other report modalities. For example, in a single trial a subject might press the button but fail to say 'yes'. Furthermore, subjects were often unaware that they had failed to produce consistent responses across the three response modalities.

Although it is certainly possible that the subjects in these studies were conscious of the stimuli that they reported, the fact that their reports were in some sense 'autonomous'—that is, could not be matched by behaviour implicating other response modalities—does entail, I think, that it is *less* plausible than it would otherwise have been. The more widely available the contents of a representation are, the more comfortable we will be in ascribing it to the agent rather than one of their components. The agent *as such* seems to recede when dealing with representations that are able to drive only a restricted range of consuming systems, and with it our evidence that we are dealing with consciousness may also recede. Arguably this will leave the borderlands between personal-level agency and sub-personal agency murky, but perhaps that is how it should be.

These comments are obviously highly abstract, and in many cases they will offer us little concrete guidance about what kinds of conscious states, if any, to ascribe to the creature in question. We will often have to muddle through as best we can, and in many cases we may simply not be in a position to say anything about the nature of the subject's consciousness with any significant degree of warrant.

5.2 The negative moment: representational disunity

I turn now from the question of how the presence of consciousness might be established to the question of how the presence of disunity within consciousness might be established. What would count as evidence that a subject has conscious states that are not phenomenally unified?

One line of argument for phenomenal disunity appeals to failures of representational integration. Here, we need to distinguish different forms of representational integration. Perhaps the most basic form of representational integration takes the form of conjunction. In Chapter 3 we examined the idea that phenomenal unity goes together with the closure of phenomenal content under co-instantiated conjunction. According to closure, if experiences with contents <p> and <q> respectively are phenomenally unified then the subject will also have an experience with the content <p&q>. Although I rejected the closure-based analysis of phenomenal unity, I did allow that phenomenal unity will typically be accompanied by the conjunction of phenomenal content, particularly when the states in question belong to the same perceptual modality.

But there are other forms of representational integration besides conjunction. Consider Nagel's gloss on the unity of consciousness:

> Roughly, we assume that a single mind has sufficiently immediate access to its conscious states so that, for elements of experience or other mental events occurring simultaneously or in close temporal proximity, the mind which is their subject can also experience the simpler relations between them if it attends to the matter . . . The experiences of a single person are thought to take place in an experientially connected domain, so that the relations among experiences can be substantially captured in experiences of those relations. (Nagel 1971: 407)

Examples of the idea that Nagel has in mind are not hard to find. An experience of two colour patches is typically accompanied by an awareness of their relative intensities; an experience of two sounds is typically accompanied by an awareness of their spatial relations; an experience of two bodily sensations is typically accompanied by an awareness of whether or not they occur in the same limb. Even when one is not actually aware of the simpler relations between the contents of unified experiences, one usually has the capacity to become aware of those relations.

What holds of the simpler relations between perceptual features also extends to the awareness of categories and gestalt relations that are mediated by e_1 and e_2. Suppose that e_1 and e_2 are representations (as) of the first and second halves of the word 'cobweb'. A subject who enjoys both e_1 and e_2 and who has the ability to recognize the word 'cobweb' as such will typically enjoy an experience (as) of the word 'cobweb'—at least, if e_1 and e_2 are phenomenally unified. If these experiences are not phenomenally unified then the subject will be restricted to experiences of the words 'cob' and 'web' and will not experience the perceived word as 'cobweb'.

With these thoughts in mind we can now see how we might argue for phenomenal disunity by appealing to failures of representational integration. Such an argument will require a principle connecting phenomenal unity with representational unity, along the lines of the following:

> *Representational Integration Principle* (RIP): For any pair of simultaneous experiences e_1 and e_2, if e_1 and e_2 are phenomenally unified then, *ceteris paribus*, their contents will be available for representational integration.

Developing a plausible argument from representational disunity to phenomenal disunity requires three things. Firstly it requires showing that the subject does indeed have experiences e_1 and e_2. Secondly it requires showing that these experiences are not representationally integrated. And, thirdly, it requires showing that the various *ceteris paribus* clauses cannot be activated. The first two tasks raise issues that have already been dealt with in §5.1, so I will focus here on the third: when might failures of representational unity fail to provide evidence of phenomenal disunity?

