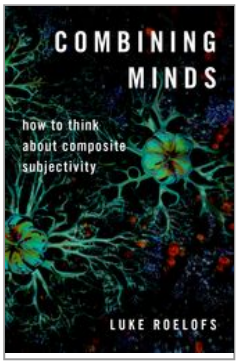

Combining Minds: How to Think about Composite Subjectivity
Luke Roelofs

Print publication date: 2019
Print ISBN-13: 9780190859053
Published to Oxford Scholarship Online: February 2019
DOI: 10.1093/oso/9780190859053.001.0001

Composite Subjectivity in Organisms, Organs, and Organizations

Luke Roelofs

DOI:10.1093/oso/9780190859053.003.0006

Abstract and Keywords

This chapter looks at four potential cases of mental combination, to examine what the theory sketched in the previous chapter might say about them. It starts with the “Nation-Brain” thought experiment, originally offered as a *reductio ad absurdum* of functionalism, where a few billion people agree to collectively simulate a single human mind. It then considers actual human social groups, the ways that they differ from this thought experiment, and the significance of these differences for questions of collective mentality. It next considers the split-brain phenomenon, where patients with a severed corpus callosum seem at times to exhibit two distinct consciousnesses in one head, and then finally comes back to the ordinary human brain, where two cerebral hemispheres, each capable of supporting consciousness without the other, are able to establish richly unified consciousness through their intact corpus callosum.

Keywords: functionalism, collective mentality, Nation-Brain, split brain, human brain, unity of consciousness

CHAPTER 5 OUTLINED functionalist combinationism, but fully fleshing it out is well beyond the scope of this book. The functional and informational relations that play central roles in the theory are complex and multifarious and must be simultaneously viewed under three aspects: as objective physical structures, as structures experienced by a single composite subject, and as external relations between subjects. I do not claim to have rigorously defined or exhaustively analyzed these structures under any of their three aspects. I have simply tried to connect together the rough understanding we already have of those aspects—of

information integration from observing physical systems, of conscious structure from our own consciousness, and of mutual co-presentation from our interactions with other people.

The best way to enrich the theory is to apply it to some cases studies, in the hope of showing how functionalist combinationism can illuminate some of the most perplexing debates in philosophy of mind. For this chapter I have chosen four case studies:

(p.190)

1. Ned Block's Nation-Brain thought experiment.
2. Actual human nations considered as candidates for consciousness.
3. The split-brain phenomenon.
4. The ordinary human brain understood as comprising multiple parallel processes determining conscious content partly independently of each other.

The first two involve a set of discrete parts, each uncontroversially an intelligent subject, together forming a system that shows some indications of subjecthood. The second two involve a system which is uncontroversially, or at least plausibly, an intelligent subject, composed of (overlapping) parts which individually show some indications of subjecthood.

6.1. Composite Subjectivity in the Nation-Brain

My first case study is the "Nation-Brain" thought experiment described by Block (1992) as an argument against functionalism. This is a best-case scenario for composite subjectivity, for it is stipulated to have conscious parts and yet be functionally identical to a human brain. This functional structure is realized not by neurons but by people: we have somehow persuaded the citizens and government of some sufficiently populous country to simulate a particular human brain using radios.

Each citizen is given a device which can send and receive signals to and from the other citizens' devices and given detailed instructions of the form "If you receive signals from such-and-such devices, send out signals to such-and-such other devices," for as many pairings of input and output as are needed to perfectly simulate the chosen brain. An appropriate human-like robot body is then wired up to be controlled by the citizens in a way that parallels the sensory and motor connections between body and brain. This body (call it the system's "avatar") walks around, responds to its environment, and expresses complex thoughts just like a human would, in virtue of several billion people pressing buttons.

Block thinks it counterintuitive that this vast group of button-pressing citizens must *itself* be conscious just like a human being. He concludes that pure functionalism is false, since the Nation-Brain and a normal brain could differ in consciousness but have all the same functional states. I am neutral on the truth of pure functionalism, but even if it is false, I do not think the Nation-Brain shows this.[1] **(p.191)** I think some versions of the Nation-Brain would be conscious, and the intuitive rejection of consciousness in the Nation-Brain stems from tacit acceptance of anti-combination. In this section I try to show how a combinationist can overcome that intuitive rejection.

The reason I say "some versions" of the Nation-Brain would be intelligent subjects is that some versions probably would not be, for reasons unrelated to anti-combination. Block distinguishes two forms of functionalism, according to how they formulate functional descriptions: capital-F "Functionalism" (which I will call "machine-functionalism") and "psychofunctionalism." The first identifies mental states with states of a certain kind of Turing machine,[2] while the second identifies mental states with whatever functional states are posited by a projected complete psychology. That is, one understands "functional state" in terms of a certain, well-defined sort of abstract system, while the other understands "functional state" as a place-holder for whatever the best scientific models of human psychology turn out to talk about.

We can accordingly distinguish two versions of the Nation-Brain: machine-functional and psychofunctional. The former matches a human brain just in those respects which would be captured by a Turing-machine table, the latter in any respects that psychology might have cause to appeal to. Since a Turing-machine table identifies its states solely by the output and state-changes they produce for any given input, the machine-functional Nation-Brain can be set up to give each citizen a very simple job:

> On one [giant satellite] is a bulletin board on which is posted a state card, i.e., a card that bears a symbol designating one of the states specified in the machine table. . . . Suppose the posted card has a "G" on it. This alerts the [citizens] who implement G squares—" 'G-men" they call themselves. Suppose the light representing input I17 goes on. One of the G-men has the following as his sole task: when the card reads "G" and the I17 light goes on, he presses output button O191 and changes the state card to "M." This G-man is called upon to exercise his task only rarely. (Block 1992, 70)

Clearly this is not how the human brain is organized: there is no single "bulletin board" which all neurons monitor, nor is there a single list of inputs such that **(p.192)** at any time there is one and only one input. A psychofunctional Nation-Brain would have to more accurately mirror the brain's setup, for instance by having many citizens active in parallel, and by having each citizen communicate

with a great many others—up to several thousand, if citizens are to play the role of neurons.

The reason for distinguishing these two forms is that even a functionalist combinationist could think that the machine-functional Nation-Brain is not an intelligent subject, on the basis that there is no intelligent *internal* functioning and very little integration of information. Each citizen need only detect two things (the input and the state) and need never learn anything of the doings of the other citizens. Changing the instructions given to one citizen makes no difference to the other citizens, except via changing what state is displayed on the "bulletin board." Without integrated information, there would not be enough conscious structure for the Nation-Brain's states to play their roles in intelligent functioning *in virtue of* their phenomenal character and representational content. Without conscious structure or coherence, the system would be what in chapter 5 I called a "Blockhead."

So let us focus on the psychofunctional Nation-Brain. Many people find it deeply implausible that this collection of button-pressing people should be conscious. Why? To some extent it may just be sheer weirdness, combined with the sense of vertigo elicited by supposing any visible, tangible object to literally be a conscious subject. (As Huebner [2014, 120] says, "It is hard to imagine that collectivities can be conscious; but, it is just as hard to imagine that a mass of neurons, skin, blood, bones, and chemicals can be . . . conscious.") But intuitive resistance to Nation-Brain consciousness seems to go beyond this (see, e.g., Knobe and Prinz 2008; Huebner et al. 2010), and I believe the reason is that the Nation-Brain is made of conscious citizens. This opposition to Nation-Brain consciousness based on its parts being conscious can be articulated into at least four distinct objections:

1. If the Nation-Brain is conscious, its consciousness must be distinct from the consciousness of each citizen, so where is this "extra" consciousness?
2. The citizen's experiences are not unified with one another, so the Nation-Brain's consciousness is massively disunified, and thus nothing like ours.
3. The citizens can report their experiences; if the Nation-Brain shares their experiences, why can't it report them too?
4. Consciousness in the Nation-Brain would be explanatorily redundant.

(p.193) Defending combinationism therefore means defending the consciousness of the Nation-Brain, and in the remainder of this section I will consider and rebut the above four objections.

6.1.1. Where Does the Extra Consciousness Fit?

Consider first the following sort of objection:

> *The Nation-Brain is nothing over and above the citizens, so it cannot have an extra consciousness in addition to theirs. Yet its consciousness is not identical to any of theirs, so it is not conscious at all.* (See, e.g., Searle 1990, 404; List and Pettit 2011, 9)

This objection could be summed up by saying that consciousness belonging to the whole system has "nowhere left to fit." Its experiences are not those of any particular citizen, since it may feel and think very different things from any of them. Yet it also cannot be somewhere "outside" them, since it is simply them and nothing more.[3] If the Nation-Brain were conscious, it would be a mysterious emergent mind somehow floating free above those of the citizens, or else it would be one of them. But the combinationist theory developed over the past few chapters shows that this need not be the case: the consciousness of the Nation-Brain is entirely grounded in and explained by that of its citizens, and all of its experiences are composed of subsets of their experiences.

Which of their experiences? To answer this question we must look primarily at what information is integrated across the Nation-Brain, which means looking at what information from each citizen is broadcast to others. What is primarily broadcast is the citizen's motor experiences of pressing buttons; these in turn carry information about their causes, namely the signals that citizen had received (i.e., certain perceptual experiences, seeing the lights flash or similar), and whatever mediated between the two (i.e., certain thoughts, memories, and decisions, such as "Ah, the gamma pattern—that means I need to signal to devices, um, let me see, 2435, 5784, and 2049!"). Call these three things together—the perceptions, the mediating thoughts, and the resultant actions—the citizen's "radio experiences." My first claim about the psychofunctional Nation-Brain is that those of its experiences which play a role in its intelligent functioning are composed exclusively out of the citizen's radio experiences.

(p.194) This excludes most of the experiences the citizens have (everything they think and feel and perceive that is unrelated to this zany scheme we have somehow compelled them to participate in) because information about them is not integrated and they consequently cannot be accessed by the whole for any cognitive purpose. Whether these informationally isolated experiences nevertheless still belong to the Nation-Brain depends on the choice discussed in section 5.5 of chapter 5, between "inclusionary" and "exclusionary" approaches to combinationism, divided by whether being an intelligent subject precludes having inaccessible experiences. If it does, then the Nation-Brain shares only the citizens' radio experiences, and no others. If it does not, then the Nation-Brain's experiences are simply all of those of its parts: a few billion consciousnesses experienced all at once.

Let us dwell on the inclusionary option a bit. This description seems ludicrous at first. Surely, someone might object, there is no coherent experience answering to the description “experiencing the full mental life of a few billion different people at once.” In a sense this objection is correct: there is no *coherent* experience here, because there is not enough information integration among these billions of conscious fields to make them align in any definite way. For the most part, they blend together, their joint content massively indeterminate, their joint quality nothing more than a constant inarticulate background blur. Only a few of these experiences are informationally integrated (the radio experiences), and so only a few of them combine in determinate ways so as to “line up” and reinforce each other. These experiences, and only they, coalesce out of the haze: the architecture of the system manages to bring a particular pattern into stable clarity, and the one it brings to clarity is the only one that can subserve intelligent functioning. And if the exclusionary approach is right, then this stable clear pattern is the entirety of the whole’s consciousness.

6.1.2. How Can the Whole Have Unified Experiences?

Consider next the following sort of objection:

> *The radio experiences of the citizens are not unified, since they belong to many different people. Thus if the Nation-Brain shared them, its consciousness would be massively disunified, and thus nothing like our own.*

What is true is that the informational interactions among different citizens’ radio experiences subserve intelligent functioning only at the Nation-scale, not at the citizen-scale. But the pure functionalist and the panpsychist will likely have different ideas of what this means. The panpsychist, appealing to the arguments of **(p.195)** chapter 3, can say that we need to distinguish phenomenal unity (i.e., subsumption by a state that is phenomenally conscious—a division-invariant relation) from richer relations like access-unity or global consistency (which are division-relative). Regarding phenomenal unity, the experiences of the citizens were already unified even before the Nation-Brain was established, just on account of their being interacting parts of a fundamentally conscious world. But the interactions among them were so information-poor that this unity did not enable access-unity (which is needed for an introspective impression of unity) at any scale above that of a single citizen: hence no subject was able to report the unity between different citizens’ experiences. And this is still the case for the citizen’s non-radio-related experiences. But with the radio devices being used, the citizen’s many experiences are now subsumed by a state which is able to subserve intelligent functioning, and in particular produce introspective impressions of itself as a single state, at the nation-scale.

The pure functionalist, by contrast, will want to explain phenomenal unity by reference to division-relative relations like access-unity, and so for them phenomenal unity will also be division-relative. Since the radio experiences are functionally related in the right ways at the nation-scale but not at the citizen-scale, this view implies that the experiences will be phenomenally unified relative to the one scale but not the other. Since the radio connections are integrating information, they will still be "pre-phenomenally bound" relative to all scales, but only at the scale of the whole Nation-Brain is this accompanied by phenomenal unity, and so only at this scale does this constitute representational unity.

What does this mean at the level of the individual citizens? If phenomenal binding involves informative mutual co-presentation, then my proposals in chapter 5 suggest that each citizen's radio experiences will co-present the others', which it seems plausible they will. Each citizen experiences the flashing lights and pressed buttons as manifestations of something larger, namely the radio cascades that cause or result from them. Each citizen, that is, will have some degree of awareness of the "thoughts" flowing through the system, so that their radio experiences are all mutually referential. This awareness of context will likely not be very informative, and may fall well short of what the citizens would be capable of if they dedicated themselves to understanding each other's participation in the Nation-Brain in detail. (This would be analogous to some of my neurons dedicating themselves to studying neuroscience, if that were possible.) Indeed, measured by degree of information shared, the unity between the radio experiences of one citizen and the next will be significantly less than that between the radio experiences of one citizen and other experiences of that citizen. But this does not impugn the relative unity of the Nation-Brain's consciousness, because the information integration **(p.196)** between two citizens' radio experiences is equal (by stipulation) to that between the experiential contributions of two neuron firings in a single citizen's brain. And since there are as many citizens as neurons, this is the relevant comparison.

6.1.3. Why Can't the Whole Report Its Parts' Experiences?

A third objection might go as follows:

> *If the Nation-Brain were undergoing its citizens' radio experiences, then it would be able to report them, and we could ask it what lights particular citizens were seeing or what buttons they were pressing. But since it is functionally equivalent to a human being, we know it cannot tell us.*

All of the Nation-Brain's experiences that subserve intelligent functioning are composed of experiences of seeing some lights flash, recalling some instructions, deciding to follow them, and pressing some buttons. Yet if we ask it, via its robotic avatar, whether it feels as though it is seeing lights flash and pressing buttons, it will usually say no, just as the human being it is modeled

after would usually say no.[4] Yet the radio experiences have their phenomenal character essentially: what it is like to undergo each of them cannot vary between the distinct subjects who undergo it. So it might seem that the Nation-Brain must both feel like it is seeing lights and pressing buttons, and not.

The solution to this seeming paradox comes from the doctrines of confusion and blending laid out in chapter 4. Just as each distinguishable experience of ours corresponds to a great many synaptic firings, which we cannot distinguish from one another and which consequently are experienced together as a single blended quality, so the Nation-Brain experiences cascades of many radio experiences as phenomenal blends, unable to distinguish different citizens' experiences from one another (and certainly unable to discern their internal components—particular seeings-of-this-light and pressings-of-this-button) and thus unable to discern or report that they have the character of visually perceiving something and pressing buttons in response.

But the Nation-Brain's sensation is not just any confused blur: it is a blur with a specific quality, based on the quality of the radio experiences and the particular ways they cause one another. Whatever a particular distinguishable sensation feels like for the Nation-Brain, it feels that way because of what the radio experiences **(p.197)** feel like, filtered and blended based on the way information is integrated among the relevant citizens. Most likely, none of its sensations will feel quite the same as our sensations do; in particular, if panpsychism is true, then our experiences are blended out of whatever experiences a neuron has when it fires, while its experiences are blended out of light-seeing and button-pushing experiences. Yet because the informational structure is the same in both cases, there will be structural similarities in the number of qualities that can be distinguished, in which qualities are similar to which, and so on.

6.1.4. Why Bother with Consciousness in the Whole?

A final objection might run as follows:

> *The consciousness of the Nation-Brain is superfluous: it does no causal or explanatory work that is not already done by the consciousness of its component citizens*.

This objection takes the possibility of reduction (deducing every fact about the Nation-Brain from facts about its citizens and their relations) as a reason for elimination (denying that the Nation-Brain is conscious). If this sort of reasoning were valid, however, the possibility of deducing every fact about, say, an organism from facts about its component molecules would be a reason for denying that organisms exist (cf. Huebner 2014, 126–182). I take it that such a deduction is possible in principle, but that organisms obviously exist, and conclude that reduction is not enough for elimination.

This objection would be well-motivated if the Nation-Brain's consciousness were not fully grounded in that of the citizens; if they were metaphysically independent things then we would have to ask which of the two was causally responsible for any given result. But in fact this question makes no more sense for the Nation-Brain and its citizens than it does for a brain and its neurons. Of course there may remain some questions of "whole or parts" that do make sense, such as whether an explanation pitched at one level is more illuminating or more efficient or more satisfying than an explanation pitched at another level. But in this sense it is clear that the Nation-Brain will sometimes be the right level at which to explain things, since explanations that mention only the citizens will be swamped with irrelevant detail, just like explanations of human action that mention only neurons. Answering "Why did the avatar ask for my phone number?" with "Well, citizen 2567348 pressed buttons 32, 3, and 75, and that sent signals to citizens 5543, 890897, and 43455, who then . . ." may be a strictly true answer, but it is not a **(p.198)** useful or satisfying explanation compared to "Because the Nation-Brain wants to date you."