Begin by noting that some creatures simply won't have the cognitive machinery required to integrate the contents of their mental states in the appropriate manner. The degree to which a creature's states can be integrated will be a function of its integrative abilities—on what categories and concepts it has. Someone who cannot recognize elephants will not be able to synthesize experiences of the front and back ends of an elephant so as to recognize the presented animal as an elephant. Similarly, someone who suffers from prosopagnosia will be unable to synthesize the perceptual features of a face into a face percept. Of course, if certain forms of feature binding are necessary for perceptual consciousness—as has sometimes been claimed—then anyone with any perceptual experience at all must possess certain integrative capacities. But even if some forms of feature binding are essential to perceptual consciousness—it is clear that there are many others forms of binding that are not required for the possession of consciousness as such.

Secondly, even when the creature in question does have the machinery to integrate the contents of e_1 and e_2, it might not have the capacity to deploy that machinery on the occasion in question. Certain kinds of integration might require that the experiences in question have a certain temporal duration (which they might not have), or that the creature can attend to the relevant objects and properties (which they might not be able to). We also need to be particularly mindful of the role played by background states (or 'levels') of consciousness. Although we have some basic familiarity with the kinds of capacities for representational integration that creatures possess in the context of normal wakefulness, this is only one of many background states of consciousness and it would be naïve to expect that the kinds of representational integration and coherence that we find in ordinary waking consciousness also characterize all other background states of consciousness. In light of these considerations we would do well to exercise especial caution when arguing from representational disunity to phenomenal disunity outside of normal attentive wakefulness.

Let us take stock. The lack of representational integration—and closure, in particular—will often provide us with a reason to think that the states in question are not phenomenally unified, but just how strong that reason is will depend on the details of the case, and we certainly shouldn't expect that representational integration will follow from phenomenal unity with strict necessity.

5.3 The negative moment: access disunity

A second—and perhaps more potent—line of argument for phenomenal disunity concerns the *uses* to which a subject's conscious states can be put. In general, the contents of each of a creature's conscious states are available to the same range of cognitive and behavioural systems. But suppose that we came across a subject who appeared to be in two conscious states at the same time (e_1 and e_2), where the contents of these two states were not available to the same systems of cognitive and behavioural consumption. For example, the contents of e_1 might be available for verbal report but not memory consolidation, whereas the contents of e_2 might be available for memory consolidation but not verbal report. It would be tempting to take this selective accessibility—this breakdown in 'access unity'—as evidence that e_1 and e_2 were not phenomenally unified.

The key component in any argument from access disunity will be a principle that links phenomenal unity to access unity, such as the following:

> *Conjoint Accessibility Principle* (CAP): For any pair of simultaneous experiences e_1 and e_2, if e_1 and e_2 are phenomenally unified then, *ceteris paribus*, their contents will be available to the same consuming systems.

Assuming CAP, we can use the fact that the contents of e_1 and e_2 are not available to the same consuming systems as evidence that they are not phenomenally unified. As with arguments from representational disunity, arguments from access disunity will rarely—if ever—be demonstrative.

Developing a plausible argument from access disunity to phenomenal disunity requires three things. First, it requires showing that the subject does indeed have experiences e_1 and e_2. Secondly, it requires showing that these experiences are 'access disunified'—that their contents are not available to the same consuming systems. And, thirdly, it requires showing that the various *ceteris paribus* clauses cannot be activated. Given that the first task raises issues that were dealt with in §5.1, we need now to examine the second and third tasks.

Let us start with the question of what it is for states to be 'access disunified'. Take a creature with two experiential states (e_1 and e_2) and five consuming systems ($CS_1 \ldots CS_5$). States e_1 and e_2 will be fully access unified if their contents are available to all and only the same consuming systems, and they will be fully access *dis*unified if their contents are not available to any of the same consuming systems. But suppose that the contents of e_1 and e_2 are available to *some* of the same consuming systems but not others. For example, e_1 might be available to CS_1–CS_4, and e_2 might be available to $CS_2 \ldots CS_5$. We might think of such states as *partially access unified*. What should we make of such states? We could take the fact that the contents of both e_1 and e_2 are accessible to CS_2, CS_3, and CS_4 as evidence that they *are* phenomenally unified, but we could equally invoke the fact that some consuming systems have access to only one of these two experiences as evidence that they are not phenomenally unified. Neither response seems entirely appropriate.

One might attempt to defuse the force of this worry by arguing that partial co-accessibility is unlikely to occur, and hence that this issue can be set to one side notwithstanding its theoretical interest. Although tempting, I suspect that this kind of optimism is misplaced; in fact, given the complex nature of the architecture underlying cognitive and behavioural control there is reason to expect that breakdowns of access unity will usually be partial rather than complete. If one looks hard enough, one is likely to find some kind of cognitive task on which otherwise disunified states can be jointly brought to bear. Rather than being a theoretical problem that can be safely brushed under the carpet, partial co-accessibility needs to be confronted (as we will see).