In summary, functionalist combinationism suggests that there is indeed something it is like to be the (psychofunctional) Nation-Brain. It is either like being billions of different people all at once (the inclusionary option), or it is only the determinately structured aspects of the latter (the exclusionary option). Either is hard to make sense of, because simply superimposing billions of streams of consciousness will, by default, have the effect of blurring out every interesting detail of each. What gives the Nation-Brain experiences with complex determinate content is that out of this vast tissue of overlain experiences, certain patterns coalesce because of the way that certain of the billions of streams are aligned with one another. The form of these coalescing patterns depends on how information is integrated, but what fills out the pattern is the consciousness inherent in the system's parts, namely the citizens who compose it.

6.2. Composite Subjectivity in Social Groups

The Nation-Brain is a thought experiment, but we can also ask about composite subjectivity in actual nations, and more broadly in social groups like corporations, communities, and couples.[5] We can begin by asking how such entities, considered as putative conscious subjects, differ from the Nation-Brain.

6.2.1. Information Integration in Nation-Brains and Nations

First, the Nation-Brain plausibly integrates more information as a whole. By stipulation, it integrates as much information on the nation-scale as a human brain does on the brain-scale, while most actual nations arguably integrate significantly less. Consider: if we made some arbitrary intervention on one individual or small group, how much would the behavior of a randomly chosen other citizen be affected, and how much would the way it was affected depend on the particular sort of intervention we made? Some interventions would

produce noticeable effects (e.g., a particularly gruesome mass murder might cause news reports), but most would not—most changes to the mind or body of most individuals would make only a negligible difference to the lives of most other individuals. In pre-twentieth-century societies this would be even more true, and in a random large group of people (as opposed to a nation) even more.

(p.199) At the same time, actual nations make much more use of the cognitive capacities of their parts than the Nation-Brain. The Nation-Brain, as befits a hyperbolic thought experiment, treats its members as mere button-pushers, following mechanical rules and responding only to lights flashing on their handheld devices. In effect, it tries to make them behave like neurons, and thus ignores all the ways that their capacities exceed those of neurons. Actual nations can afford no such wastefulness. In real groups, more of the information in each individual's brain can *potentially* be integrated into the system's functioning, because the channels of information transfer between citizens are richer and more flexible than the button-pushing that the Nation-Brain relies on. In fact these channels are simply our familiar modes of human communication, chief among them language and the kind of face-to-face interaction in which we "read" the moods, intentions, and other mental states in someone's expression, body, voice, etc. Although our everyday lives often acquaint us with the limitations of these methods, they are really remarkable compared to virtually anything else known to exist. As Schwitzgebel (2015, 1707) says:

> People exchange a lot of information. How much? We might begin by considering how much information flows from one person to another via stimulation of the retina. The human eye contains about 10^8 photoreceptor cells. Most people in the United States spend most of their time in visual environments that are largely created by the actions of people (including their own past selves). If we count even 1/300 of this visual neuronal stimulation as the relevant sort of person-to-person information exchange, then the quantity of visual connectedness among [the USA's $3x10^8$ citizens] is similar to the neuronal connectedness within the human brain (10^{14} connections). Very little of the exchanged information will make it past attentional filters for further processing, but analogous considerations apply to information exchange among neurons.

The familiar consequence of these connections is that humans living in societies can develop an incredibly rich idea of each other's experiences, and of how their own experiences relate to theirs. We can have experiences like, for example, competitive gameplay, in which every step of our strategic thinking and feeling is inflected by our awareness of what our opponent is thinking and feeling, based both on past experience with them and on ongoing observation of their moves and demeanor. As a result, the experiences of a small set of people interacting in a real society could potentially form a complex experience with a more complex determinate content than the same people could in the Nation-

Brain, where each **(p.200)** knows the others only as the anonymous receivers of the signals they send, and the anonymous senders of the signals they receive.

Because of their capacity to make fuller use of the information-processing and information-sharing capacities of individual citizens, actual groups could in principle come to have a composite subjectivity that far exceeds an individual's in complexity and power (and which consequently also exceeds the Nation-Brain's). But in practice this potential is largely unrealized, for at least the following two reasons. First, while humans are very good at integrating information with individual others, their methods of doing so don't scale up very well. While two people might readily establish a detailed shared understanding of an environment, task, or topic, it is much harder for ten people, even harder for a hundred, and so on (cf. Shapiro 2014). By contrast, that Nation-Brain's mechanisms, which seem less impressive at small scales, are better at coordinating functions and integrating information across the entire system. (They are, in fact, precisely as good at that as our neural mechanisms.) Second, humans largely serve their own, individual goals, which only sometimes require integrating information with others, and often strongly tell against doing so. Our brains work very differently: neurons do not decide on each occasion whether signaling to others is "worth it" for that neuron.[6] So the kind of superhuman mega-intelligence which could in principle arise from human social groups would require radical transformations both in the mechanisms by which members interact (they would need to scale better) and in the incentives for which members act (they would have to secure patterns of behavior that prioritized system-scale over individual functioning). Perhaps the future of neuroscience, information technology, or social science will yield developments that might effect such transformations. I take no stand here on whether such a speculative "collectivization" of the human mind deserves to be seen as utopian, as dystopian, or as a mix of both; I intend simply to throw our actual situation into sharper relief by comparison.

6.2.2. What Is It Like to Be a Nation?

What does the generally low level of information integration in human societies mean for their collective consciousness? It may well mean that they are not **(p. 201)** conscious at all: if pure functionalism is true, there is a (vague) minimum level of functional organization and complexity required for any kind of consciousness, and it may well be that no human groups currently meet that level (though see Schwitzgebel [2015] for an argument that they do). To put the same point differently, individual experiences may never meet conditional experience inheritance's standards of control, sensitivity, and coordination relative to any actual social groups. But even then, there remains the possibility of a group becoming sufficiently better organized than our present-day groups—and if panpsychist combinationism is true, then even present-day nations will be

conscious, though perhaps they do not qualify as intelligent subjects. So let us just ask: If a human nation were conscious, what would it be like to be it?

Start with the many consciousnesses of the citizens, considered as undergone by a single entity at once. For simplicity, let us first consider just two citizens' fields of consciousness, undergone together. Each is structured in the usual way: senses which disclose a continuous surrounding space, attention focused either on some aspect of this perceived world, or on some inner thought or mental image, and so on. If we try to imagine these undergone together —"superimposed" on one another, according to my favored metaphor—how should we line them up? Is the attentional focus of one phenomenally bound with the focus of the other? Do we superimpose their visual fields so that "up" and "down" coincide, or so that one is upside-down relative to the other? Or do we superimpose the visual field of one onto the bodily sensations of the other, so that each colored object seen by one is phenomenally bound to some portion of the other's skin? Questions like this abound and cannot be given any *a priori* answer. By default, the two fields blur together, so that the content of the total field is massively indeterminate and experienced as a blend of all the qualities of the experiences of both.

When the two people are actually interacting, they begin to somewhat integrate information about their minds, and this indeterminacy is correspondingly somewhat reduced. This involves each experiencing a form of co-presentation whereby their visual and auditory perceptions of the words or behavior of the other appear as revealed aspects of something with other experiences as its concealed aspect. Functionalist combinationism says that because their experiences integrate information and mutually co-present, they are pre-phenomenally bound; if they are also phenomenally unified they will be phenomenally bound, and thus representationally unified. Under normal circumstances two people's experiences are surely never sufficiently functionally unified to count as phenomenally unified by the standards of pure functionalist combinationism, though according to functionalist-panpsychist combinationism they are in fact phenomenally unified from the very beginning (like all experiences). If the latter is correct, then they **(p.202)** do form a composite experience, with a somewhat determinate representational content. But since the average level of information integration in human societies is not very high, a great deal of indeterminacy remains. The composite experience might be compared to the feeling of looking around a foggy environment, with faint, confusing images appearing and disappearing out of the haze. (Of course it is not a visual image, so this analogy is quite limited.)

To make a nation an intelligent subject would require that it display intelligent functioning at the nation-level, which is arguably still a futuristic dream. But sometimes the mental states of members of some groups do inch a bit closer to group-level intelligent functioning, by guiding the activities of other individuals

in unison with their individual subjects' activities, giving rise to a range of phenomena which have been studied under the heading of "collective intentionality."

6.2.3. Collective Intentionality

The study of collective intentionality starts from the observation that in everyday life we routinely speak of, for example, a crowd trying to topple a statue, a company trying to maximize its profits, or an army trying to find out where its enemies are hidden. Obviously some of this everyday talk is simply anthropomorphism, or self-serving ideological blather (e.g., a politician attributing their popularity to having "tapped a reserve of almost-forgotten hope in the nation soul"), but recent philosophers have devoted considerable effort to defining conditions under which it is well-motivated and rigorously defensible. I will review three particular forms of collective intentionality which are studied in this literature, noting briefly how functionalist combinationists might think about the consciousness involved: I will label them "joint intentionality," "authority relations," and "joint commitment."

By "joint intentionality" I mean cases where multiple people all have mental states of the same sort with the same content, and stand in relations which allow the set of their matching states to play the same role in guiding behavior for all of them as it does for each of them. This has been discussed most fully in relation to joint intentions and their role in joint action. Intuitively, there is a difference between people doing something together and their doing the same thing independently, even if the bodily movements involved are identical. Even in simple cases, we can tell the difference between two people walking somewhere together and their merely walking to the same location beside each other (Gilbert 1990).[7]

(p.203) Several rival accounts of joint intention have been offered,[8] but the one I prefer is Bratman's (1997, 2009, 2014), on which for people to jointly intend to do something is (roughly) for them each to intend that they all do it, for them each to intend this *because* the others do, and for each to know all this of the others. In short, a joint intention is a set of corresponding interdependent intentions. For our purposes the significant thing is that joint intentions involve information integration: since each must know that the others intend, and base their own intention on this knowledge, the mental state of each must reflect (carry information about) the mental state of the others. Each individual's experience accordingly features a highly informative form of co-presentation: they are indirectly aware of the others' intentions, as things-I-am-not-directly-aware-of, yet also have a fairly specific idea of the nature of those intentions, and base their own intentions on them.

According to the theory laid out in chapter 5, this means that from the whole's perspective, the experiences that relate to this joint intention (experiences of hoping, acting, monitoring progress, etc.) will connect in a relatively determinate way, potentially giving rise to a composite experience that is not simply a blend. Moreover, these interdependent experiences of the participants then serve a recognizable cognitive function, namely guiding action. So in cases of joint action, the massively confused experience of the group whose members act together begins to resolve itself a little. To put it very roughly, the more people who are involved, and the more sensitive each is to the others' progress, the more reasonable it becomes to say that the whole system (a nation, or whatever social group we are considering) consciously intends their goal and acts to achieve it.

What about states other than intentions? There is less work on this, and it is harder to identify a plausible set of sufficient or necessary conditions (but see Gilbert 2002; List 2018). Indeed joint intentions are likely to be the most common joint mental states, since intentions are the most directly action-guiding states, so people will most easily converge there when they need to work together. But people can also potentially be regarded as having, say, joint beliefs and desires, maybe even joint emotions, when people's individual instances of these states are connected in a way that lets them play the cognitive role typical of belief (or desire, emotions, etc.) for the whole group.

The second form of collective intentionality involves some individuals acting in a way that is guided by the mental states of other individuals, independently **(p. 204)** of whether they themselves have a corresponding mental state of that sort. For instance, if A will obey any instruction B gives them, then B's intentions can guide both of their actions, independently of what intentions A has.[9] At the extreme, a whole nation might come to behave in a consistent, intelligent way because everyone there follows without question the instructions of a single dictator.[10]

Such "authority relations" allow the experiences of individuals to play cognitive roles in large-scale social systems. Yet the information integration involved might nevertheless be very modest, because those who obey get only the information about their particular tasks, and those who command get only the information of assent. The co-presentation involved is less informative: each knows that the other has some hidden inner life, but knows less of the content of this inner life.

A composite subject whose component subjects coordinate in this manner might consequently have a different sort of consciousness than one whose parts relied on joint intentionality. The former might be "phenomenally impoverished" relative to the latter, rather like if we sometimes had perceptions or emotions which felt "faint," dull, and affectless, but which still had the same effect on

what we did and thought. In the extreme case of dictatorship described above, the nation would think and act on the basis of states it barely felt. That is not to say that these states are not conscious—the dictator is conscious, after all, and their experiences are inherited by the whole, so that the nation thinks and acts based on them. But the phenomenology of these experiences is not in proportion to their role. Thus even if a nation were to bend to the dictator's will perfectly, ensuring intelligent functioning, it would arguably not be an intelligent subject (or perhaps would be only a borderline case of one) for the same reason that the machine-functional Nation-Brain is not: its intelligent functioning is not based on conscious structure. It would be somewhere in between an intelligent subject and a "Blockhead."

A third form of collective intentionality arises when people jointly undertake to "act on" states that they all understand but which they may not individually feel. For instance, a committee might resolve that, when acting and speaking in their capacity as members of that committee, they will act and speak as though some proposition P is true, independently of whether most or any of them think that P is actually true. I will call this "joint commitment," borrowing the phrase **(p. 205)** from Gilbert (1990, 2000, 2002, 2009), though applying it somewhat more narrowly than she does. On her account, any collective intentional state (a belief, an intention, even an emotion) exists when two or more people jointly commit to acting in accord with that state. I do not know if this is a good way of thinking about collective intentionality in general, but it does seem a good model for cases like the committee.

Joint commitments fit better than either joint intentionality or authority relations for interpreting things like laws, charters, contracts, and other written documents which coordinate people's thoughts and action. If a number of people have committed to a document which mandates that certain steps be taken to accomplish a certain goal, that can explain their acting to accomplish that goal by means of those steps, even if none of them particularly care about that goal for its own sake.

The component experiences involved in joint commitment are in some respects not that different from those involved in authority relations: the individuals allow their behavior to be guided by a certain mental state, while relating to it as something external and not their own. The difference is that these "other-directed" states do not have any individual's mental state as their focus: there is no one who commands, only many who obey. What was said above about collective intentionality based on authority applies even more here: the system functions intelligently but may have impoverished phenomenology.

6.2.4. The Instability of Real-World Group Minds

These three forms of collective intentionality are often combined (e.g., a committee might act jointly to write a charter to direct the many other individuals over whom it has authority). Ordinary human societies are a patchwork built up from diverse forms of social interaction, and so if a nation were conscious it would have some degree of conscious structure, though a very different sort of structure from our own. In some cases particular institutions or communities may even be engineered to have a sufficiently stable and consistent set of dispositions that it makes sense to describe them as rational agents with a single set of beliefs and desires which they consistently act on (cf. List and Pettit 2011; Tollefsen 2014).

But the extent of collective intentionality should not be overstated. In particular, episodes of collective intentionality tend to be less stable the larger in scale they are, and so insofar as a nation could be said to have "thoughts," they will often be too unpredictable and unstable for us to speak of stable dispositions like beliefs or desires. Because the arising and subsiding of particular episodes of collective intentionality depends so heavily on the interplay of unpredictable factors **(p.206)** that the whole has no cognitive access to (i.e., the private decisions of individuals about how to interact with each other), the whole's stream of consciousness will lack coherence and stability. Moreover, even if a given collective state managed to persist for a long time, the *role* it could play would depend on its relation to other collective states; even if one institution, say, manages to retain a stable and cohesive plan of action for years, they will get little done except insofar as they can persuade, compel, or manipulate other parts of society. It is as though a person, even when they did manage to hold a single idea in mind for some length of time, then found that at different times this idea might or might not be reportable, or usable for planning, or suitable for recording into memory, or might be able to guide only behavior involving the left half of their body.

The resulting degree of instability, in both the arising and subsiding of thoughts and experiences and in the roles they play, makes the structure of any possible nation-consciousness barely imaginable to compulsively coherent beings like us humans. Perhaps the best analogy is with "lowered modes" of consciousness, such as delirium, dreaming, epilepsy, or coma. Bayne and Hohwy (2016) discuss the challenges of defining and distinguishing such modes, proposing to do so by a mixture of whether complex contents can be represented, whether attention is controlled and distributed normally, and whether the contents of consciousness are available for the normal range of cognitive and behavioral uses (cf. Bhat and Rockwood 2007). A social group's consciousness, considered as a whole, will likely count as radically deficient on all three scores.

I said in chapter 1 that I think the widespread intuitive resistance to the very possibility of collective consciousness is partly driven by anti-combination, and thus mistaken. But I think resistance to the idea that actual human societies could be conscious has another, more correct basis. Even if human societies are, in the strict and fundamental sense, conscious (as functionalist-panpsychist combinationism says), their consciousness is rarely useful to think about. To explain why any particular event happens, it will almost always be more useful to appeal to facts about individuals. The instability of the "nation-mind" means that there are few useful generalizations that can be made about its cognitive functioning: whether it will have a certain thought, and what role that thought will play, depend too heavily on the unpredictable complexities of individuals or small groups and their alliances, friendships, feuds, moods, and estimations of personal cost and benefit from any given collective act. And if we need to make reference to the details of individuals and smaller groups anyway, we might as well just use them as our explanation.

Yet this lack of explanatory usefulness is not a reason to deny the *existence* of collective consciousness; it is only a lack of one sort of reason to posit it. If the **(p.207)** existence or possibility of collective consciousness follows from the best metaphysics of consciousness—in particular, if it follows from the best explanation of what makes *us* conscious—then we should shrug and observe that many things exist which it is rarely useful to talk about.