A second response to the challenge of partial unity would be to suggest that degrees of access unity can be equated with degrees of phenomenal unity: the more access unified two states are, the more phenomenally unified they are. But this response can also be set to one side. For one thing, it is highly doubtful whether phenomenal unity can come in degrees. Two conscious states are either phenomenally unified with each other or they are not, and it is not possible for one pair of experiences to be more phenomenally unified with each other than another. And even if phenomenal unity can come in degrees, there is no reason to assume that degrees of phenomenal unity will be correlated with degrees of access unity.

In my view, the most reasonable response to partial unity is epistemic: the strength of any argument from access disunity will be a function of the degree to which the contents of the relevant states are co-accessible: the less co-accessible the contents of the states, the stronger our evidence for thinking that they are not phenomenally unified.[5] We should allow that phenomenal unity is *compatible* with some degree of access disunity—indeed, in some cases it might even be reasonable to suppose that phenomenal unity coexists with a high degree of access disunity—but the more radical the access disunity the better our evidence for phenomenal disunity. In some situations it might be rational to think that phenomenal unity coexists with a high degree of access disunity (see further §5.5), but in such cases one would be under an obligation to explain why phenomenally unified states are not also access unified.

Let us turn now to the third issue raised by the argument from access disunity: the question of *ceteris paribus* clauses. Under what conditions might things not be equal?

First, any argument from access disunity needs to factor in the possibility of processing bottlenecks. Consider a complex phenomenal state *(*e_3*)* and a simpler phenomenal state *(*e_1*)* that is subsumed by it. Although e_3 and e_1 will be phenomenally unified, they may not be access unified due to the fact that e_3 is more complex than e_1. (Given the disparity in the 'size' of their contents, certain consuming systems might be able to access the contents of e_1 but not those of e_3.) Of course, the same point holds with respect to experiences that are not part-whole related: two experiences might not be co-accessible due simply to the fact that one of them is more complex than the other. In light of this, arguments from access disunity will be most secure when dealing with

[5] Thanks to Ian Phillips for pushing me to say more here.

conscious states that are roughly of the same 'size'. The greater their size—especially their combined size—the more reason we will have for supposing that failures of access unity might be due to processing bottlenecks rather than failures of phenomenal unity.

A particularly important processing bottleneck is implicated in the psychological refractory period (Pashler 1992, 1998; Spence 2008). When two stimuli that are presented in close temporal succession require different responses, the response to the first stimulus almost always retards the response to the second stimulus, suggesting that there is a central bottleneck which prevents people from 'doing two things at once'. However, there are contexts in which the psychological refractory period can be minimized and perhaps even eliminated altogether. For example, the refractory period is notably reduced when the stimuli can be grouped together as aspects of a single object (here, subjects appear to be able to 'compile' the two responses into a single response), and when the responses involve high degrees of 'stimulus-response' compatibility, for example moving one's eyes in the direction of a visual stimulus.[6]

A second respect in which 'things might not be equal' involves differences in the representational format of the states in question. The means by which a state can influence thought and behaviour depends in no small part on the kind of state that it is. Suppose that we have two states, e_1 and e_2, where e_1's content is conceptual and e_2's content is non-conceptual. Given the link between conceptual content and reasoning, it would be no surprise to discover that e_1's content was able to drive (say) the mechanisms of belief-revision in ways that e_2's content was not. Conversely, assuming a link between non-conceptual content and action, it would not be surprising to discover that e_2's content could drive online behavioural control in ways that e_1's content could not. The lesson to be learnt from this is that failures of access unity will provide better evidence of failures in phenomenal unity when the states in question share a common representational format. Discovering that e_1 and e_2 couldn't (say) be jointly reported would give us better reason to think that they are not phenomenally unified if they were both conceptual states as opposed to one of them being conceptual and the other non-conceptual.

A final point that demands attention concerns the individuation of consuming systems. The problems here are not just the practical ones of knowing the neuropsychological means by which a particular behavioural response was produced, but the theoretical challenge of knowing what to do with

[6] For studies that establish the former point see Fagot and Pashler (1992) and Schumacher et al. (2001); for studies that establish the latter point see Kornblum et al. (1990), Greenwald & Shulman (1973), and Greenwald (2003).

such information. What does it take for actions to count as manifestations of a single consuming system as opposed to manifestations of distinct consuming systems?