6.3. Composite Subjectivity in the Split-Brain Patient

The results of severing the corpus callosum, which connects the two cerebral hemispheres, are among the most perplexing phenomena in the study of the brain: a person who seems most of the time perfectly normal but at other times operates as though two quite separate agents lived in their head, responding to inputs from different subsets of the body's sense organs, using different subsets of the body's motor capacities. Since its first description, this phenomenon has prompted philosophical discussion: the options canvassed include:

1. The two hemispheres are distinct subjects, with two separate streams of consciousness, who somehow coordinate their activities perfectly almost all the time (call this the two-subjects view, with the other seven being different one-subject views; see, e.g., Puccetti 1973, 1981; Schechter 2009, 2015).
2. There is one subject, who undergoes two distinct streams of consciousness, such that all the elements of each stream are unified with each other but not with those of the other stream (call this the two-streams view; see, e.g., Tye 2003; Bayne 2010).
3. There is one subject, who undergoes a single unified stream of consciousness but has an impaired ability to access all its contents together (call this the one-stream view; see, e.g., Bayne and Chalmers 2003; cf. Schechter 2010).

4. There is one subject, whose consciousness is partially unified, so that experiences based in different hemispheres, though not unified with each other, are both unified with certain core experiences possibly based in subcortical structures (call this the partial-unity view; see, e.g., Lockwood 1989, 1994; Hurley 1998).
5. There is one subject, who has two simultaneous streams of mental processing, of which one is conscious and the other unconscious at each moment, with consciousness switching between the two (call this the switch view; see, e.g., Levy 1990; Bayne 2008, 2010; cf. Schechter 2012b).
(p.208) 6. There is one subject, who has two streams of consciousness under experimental conditions when stimuli are carefully segregated between sense organs, but one stream the rest of the time (call this the contextualist view; see, e.g., Marks 1980; Tye 2003).
7. The left hemisphere (or more precisely, whichever one has language capacities) is a conscious subject, and the other hemisphere is an unconscious automaton (for discussion see Nagel 1971, 402–404; Tye 2003, 117).
8. The left (language-using) hemisphere is a conscious subject, and the other hemisphere, though conscious, does not have sufficiently unified experiences to qualify as a subject (for discussion see Nagel 1971, 403–405).

Functionalist combinationism does not by itself support one of these interpretations over the others; rather, it partially dissolves the seeming incompatibility between them, allowing us to see them as alternative ways of describing the same facts.

The upshot of combinationism is that we should think of the questions “How many subjects?” and “How many streams of consciousness?,” posed about the split-brain case, in something like the way that we think of the question “How many brains?” This question is not easy to answer, but it is easy to leave unanswered: it does not demand a single definite answer because when we recognize the respects in which it is appropriate to describe the situation as involving one brain, and also recognize the respects in which it is appropriate to describe the situation as involving two brains, we have understood how things stand. It would be a mistake to insist on seeking for some further, objective fact —the number of brains. We can see that this would be misguided because we can see that there is a single physical system here which has many of the features associated with our concept “brain,” and that it has two (overlapping) physical parts which each have many of the features associated with our concept “brain,” and that this is all there is to the system. The fact that brain-like things can be parts of other brain-like things is crucial to this equanimity: combinationism allows us to take the same attitude toward subjects.

More precisely, I believe that views 1–5 of the above all contain important elements of truth, and that functionalist combinationism lets us affirm these all together. In this section I will first summarize the experimental results pertaining to split-brain patients, then review the ways that a functionalist combinationist could partly agree with each of the five noted views.

(p.209) 6.3.1. A Quick Overview of the Split Brain

First, the known facts. The cerebral hemispheres are connected by a bundle of nerve fibers called the corpus callosum, as well as by certain other connections, and most fundamentally by their both being linked to the nonhemispheric parts of the brain. The “split brain” results from severing the front part of the corpus callosum (with or without cutting certain other connecting areas), but leaves much of the subcortical brain intact; we have no evidence that the entire brain can be bisected without killing the patient. Thus there is an important sense in which it is misleading to think of the split-brain patient as having two complete and separate half-brains: what they have two of is what I will call “hemisphere systems,” each consisting of the unsplit nonhemispheric parts and one or other cerebral hemisphere. Since the hemisphere systems both contain the nonhemispheric parts, they overlap and are not discrete from each other.

From cases where an entire hemisphere is destroyed or removed, we know that what remains—a single “hemisphere system”—can support more or less normal human consciousness by itself (albeit with various incapacities relating to, for instance, controlling both sides of the body). This is important, because it means that even non-panpsychists have good reason to regard hemisphere systems as themselves candidates for being (intelligent) conscious subjects. Hemispheres by themselves, without the nonhemispheric parts, seem incapable of doing anything interesting, and so we might doubt that they are conscious subjects (unless we are panpsychists and prepared to ascribe some form of subjecthood very widely).

Most of the time split-brain patients show few abnormalities in thought or action: the experiments that make the split-brain patient so perplexing involve making efforts to present different stimuli to the sensory channels which primarily feed to each hemisphere; for the left hemisphere this includes touch over most of the right side of the body, the right ear, the right half of the visual field, and the left nostril; for the right hemisphere, the reverse. Under these circumstances, we find striking dissociations, in particular failures of access and representational unity. Access-unity fails in the sense that certain contents are unavailable for guiding the same processes as other contents; for example, when we show the word “key” to the right hemisphere, it cannot be verbally reported (because the left hemisphere controls language), but if we ask the patient to pick up the object they saw the word for, and let them use their left hand (controlled by the right hemisphere), they will readily pick out a key. Representational disunity here means the apparent lack of any consciousness of

conjoint contents spread across both hemispheres; for example, if the patient is shown the word “keyring,” with **(p.210)** “key” going to the right hemisphere and “ring” going to the left, they will pick out a key (with their left hand) and a ring (with their right hand), but not a keyring. They will even report having seen only the word “ring,” and deny having seen either “key” or “keyring.” What are we to make of this?

6.3.2. The Truth in the Two-Subjects View

Let us begin with the two-subjects view. On this account there are two conscious beings at work in any given split-brain case: one which undergoes the experiences corresponding to neural activity across the left hemisphere and the nonhemispheric parts (call it “Lefty”), and one which undergoes the experiences corresponding to neural activity across the right hemisphere and the nonhemispheric parts (call it “Righty”), with any experiences corresponding to activity just across the nonhemispheric parts being shared by both subjects. Let us use the term “hemisphere subjects” for the type of thing that Lefty and Righty are. Lefty perceives everything coming into the right-hand set of sense organs, and responds as best it can using the right-hand set of motor organs, while Righty does the same with the left-hand sets.

We can tell that Lefty and Right are distinct, on the two-subjects view, because under experimental conditions we can in effect address “questions” (stimuli) to them independently and receive different sets of “answers” (responses). But most of the time their distinctness is hidden by the near-perfect synchronization of their activities, accomplished by some mixture of subcortical connections (that allow the activity in one hemisphere to affect that in the other), subcortical processing (neural activity in their shared part, which likely plays a significant role in integrating things like motor control and emotion), and detection of peripheral sensory cues (e.g., Lefty picks up on what Righty is doing, by seeing the left hand’s movements in the right visual field). Nagel (1971, 406) considers this view, pointing out that sometimes two people do seem to work “as one,” as in “pairs of individuals engaged in a performance requiring exact behavioral coordination, like using a two-handed saw, or playing a duet.” Perhaps the two hemispheres are like this? But Nagel worries that this position is unstable: “If we decided that they definitely had two minds, then [why not] conclude on anatomical grounds that everyone has two minds, but that we don’t notice it except in these odd cases because most pairs of minds in a single body run in perfect parallel” (409). I think Nagel’s worry here is cogent (though for criticism see Schechter 2018): if we accept that there can be two distinct subjects despite its appearing for all the world as though there was only one, we seem to lose any basis for confidence that the same thing is not happening in other cases. He continues:

(p.211)

> In case anyone is inclined to embrace the conclusion that we all have two minds, let me suggest that the trouble will not end there. For the mental operations of a single hemisphere, such as vision, hearing, speech, writing, verbal comprehension, etc. can to a great extent be separated from one another by suitable cortical deconnections; why then should we not regard each hemisphere as inhabited by several cooperating minds with specialized capacities? Where is one to stop? (Nagel 1971n11)

Where indeed? For a noncombinationist, the postulate of two minds is an alternative to the postulate of one mind, and the postulate of a dozen minds is an alternative to the postulate of two. And so the noncombinationist has to choose one specific number, while having just given up any basis for being confident in their choice. But while this is a problematic result for noncombinationists, it is no problem at all for the combinationist, whose ontology already allows for multiple distinct subjects making up larger composite subjects. The combinationist has no problem saying that the split-brain patient is two, or more, subjects because they have no problem saying that each of us is two, or more, subjects. That adds to, rather than replacing, the knowledge that each human being as a whole is a subject, just as our learning that our cells are to all intents and purposes organisms can add to, rather than replacing, the knowledge that each human being as a whole is an organism.