Differences between motor systems might provide us with a rough guide to the individuation of consuming systems, but it would be most unwise to assume that consuming systems bear a one-to-one relation to motor systems. Most obviously, a single consuming system can be implicated in multiple behavioural responses. For example, introspective reports can be expressed by what one says, which buttons one presses, or where one looks. Less obviously, a single type of motor response can involve the activity of quite distinct systems. In some contexts a hand movement might realize a grasping action, in other contexts it might realize an attempt to communicate. The gross behavioural profiles of these two actions might be identical but the consuming systems implicated in them will not be. Indeed, grasping and communicating involve mechanisms that can be differentially disabled: some patients with somatosensory processing deficits are able to grasp the affected limb but unable to point to it when asked to indicate the site of damage (de Langavant et al. 2009), whilst other patients show the converse dissociation (Anema et al. 2009). The individuation of consuming systems is ultimately an empirical matter. Although pre-theoretical intuitions might be of some help in delimiting the rough borders between them, the fine-grained distinctions between consuming systems will be revealed only as the details of our cognitive architecture are filled in. An implication of this is that arguments from access disunity will be hostage to fortune, for our current assumptions about the borders between consuming systems might be quite wide of the mark.

Let us take stock once more. Access disunity provides us with an argument for phenomenal disunity, for there is reason to think that the contents of phenomenally unified states will generally be co-accessible to the subject's various consuming systems. The strength of such arguments, however, may often be difficult to determine, for not only is it plausible to suppose that phenomenal unity can coexist with some degree of access disunity, it might also be difficult to determine the degree to which the states in question are access disunified in the first place.

5.4 Probe-dependence

In the previous sections I examined two ways in which one might make a case for phenomenal disunity: by appealing to representational disunity or by appealing to access disunity. I also noted that neither type of argument is straightforward, for in each case the crucial principles involved—namely RIP and CAP—involve 'other things being equal' clauses, and in each case there is more than one way in which things may not be equal. I turn now to a further matter that complicates the evaluation of potential counter-examples to the unity thesis.

Most objects of study are independent of our attempts to study them. The number of bricks in a wall, the number of passengers on a train, and the number of monkeys up a tree are not typically dependent on the tools that one uses to measure them. Although it is natural to assume that consciousness is likewise independent of our attempts to detect it, there is reason to think that this assumption is false, and that the very contents of a subject's conscious states may be modulated by the tools that one uses to identify them. In other words, consciousness exhibits what we might call 'probe-dependence'.[7]

Probe-dependence can be illustrated by the phenomenon of extinction, a mild form of perceptual neglect that usually affects the left side of the patient's visual field (Vuilleumier & Rafal 2004; Mattingly et al. 1997). Patients with 'extinction' may be aware of stimuli that occur in their left hemi-field when such stimuli are presented singularly, but they will cease to be aware of them when they are presented concurrently with right hemi-field stimuli. However, the degree to which a patient manifests extinction can depend on just how they are tested. In one particularly striking case, Halligan and Marshall (1989) examined the ability of a patient to bisect a series of lines at the mid-point. When the patient was asked to bisect the lines using his *right* hand he veered into the right side of his visual field, suggesting that he was not aware of the left half of the lines. However, this extinction was no longer apparent when he was required to bisect identical lines using his *left* hand. In another study, five patients with neglect for the left side of visual space were asked to give same/different responses to pairs of line-drawn pictures of animals (Vallar et al. 1994). The two pictures were shown one above the other. One of the two pictures in each pair represented an intact animal, while the other represented a chimera in which the left half of one animal had been replaced by that of another animal. All patients misjudged the two members of the pair as identical to each other, but three of the five patients noticed the incongruous drawing when asked which of the two drawings more properly corresponded to the name of the animal depicted in the non-chimeric picture. A number other studies of extinction and neglect have found that the severity of a patient's symptoms are a function of how they are 'probed'.[8]

[7] The following is indebted to Dennett (1991), Hurley (1998) and especially Bisiach (1997).

[8] See Bisiach et al. (1989); Bottini et al. (1992); Duhamel & Brouchon (1990); Halligan et al. (1991); Joanette et al. (1986); Ricci & Chatterjee (2004); Smania et al. (1996); Tegnér & Levander (1991).

What accounts for such findings? One possibility is that although patients are always aware of the so-called 'extinguished' or 'neglected' stimulus, this awareness is difficult to tap. It is there, but is revealed only when patients are probed in particular ways. Another—and to my mind rather more attractive—possibility is that the very fact of asking the patient to respond to the stimulus in one way rather than another influences whether or not they are conscious of it. Asking the patient to (say) use their left hand rather than their right hand to bisect a presented line may activate the patient's right hemisphere and thus facilitate the awareness of objects in the patient's left visual field. Similarly, requiring patients to name an object might trigger right hemisphere activity, leading to the temporary amelioration of left visual field neglect.

The probe-dependence of consciousness is not restricted to brain-damaged patients but is also to be found in normal subjects of experience. Consider, for example, the Colavita effect (Colavita 1974; Spence et al. forthcoming). In the basic Colavita paradigm subjects are presented with a random series of visual, auditory, and audio-visual stimuli, and are required to make one response to the audio stimuli and another to the visual stimuli. Surprisingly, visual stimuli often 'extinguish' auditory stimuli on bimodal trials, despite the fact that subjects have no difficulty in detecting these auditory stimuli on unimodal trials. An account of the Colavita effect needs to explain why only one stimulus is reported on certain bimodal trials, and why it is the visual stimulus that invariably 'extinguishes' the auditory stimulus (rather than vice versa).

Spence (2009) provides a 'probe-dependence' account of the Colavita effect that contains plausible answers to both questions. His account draws on the finding that subjects who are instructed to respond to visual targets will be aided by the presence of an accessory sound (that is, subjects are quicker to respond to visual stimuli on bimodal trials than on unimodal trials), but the presence of an accessory visual stimulus *retards* responses (Sinnett et al. 2008). This suggests that on bimodal trials subjects initiate their response to the visual stimulus before they are in a position to respond to the auditory stimulus. So, under time-pressure to respond, vision-only responses will be expected to occur more often than audition-only responses. Moreover, Spence suggests, the representation of the auditory event will fail to generate an auditory experience on certain bimodal trials precisely because it is not responded to—or rather, because the subject has already responded to the visual stimulus. By 'responding' to the visual stimulus the subject's perceptual system has in effect 'decided' that there was only the one stimulus, and allows the representation of the auditory stimulus to fade into neuronal obscurity.

Probe-dependence might also account for some of the examples of agentive fragmentation discussed earlier in this chapter, although the devil will be in the details. It is possible that the dissociation between response modalities observed in Zihl and von Cramon's (1980) blindsight study (see p. 104) might have been a function of the pre-motor programming of reports. Perhaps the motor preparation required for verbal reports had the effect of reducing the sensitivity of the patients to stimuli in their blind field, whereas the motor preparation required for eye-blink or key-press reports did not. A similar kind of story might account for the dissociation between report modalities found by Marcel (1993). Perhaps the information concerning the stimulus began to activate the subjects' eye-blink responses prior to activating their verbal or button-press responses, with the result that subjects not only failed to produce consistent reports across these three modalities, but they also failed to monitor the fact that their responses had not been consistent. I should add that these comments are primarily intended to serve as examples of how probe-dependence might bear on agentive fragmentation, rather than as serious proposals about how to account for these findings.

Some commentators associate the notion of probe-dependence with an anti-realist conception of consciousness. Having suggested that 'probing the stream of consciousness at different times and places produces different effects', Dennett goes on to say that it is a mistake to suppose 'that there must be a single narrative (the "final" or "published" draft, you might say) that is canonical—that is the actual stream of consciousness of the subject, whether or not the experimenter (or even the subject) can gain access to it' (1991: 113). Whether or not this assumption is mistaken—and I would argue that it is not—the probe-dependence of consciousness certainly doesn't give us any reason to reject it. As far as probe-dependence is concerned, facts about consciousness could be as robust and determinate as you like.

So much for probe-dependence—what bearing does any of this have on the evaluation of the unity thesis? Probe-dependence requires that we must exercise caution in using any one response as our measure of consciousness, for the presence of the very thing being measured might be modulated by the nature of the probe(s) that we have employed. Requiring a subject to attend to a certain region of perceptual space or to prepare to initiate a particular motor response may affect the balance of activation between the two hemispheres, thus modulating how the patient's awareness is (say) distributed across their visual field. The experiences e_1 and e_2 might *appear* to be available to different consuming systems, but that appearance might be an illusion generated by the probe-dependence of consciousness. Suppose that we show a subject a light at the same time as we play a sound to him. We want to know whether he experienced both the sound and the light, and—if so—whether or not these two experiences were phenomenally unified. How are we to test him? Requiring him to produce a verbal report might bias him to report the light (and perhaps also extinguish the sound), but requiring him to produce a button-press report might bias him to report the sound (and perhaps also extinguish the light). We might be tempted to think that these two experiences were simultaneous but not phenomenally unified with each other, but in fact there may have been no single trial on which the subject was simultaneously aware of both the light and the sound. So, in evaluating any argument from access disunity we need to keep one steely eye on just how consciousness is being probed.