The presence of two subjects in the split-brain case is thus changed from a uniquely paradoxical fact into part of the unremarkable background. What is distinctive about the split-brain case, relative to the normal case, is not the presence of two subjects, but the change in their mode of interaction. Where before they accomplished a remarkable degree of information integration through multiple mechanisms including, in particular, the corpus callosum, now they must rely on fewer of those mechanisms and correspondingly integrate information less completely and less reliably, with more dependence on external feedback (like seeing the left hand in the right visual field) and redundancies of information in their environment (like a scene which looks the same way when it falls in the right visual field at one moment as it does when it falls in the left visual field a moment later). In experimental contexts, they are deprived of these environmental feedbacks and redundancies, and we observe the resulting failures of integration.[11]

(p.212) 6.3.3. The Truth in the One-Subject View

But one-subject views are not for this reason mistaken. There is a good basis for calling the split-brain patient a single subject, and for talking about what “he” or “she” is experiencing—just as there is a good basis for calling someone with an un-split brain a single subject. The most direct argument for a single subject is that this is overwhelmingly how both the patients themselves and all those who interact with them think of the case. That is a relevant consideration not because what people spontaneously think is always right, but because it reflects

the explanatory and predictive usefulness of thinking of them as a single subject —as a single planner, a single rememberer, a single holder of beliefs and desires, and so on. In particular, when we count people (in a vote, say, or to apportion resources), it will not be a good idea to count the split-brain patient as two, or three, or any greater number. The fact that “one subject” is the most explanatorily and predictively useful form of description is a powerful, though defeasible, reason to think that it captures something about reality.

Once again, the combinationist has no problem making sense of this. Although the loss of the corpus callosum impairs the hemisphere subjects’ ability to integrate information between them, it still leaves them with plenty of channels for doing so, and so there is still a composite subject composed of them whose experiences have highly complex, determinate content. Moreover, for some reason yet to be fully understood, the cognitive uses made of this content are still largely cross-hemisphere; for example, the patient makes plans that will be carried out with both of their hands, and thus presumably with some involvement by both hemispheres, and things they have learned are available to guide the actions of both hemispheres. The main reason not to regard the whole system as a single conscious subject is simply that there are also good reasons to regard it as two subjects, and for a combinationist that is no reason at all, since the two subjects can compose a third, sharing their experiences.

6.3.4. One-Stream, Two-Streams, and Partial-Unity Views

But what about the unity of consciousness? Even if we have answered the question “One subject or two?” with “Both!,” there is still the question of whether the streams of consciousness associated with the two hemispheres are unified with one another (as the one-stream view holds), not unified with one another (as the two-streams view holds), or somewhere in between (as the partial-unity view holds). **(p.213)** There are compelling arguments that seem to point in all three directions. First, there is evidence of disunity that seems to tell against the one-stream view: the above-mentioned phenomena of functional and representational disunity. The argument against the one-stream view is simply this: surely if experiences in the two hemispheres were phenomenally unified, they would also be representationally unified and access-unified? If phenomenal unity doesn’t secure such connections, what is the point of it (Bayne 2008, 286, 2010, 197–199)?

On the other hand, many functions do seem to remain representationally and behaviorally integrated in split-brain patients, though the details vary among individuals (Bayne 2008, 287–290; Schechter 2015; Pinto et al. 2017). Some can compare stimuli presented to the two hands, but not to the two halves of the visual field; that is, they can indicate whether their two hands are feeling similar things, but not whether the same images are appearing on both sides of their visual field (Gazzaniga and Freedman 1973). Since they can also indicate whether the right hand is feeling something similar to what is seen in the right

visual field (and the same on the left), this seems to show an intransitive pattern of comparative capacities. There is also the fact that some sensory information is not hemisphere-specific, such as the sense of touch in the head and neck. This poses a challenge to the two-streams model, since if the best explanation for representational and access-disunity is phenomenal disunity, it seems that the best explanation for representational and access-unity is phenomenal unity. When normal people judge whether two stimuli are similar, they do it by having experiences of each, and having a subsuming experience that unifies both. Why think that split-brain patients do it in a different way?[12]

The upshot of these two sets of evidence seems to be that some experiences of the split-brain patient are phenomenally unified, and others phenomenally disunified, as the partial-unity view says. But the partial-unity view is vulnerable to its own "objection from inconceivability," based on its treating phenomenal unity as nontransitive (Bayne 2010, 43–45). What is it like to be the split-brain subject, on the partial unity view? Suppose there are three experiences, A, B, and C, where A and B are unified, and B and C are unified, but not A and C. If the split-brain patient has one phenomenal field that contains A and B, that field cannot also contain C (since A and C are not unified). Yet if they have another phenomenal **(p.214)** field that contains only B and C, we leave out A, and moreover they seem to be experiencing B twice over. (How would their consciousness be different if B were two distinct experiences with the same character, one unified with A and one unified with C?) In short, there is no set of experiences that can be identified as "what it is like to be" that subject. The set {A, B, C} is not fully unified, and so has no overall phenomenal character, while both {A, B} and {B, C} leave something out. So there appears to be no coherent way to conceive of this subject's experience from the inside.

The choice among these three views largely depends on the question of how phenomenal unity relates to other unity relations, and combinationism is compatible with different answers to this question. Phenomenal unity might be distinct from, and dissociable from, other unity relations; panpsychist combinationism takes this view, and concludes that phenomenal unity is likely pervasive in the material world. This suggests the following view: all the split-brain patient's experiences are phenomenally unified (just as all experiences are), but richer unity relations (access-unity, representational unity) hold only in the "Y-shaped" pattern claimed by the partial unity account. Experiences based in the nonhemispheric parts are richly unified with those based in the right hemisphere and with those based on the left hemisphere, but the latter two are not richly unified with one another (though they would be in an intact brain).

The view suggested by functionalist-panpsychist combinationism is technically a "one-stream" view, since it affirms phenomenal unity between the hemisphere subjects, but it is in some ways more like a partial-unity view. Phenomenal unity in the split-brain seems less significant if such unity is everywhere, and the more

significant sorts of unity are admitted to hold only within a hemisphere, or between a hemisphere and the nonhemispheric brain areas.

Considered as a partial-unity view, functionalist-panpsychist combinationism provides a distinctive answer to the objection from inconceivability, concerning what it is like to be the split-brain patient as a whole. They experience the whole set of experiences belonging to both hemispheres, but in such a way that the hemisphere-specific experiences are only weakly, if at all, phenomenally bound. Since they are are not “lined up” with one another, they have no determinate conjoint content. Of course, some information is integrated across the hemispheres, and some experiences may be primarily dependent on subcortical structures; these will have definite content and structure from the whole’s perspective, and so can be unproblematically ascribed to them. But even though these are continuous with the experiences specific to the hemispheres, the latter are blurred out by their low-information interactions, so that to the extent that the experiences based in the two hemispheres differ, the split-brain patient undergoes neither with any distinctness. **(p.215)**

What if phenomenal unity is not a distinct, primitive relation but instead more tightly bound up with functional and representational relations (as pure functionalist combinationism says)? Then whether the split-brain patient has unified consciousness depends on the details of those relations—what kinds of interaction, what kinds of information transfer, what kinds of mutual accessibility are required to set up phenomenal unity? Are they transitive or not? Is there a single set of necessary relations, or are there many different ways to get phenomenal unity? To these questions of detail pure functionalist combinationism offers no specific answer, and so it does not by itself decide among one-stream, two-streams, and partial-unity accounts.

What combinationism adds to the debate is that it treats conscious unity as both a between-subjects relation and a within-subject relation. This means that one-stream and partial-unity views become compatible with two-subjects views as well as one-subject views (and, indeed, with three-subjects views). Each hemisphere system might be a conscious subject and nevertheless have all or some of their experiences phenomenally, functionally, or representationally unified with those of the other hemisphere system.

What is this like—what is it like to be a hemisphere subject in (say) a phenomenally unified split brain? Functionalist combinationism suggests that it is like undergoing a set of experiences all richly unified with one another (experiences based in that one hemisphere and in the nonhemispheric parts), but also having an unusual sort of co-presentation, reflecting the impact of the other hemisphere’s experiences. Since these other experiences are richly unified with experiences based in the nonhemispheric parts, but not with experiences based in the other hemisphere, they will be informatively co-presented by the

former but not by the latter. For an idea of what this might mean, consider an example where the patient feels a touch moving across their face (processed subcortically, at least at first) while also seeing the object that touches them as it moves from their left visual field to their right visual field (processed first in the right hemisphere, then in the left). Assuming that there is no integration of contents between the hemispheres here, and assuming that the moving object is in fact seen *as* the source of the touch, one hemisphere subject, Righty, will be conscious at first of being touched by a thing they can see, and then of being touched by the same thing, which they cannot see. This much is not so strange; after all, sometimes we see and feel something, and then feel it continuing to touch us while out of sight. What is strange is that even when Righty has no visual experience of the object, they still feel it *as* something also seen: their tactile experience co-presents a visual experience of that object, and does so quite informatively—the character of Righty's tactile experience reflects that of the visual one just as much when Righty is not undergoing **(p.216)** that visual experience as when they are. Yet at the same time the visual experiences they do have (of items in the left visual field) do not co-present any further visual experiences: they carry no information about what is appearing in the right visual field, even though what is appearing there is precisely the object which is "felt as also seen."