5.5 Imperial versus federal models

A final issue that complicates any evaluation of the unity thesis has hovered in the background of the entire chapter, but must now be brought into the clear light of day. The issue in question concerns the cognitive architecture of consciousness.

Painting with a broad brush, there are two ways in which one can think of the structure of consciousness: in 'imperial' terms or in 'federal' terms. The imperial approach conceives of consciousness in terms of a centre into which all processing flows and from which all control emanates. The contents of consciousness—and perhaps only the contents of consciousness—are located in a domain-general 'reservoir of experience'. This reservoir need not have a particular anatomical address—it might be spread across various neural regions—but it is a functional unit of sorts. Any content that it contains will be available to the same range of systems of cognitive and behavioural consumption. Certain forms of control might be decentralized, but on the imperial conception of things any such decentralized control must be located firmly outside of consciousness. Conscious control, according to this view, resides only within an imperial centre.

Advocates of a federal conception of consciousness operate with a very different view of things, according to which conscious states exert influence in the form of multiple 'domain-specific' circuits. Instead of being routed through a single domain-general workspace or central executive, the federalist thinks of conscious control on the model of a political system in which order is maintained by 'loose federations of centers of control and integration', to use Block's (1997: 162) evocative phrase. According to this approach, the ability of a state to drive various cognitive and behavioural programmes may be a function of the kind of state it is, with certain types of conscious states more readily available for some forms of cognitive and behavioural control than others. Of course, content-specific control might be hidden from sight in everyday contexts due to various mechanisms of integration, and it may take brain damage or the constraints of laboratory-induced cognitive load to reveal it.

Imperial models of one form or another are exceedingly common, with the 'centre' being variously identified with working memory, the global workspace, or the supervisory attentional system. But despite its widespread endorsement, the evidence in support of the approach is less than decisive. Certain versions of analytic functionalism notwithstanding, there is no *a priori* reason to assume that the contents of each of a subject's conscious states *must* be available for the same forms of cognitive and behavioural control. If the imperial approach to consciousness is not a conceptual requirement, then the case for it must be made on empirical grounds. Has that case been made? No doubt opinions will differ on this question, but my own view is that it has not. The studies that are cited in its support invariably operationalize consciousness in terms of global availability, which is to *assume* the truth of the imperial model rather than to provide evidence for it.

To cast doubt on global availability conceptions of consciousness is not to deny that there is an intimate connection between consciousness and global availability when it comes to neurologically intact adult humans in the normal waking state. The question, however, is how tight that connection remains once we step outside that rather restricted domain. For all we know, the correlation between consciousness and global availability breaks down when dealing with non-human animals, neonates, or adult members of our own species who have suffered trauma of some kind or another. We should remain alive to the possibility that the correlation between consciousness and global availability seen in the context of the normal waking state is generated by 'imperial' mechanisms of integration that are *superimposed* on top of the relatively more 'federal' mechanisms responsible for the construction of consciousness itself. The appearance of a 'centre' may be generated by the interaction of a number of modular-ish channels that are 'patched together' by working memory, inner speech, and other domain-general mechanisms of integration. When all goes well, the contents of these channels are widely—indeed, perhaps even *globally*—available for the control of thought and action, but control may 'go local' when these patches are put under pressure. The imperial conception of consciousness might capture an *idealized* conception of the relationship between consciousness and control, but we mustn't lose sight of the fact that we are often far from ideal subjects.

It seems to me that certain very general considerations provide reason to think that consciousness has at least a partly federal structure. For one thing, the selectional pressures driving the evolution of consciousness are likely to have been content-specific. This is perhaps most plausible with respect to bodily sensations. It would not be unreasonable to suppose that pain experiences are in the business of driving damage-avoidance behaviour, hunger experiences are in the business of driving grazing behaviour, and full bladder experiences are in the business of driving bladder-emptying behaviour. Similar kinds of functional specialization may also have shaped the architecture of perceptual experience. The various perceptual modalities are not merely specialized in the detection of particular kinds of information, they are also (and relatedly) specialized in the production of certain kinds of behaviours. For example, olfaction appears to be prepotent for memory whereas audition is prepotent for the perception of danger.

In fact, we might already have seen manifestations of our federal architecture earlier in this chapter. In discussing the psychological refractory period, I noted that it is generally impossible to produce two responses at once (§3.3). However, exceptions to this rule can be found in contexts of high 'stimulus-response' compatibility, as when subjects are required to move their eyes in the direction of a visual stimulus. Perhaps high SR-compatible actions can be executed without mutual interference because they involve distinct circuits of consciousness.

Even apparently imperial treatments of consciousness might turn out to have something of a federal structure on closer examination. Consider, for example, accounts that tie consciousness to working memory. Such accounts have an imperial sound to them, for working memory is often thought of as a single faculty. However, many models posit some kind of domain-specific fractionation within working memory. For example, Baddeley's influential account takes working memory to have four components: a central executive and three temporary storage systems (a visuospatial sketchpad, a phonological buffer, and an 'episodic buffer') (Baddeley 2003, 2007). Even the so-called central executive may turn out to have domain-specific structure. Although some theorists take executive control to be structured along *functional* lines, with different subsystems responsible for (say) manipulating, storing, and then distributing representations, others argue that it is structured along *domain-specific* lines, with subsystems of executive control being individually responsible for the manipulation, storage, and distribution of particular kinds of representations.[9] There is a similar debate about the structure of response-selection. Whereas some accounts treat response-selection as a unitary mechanism that schedules responses to various stimuli irrespective of their content, others hold that response–selection can be broken down along modality-specific lines, with different components of the cognitive architecture involved in scheduling

[9] For examples of the former approach see Owen et al. (1998); Owen et al. (1999); and Petrides (1995); for examples of the latter approach see Levy & Goldman-Rakic (2000); Adcock et al. (2000), and Gilbert et al. (2006).

responses to some kinds of stimuli but not others.[10] My concern here is not to weigh into these debates, but to point out that some of the mechanisms implicated in apparently imperial accounts of the architecture of consciousness might turn out to have an underlying federal structure.

So much for the contrast between imperial and federal conceptions of consciousness—how might this debate bear on the evaluation of putative counter-examples to the unity thesis? It is clear that representational and access disunity arguments against the unity thesis will be most straightforward on the assumption that one is dealing with a subject whose consciousness has an imperial structure. Should that assumption be challenged, then the inference from representational or access disunity to phenomenal disunity will be that much more problematic. If conscious states are grounded in content-specific circuits of control and integration, then we shouldn't expect conscious states of different kinds to be generally available for representational integration or joint behavioural and cognitive control. To put the point slightly differently, when it comes to federal subjects we will be justified in inferring a break-down in phenomenal unity from a break-down in representational unity or access unity only if we are also justified in assuming that the states in question are 'circuit-mates'. This is an important point, for establishing that the states in question are circuit-mates might be far from easy.

At this point a critic might be tempted to suggest that to endorse a federal conception of consciousness is not to complicate the evaluation of the unity thesis but to give up on it. After all—she might say—how could conscious states be unified unless consciousness has an imperial structure—unless there is a single workspace in which it 'all comes together'?

The question is a fair one, for there clearly are functional architectures that would prevent consciousness from being phenomenally unified. Nonetheless, the suggestion that the unity thesis *requires* an imperial conception of consciousness should be resisted. The unity thesis does indeed demand that any conscious subject have a single total state of consciousness, but this state need not be the product of a single workspace—even a virtual workspace. The unity of consciousness does not require that all conscious content flow towards a single location in functional space, nor does it require that the contents of each of a subject's experiences will be available to the same range of consuming systems. For all we know, the unity of consciousness may involve the integration of states that are grounded in the activity of multiple, domain-specific, circuits. To assume that phenomenal unity requires an 'imperial' conception of conscious

[10] For examples of the former approach see Pashler (1994) and Jiang & Kanwisher (2003); for examples of the latter approach see Meyer & Kieras (1997) and Schumacher et al. (2003).

control is to assume that the personal-level structure of consciousness must be preserved or mirrored at a sub-personal level of analysis. That assumption could turn out to be correct, but it has little *a priori* warrant. Indeed, to suppose that the unity of consciousness demands a centralized workspace is to make the same mistake that Descartes made when he located the seat of the soul in the pineal gland on the grounds that it is the only undivided organ in the brain.