Neither the perspective of the whole split-brain subject (with massive indeterminacy between two very complex and definite states) nor that of the hemisphere subjects (with a highly informative co-presentation of something that is not actually experienced) is familiar to us or easy for us to imagine. They both differ from our everyday experience, in which we are neither split internally (unlike the split-brain patient) nor tightly linked to an external subject (unlike the hemisphere subjects). This strangeness, together with the anti-combination intuition, contributes to making the split-brain case so intensely puzzling.

6.3.5. The Truth in the Switch View

Finally, what about the switch view? This takes its cue from evidence showing that when given certain combinations of left-side and right-side stimulation, split-brain patients seem able to respond only to those on a particular side, with no response to the other stimulus even by the relevant hemisphere's motor organs, even though there would be such a response were the other stimulus presented by itself (Levy 1977, 1990; cf. Teng and Sperry 1973). Following Levy, Bayne (2008, 2010) takes this to support the view that consciousness switches between largely independent streams of processing in the two hemispheres, so that the split-brain patient is first conscious of things on one side, then of things on the other, and so on—but never has any disunified experiences at any single moment. While this view can account for many of the data, it has the implausible implication that much of the content of a single subject's consciousness switches

discontinuously several times a second; a version of the view which avoided or softened this implication would be preferable.

Fortunately, functionalist combinationism allows for such a version of the switch view. Since it allows for richer and poorer forms of conscious unity, it can make good sense of the experiences based in one hemisphere becoming, for some period of time, more closely unified with nonhemispheric functions and as a result having better access to various cognitive and behavioral processes, such as memory, report, and motor control. A moment later, that dominance might shift to experiences in the other hemisphere, with the first now being relatively isolated and inert. But we do not need this to be a sharp transition from consciousness to unconsciousness, or from unity to disunity; it can be a diminution **(p.217)** along a continuum, in the degree of phenomenal binding between different sets of experiences.

From the perspective of the whole, the switch might be experienced as one set of contents becoming more determinately lined up with the nonhemispheric contents (which are always highly determinate), “coming into sharper focus,” so to speak, while the other set fades and blurs a little. In everyday contexts this makes very little difference, because so much of the content in both hemispheres is very similar (with no pesky experimenters trying to segregate inputs). In experimental contexts, there is a significant difference in the experiences before and after a given switch—though even then, there will not be a corresponding experience of difference, since split-brain patients appear entirely oblivious to their own condition until confronted with evidence. This seems odd at first, until we remember that making judgments about change—*noticing* that experiences have changed—is itself one of the cognitive functions which the hemispheres seem to carry out individually, and which in this case neither can carry out by itself.

What makes the split-brain patient perplexing is not simply that it is hard to say whether there are two of something there or just one—this sort of difficulty is common and banal. We might look up at the sky to count the clouds and not be sure whether two partly connected masses of vapor should be called “one cloud” or “two clouds,” and then shrug and get on with our day; both descriptions capture something, and we are under no pressure to choose one and reject the other once and for all. What is deeply perplexing about the split-brain patient is that the things we do not know how to count are conscious subjects, and the only way to get the same equanimity about different counts of subjects as we do about different counts of clouds is to adopt combinationism.

6.4. Composite Subjectivity in the Typical Human Brain

What about those of us whose brains have not been split or otherwise tampered with? If “being conscious” is an intrinsic property, and if hemisphere subjects are conscious when they operate in isolation (as the survival of

hemispherectomy patients, who lose one hemisphere, suggests), then it seems to follow that we are each three overlapping subjects: two hemisphere subjects and a whole brain. Indeed, if constitutive panpsychism is true, then we are many more subjects than that. But let us start with the relationship between the hemisphere subjects: exploring that will cast light on how we might be even more radically composite.

Moreover, in exploring these possibilities we will discover a novel response to certain arguments advanced by Daniel Dennett (1991), which aim to show that there is no objective fact about what someone is conscious of, because there is **(p.218)** no single perfectly integrated stream of processing in the brain: functionalist combinationism allows us to accept much of the substance of Dennett's proposal (what he calls the "multiple drafts" theory) while remaining realists about consciousness.

6.4.1. Phenomenal Binding between Hemisphere Subjects

To understand the relation between hemisphere subjects, consider again the case where the words "key" and "ring" are flashed to the left and right visual fields. A split-brain patient has a visual experience of "key," which they understand as being about keys, and a visual experience of "ring," which they understand as being about rings; what the corpus callosum accomplishes for a normal subject is that these two visual experiences form a single experience of "key ring," which they understand as being about keyrings. To see what this means for the hemisphere subjects, we must crack open this complex experience and analyze how its parts are related.

According to functionalist combinationism, the normal brain's representational unity among the experiences of "key" and "ring" is a matter of informative co-presentation: each experience referring to the other's content as something continuous with its own, but not given directly through it. So the visual experience of the word "key" presents that word as the left-hand part of something. Rather than presenting the other, right-hand part of that thing, this experience points outside itself to (co-presents) the experience of "ring," which correlatively presents that word as the right-hand side of something, and points back to the first experience as what presents the other, left-hand part of it. And the associated awareness of the words' meanings shows the same structure, though here the relationship between the two contents is harder to label than mere spatial proximity: roughly it is that the rings are rings *for* keys, or rings whose function is somehow key-related.[13] Thus the experience of seeing "key" as being about keys is also an experience of seeing it as being about the key-relatedness of some other object, though it does not itself present that object; it merely refers outward to the experience of seeing "ring" as being about rings, which in turn is an experience of seeing "ring" as being about rings-for-some-

particular-purpose, and refers back to the first experience to specify which purpose.

(p.219) All this is experienced by the whole person. What about the hemisphere subjects? Supposing that the detection and interpretation of each word occurs in only one hemisphere, we might conjecture that they each experience only one half of the above set of experiences.[14] So Lefty, for instance, seeing only the right visual field, will have an experience of the word “ring,” presented as the right-hand side of a phrase involving another word, and will experience it as being not just about rings, but about rings that are meant for some purpose specified by the other word in the phrase. This is not so hard to imagine: we often see part of a phrase while knowing or assuming that there is more of it unseen, perhaps hidden behind a barrier, and we might in some cases even have a fairly definite idea of how the unknown meaning of the unseen word connects with the meaning of what we do see. What is perhaps a little stranger is that neither hemisphere subject does what we might do in such a situation: wonder what the other word is, and try to find out. Neither seems aware that the further experiences their own experience co-presents belong to another subject. Presumably the reason for this is that the signals they get from the other hemisphere not only give their experience the character of indicating something outside itself, but somehow satisfy, or preempt, any impulse to seek out that something else; because of the way each interacts with the other, both are oblivious to the fact that they are only half of the whole cognitive system. Although there are multiple subjects, their self-consciousness is of only one subject (cf. Uddin et al. 2005; Schechter 2018, 156–180). The explanation of this mutual obliviousness will have to wait for chapter 7’s discussion of self-consciousness.

Comparing the relation between hemisphere subjects with and without an intact corpus callosum shows us a particularly clear case of phenomenal binding. Of all the things that Lefty is conscious of, it is the experience of “ring” in particular that is phenomenally bound together with a particular experience of Righty’s, namely that of “key.” What allows for this specificity of binding is (to label it roughly) the integration of information which is accomplished by the corpus callosum: discovering the details of how this is accomplished requires detailed neuroscientific study. Without this integration of information, the elements of Lefty’s experience would have no specific tie to any specific elements of Righty’s, so that Righty would have no determinate awareness of what Lefty is experiencing, and vice versa.

(p.220) 6.4.2. The Representational Structure of Human Experience

Phenomenal binding allows for the construction of complex content from simpler content, and so in principle could construct all the complex contents of human consciousness out of very simple components. The particular example considered here, of course, involves binding fairly complex individual contents

(like the concept of a key, which depends on the idea of a lock, etc.) experienced by extremely complex subjects (large sections of the human brain) and so is not such a direct insight into the original construction of complex contents out of simpler ones. What it illustrates, however, is that contents can be combined in such a way that the result is more complex than either component. Other examples make the point more clearly: two component subjects, undergoing mutually referring experiences, one of a circle-as-left-hand-part-of-a-shape and one of a circle-as-right-hand-part-of-a-shape, would form a composite subject with an experience of an “infinity sign” type of shape. This might be true even if neither component subject by itself had the capacity to experience a shape like that but was limited to just circles.