5.6 Conclusion

Let us attempt to pull together some of the various themes developed in this chapter. Our focus has been on the question of how one might go about showing that the unity thesis is false. I began with the 'positive' moment—the task of showing that a creature enjoys certain conscious states. According to received wisdom, the only legitimate evidence of consciousness takes the form of (the capacity for) introspective report. I rejected this introspective criterion as being implausibly restrictive, and in its place I defended an agentive approach to the ascription of consciousness, according to which attributions of consciousness can be underwritten by intentional agency. Introspective report and intentional agency will typically point in the same direction as far as consciousness is concerned, but there may be occasions in which they point in different directions. I suggested that there are few rules about what to do in such contexts: in some cases it might be appropriate to invest one's faith in introspective report, in other cases it might be more appropriate to trust the 'testimony' of the subject's non-verbal behaviour.

In §5.2 I turned to the 'negative' moment—the task of establishing that some of a creature's conscious states are not unified. We saw that there are two ways in which the negative moment can be developed: the first involves an inference from representational disunity and appeals to RIP, the second involves an inference from access disunity and appeals to CAP. Neither route is unproblematic, for both RIP and CAP are hedged with *ceteris paribus* clauses, and it will often be unclear whether *cetera* are *paria*. Some failures of representational integration might be best explained by supposing that the subject lacks the capacity to integrate the contents of the states in question; other such failures might be best explained by supposing that although the subject has such capacities she is unable to exercise them at the time in question. Some failures of co-accessibility might be accounted for by invoking processing bottlenecks of various kinds; other such failures might be accounted for by differences in the representational format of the states in question. The presence of these 'get-out' clauses complicates the evaluation of putative counter-examples to the unity thesis, for it is difficult to distinguish legitimate appeals to such clauses from hollow exercises in 'ad hocery'.

§5.4 and §5.5 considered two more issues that further 'problematize' the evaluation of the unity thesis. Firstly, the very manner in which one probes a subject can have an impact on what that subject is conscious of. As we saw, the probe-dependence of consciousness cannot be ignored when asking whether a subject has certain experiences, and—if so—whether they are phenomenally unified. The second issue concerned the architecture of consciousness. Painting with a very broad brush, I distinguished two models of the architecture of conscious control: centralized (or 'imperial') models and decentralized (or 'federal') models. Deploying the arguments from representational and access disunity will be most straightforward on the assumption that consciousness has an imperial structure. By contrast, if consciousness has a federal structure—and we saw that there is some reason to suspect that it might—then the link between phenomenal unity on the one hand and representational integration and co-accessibility on the other will be that much more tenuous. Phenomenal unity might go hand-in-glove with representational integration and co-accessibility for 'circuit-mates'—that is, for conscious states that are located within the same circuits of consciousness—but there is little reason to expect this relation to hold for states that are nestled within distinct circuits.

Let me conclude by noting one final complication. In this chapter I have treated the negative and positive moments as though they are independent of each other. Although heuristically useful, this assumption is built on a fiction, for the question of whether a creature has experiences e_1 and e_2 cannot be disentangled from the question of whether those experiences—if indeed they occur—are unified with each other. Here's why. Suppose that we know that the creature is in mental states m_1 and m_2, but we want to know whether these two mental states are also conscious states. Suppose, further, that m_1 and m_2 are not co-accessible: the contents of m_1 are available to CS_1 but not CS_2, whereas the contents of m_2 are available to CS_2 but not CS_1. The fact that m_1 and m_2 are not co-accessible gives us some reason to think that *if* they are conscious then they are not co-conscious, but it might also give us some reason to think that they are not conscious in the first place. Perhaps m_1 and m_2 are examples of unconscious states that are able to drive (high-level) consuming systems. The upshot of this is that in order to produce a convincing counter-example to the unity thesis one must walk the following tightrope. On the one hand, one must show that there is sufficient disunity within a subject's cognitive economy to justify the thought that it contains states that are not phenomenally unified with each other. On the other hand, one must also establish that those states are conscious in the first place, and arguably that requires showing that their contents are at least widely available for the control of thought and behaviour. This does *not* mean that the unity thesis is 'methodologically secure'—a thesis whose falsehood could never be demonstrated—but it does mean that securing the evidence required to show that it is false is a far from trivial undertaking.

This concludes my discussion of how to evaluate the unity. I have presented two lines of argument that might be deployed against the unity thesis, and have identified some of the challenges we confront in attempting to determine the force of those arguments. In the following four chapters I put these ideas to work by examining the case for disunity, both within the normal human subject and in the context of various disorders of consciousness.