Woodward (2015, 249–250) worries that there are still too many primitive types of content involved in human representation: since they are primitive they cannot be built up by phenomenal binding, and there are too many of them to be plausibly taken as metaphysically fundamental. One response for the functionalist combinationist would be to dispute the claim that this or that particular content-type was primitive, but in fact it is not necessary to engage in this kind of semantic trench warfare. The combinationist need not regard all those contents which are primitive for human development and reflection as primitive in a metaphysical sense: the fact that we cannot analyze a concept into further components does not by itself guarantee that it is not built up out of further components, for those components may be “hardwired” together within the human brain, so that they are robustly confused: only ever represented by us in combination, and never alone.[15]

Thinking about this sort of hardwiring is also relevant to a form of conscious structure mentioned earlier: the division of human experience into five sensory modalities.[16] What explains this division? Chapter 5’s account of blending serves as a first step, since it explains how many different sorts of qualities can be experienced. But it is not a full answer, because some modalities feature more than one type of quality (e.g., vision presents both shapes and colors) and because some types of quality can appear in more than one modality (e.g., shape is detected by **(p.221)** both vision and touch). We see the contours of a fuller answer when we observe that the boundaries between modalities correspond to constraints on what can and cannot be phenomenally bound. For example, we always experience colors as colors of some shape, and never as colors of some sound.[17] We always experience loudness as the loudness of an event with a pitch, and always experience pitch as the pitch of an event with some degree of loudness. In short, there are constraints on the ways that particular experiential contents can be, or must be, phenomenally bound to particular others. These constraints constitute a division among modalities; for each modality there is a set of types of content that, whenever they occur, can or must be bound to each other but not to contents of different modalities.[18]

The distinctively combinationist spin on the above claims is simply that as well as the whole subject experiencing colored shapes there may also be component subjects whose experience is best characterized as something like “redness-as-the-color-of-something-shape-like.”[19] Their relationship to other such subjects is thus somewhat like the relationship of one hemisphere subject to the other, but more fixed, since their experiences always co-present particular other subjects’ experiences. This relationship is also somewhat like that between two people in a rigidly organized social group, each working on a particular part of a task while maintaining a sketchy background awareness of the larger task, including the portions the other is working on.

Why are our experiences bound in particular ways? Providing a detailed answer to that question is an ongoing area of research in psychology, on the toes of which I have no wish to step. But the basic type of explanation is clear and unmysterious: the causal architecture of the brain puts some neural structures in communication with others, so that information is integrated in certain ways rather than others.

(p.222) 6.4.3. Functionalist Combinationism and the Multiple Drafts Model

The constitution of composite experience by many somewhat independent component subjects gives rise to a certain puzzling possibility: that the experiences of the brain’s many parts might occur “out of sync” with the other experiences they co-present. For example, a particular object’s color might be experienced by one part before the part which detects and experiences its shape has finished its information processing. Not only is this a possibility, but it seems likely to be an actuality. After all, from an evolutionary perspective, a brain which does not force all subsystems to wait for the slowest one will react faster, which might make the difference between life and death. Yet this makes it puzzling that the composite subject seems to experience a single consistent timeline, in which all the features of the object come together at once.

The “smearing-out” of content determinations across various subsystems comes out very clearly in certain perceptual illusions, like the “cutaneous rabbit” (Geldard and Sherrick 1972; Geldard 1982) and “color phi” (Wertheimer 1912; Kolers and von Grünau 1976), where two stimuli appearing at different locations in rapid succession give rise to an experience of a single object moving from one location to another. In the “cutaneous rabbit,” for example, a tap at the wrist followed by a tap at the elbow gives rise to an experience as of multiple taps “hopping” along the forearm, even though the forearm was not touched.

What is puzzling about these cases is not simply that there is an illusion of a stimulus which was never present: that can be explained simply as brain systems making rapid, fallible “guesses” about the likely causes of inputs. What is puzzling is that the illusory experience seems to take place *before* the experience of the second actual stimulus, even though the latter is necessary for

the illusion to take place at all. Thus it seems that the brain must be detecting the second stimulus, forming a “best guess” model of the cause, and generating an experience of the intermediate, illusory taps—and doing all of this *before* there is any conscious experience of the second tap. While it is not inherently implausible that there is some temporal gap between receiving a stimulus and undergoing an experience, the length of the gap in this particular case seems implausibly large—in particular, it would exceed the time that seems to be required for the construction of experience in other, simpler cases (see, e.g., Efron 1967; Pockett 2002, 2003). This might push us toward an alternative picture: the experience of the second stimulus occurs quickly, but as soon as some brain mechanism computes that the most likely cause of two such rapid stimuli is a single moving object, there is a swift revision of memory, so that the subject falsely remembers having experienced the intermediate stages of the movement.

(p.223) Dennett (1991, 115–117) labels these two models the “Stalinesque” and “Orwellian,” since the first carefully constructs a “show trial” (a false perception) and the second revises “historical records” (a false memory). Rather than supporting one over the other, however, he holds them up as a dilemma that a good theory should save us from. If the brain is a system designed by evolution to do things as quickly and efficiently as possible, why would it “wait” to experience the second stimulus, as the Stalinesque model posits? And equally, why would it bother making rapid, undetected, revisions to its memory records, as the Orwellian model posits? In fact, he thinks, once we reject the misleading image of a central location in the brain where conscious experience is “played” for an observer (the conscious subject), we see that neither the Stalinesque nor the Orwellian model is right. What happens, instead, is that multiple subsystems of the brain are processing different aspects of the stimulus in parallel, not waiting for one another but updating each other with their results as soon as those results are produced, and revising their results when those of other subsystems turn out to conflict. In this ongoing stream of multiple, changing, partial-content determinations, there is a rapid determination of certain features of the second stimulus, and then soon after a determination that the most likely cause was a single moving object, leading to a representation of intermediate stimuli that arises after the representation of the second actual stimulus, but is represented as occurring before it. There exist simultaneously two “drafts” of history: one featuring two stimuli, the other featuring three.

Most radically, Dennett proposes that there is no objective fact as to what the subject is conscious of at any particular moment: all of these multiple, sometimes conflicting content determinations carried out by different subsystems are equally good candidates for being called “what is experienced.” Their interactions—checking for consistency and revising what is found inconsistent—are not “prepublication” checks to ensure that a consistent total content is “presented to consciousness,” because there is no further, distinct

stage where conscious experience is “played”: these messy changing parallel content determinations, these “multiple drafts,” are all there is. Our impression of a single consistent timeline of consciousness arises because if someone else *asks* us “What did you experience?,” or if we ourselves ask that question for some purpose, our various subsystems will settle on mutually consistent contents *to present as an answer*, i.e., to guide the behavior that answers the question, to be reported, remembered, or otherwise made use of. As Dennett (1991, 138) puts it, any “probe” we give the subject will elicit a single consistent timeline of experienced events, but “there are no fixed facts about the stream of consciousness independent of particular probes.” Thus his theory has been viewed, I think rightly, as in effect a sort of “eliminative **(p.224)** materialism,” a denial of the reality of conscious experience as something existing independently of any particular probe.

Panpsychist combinationism is in some ways radically at odds with the multiple drafts theory: one denies any objective fact about consciousness, the other treats consciousness as a fundamental feature of reality. And functionalist combinationism, though not committed to the fundamentality of consciousness, does presuppose that there are objective facts about it. But in another way the theories are very similar: both identify the seemingly indivisible unity of the mind as masking an underlying multiplicity of somewhat independent parallel processes. In particular, they can all avoid the awkward choice between Stalinesque and Orwellian models of illusions like the cutaneous rabbit. The multiple drafts theory says that there is no objective fact about whether the illusory intermediate touches are inserted into the story “before consciousness” (Stalinesque) or “after consciousness” (Orwellian), because there is no objective fact about what is and isn’t conscious. Functionalist combinationism says that it is all conscious, equally, but forms a composite stream of consciousness whose content is at certain points significantly indeterminate. Let me explain a little more fully what this means.

As sensory input is processed by different subsystems, each one is conscious and their experiences reflect the information they are processing. As soon as one subsystem, for instance, determines the result “red” for the color seen at a certain point, its experience of visual red is part of the composite subject’s conscious stream. But it is always going to be easier for each subsystem to reach a conclusion than for them all to establish a consistent, integrated timeline of what has been perceived; integration lags behind local content determination, if only because integration will be delayed if some of the subsystems work much faster than others. This delay of integration means a delay between the redness experience being part of the composite’s stream of consciousness, and its being fixed to a particular point in that stream—characterizing a particular object, with particular other features, that looked a certain way a moment before, in a particular modality, in sensation as opposed to emotion, etc. The process of its acquiring such a particular phenomenal location is in effect the process of its

becoming conscious, because when it is entirely unlocalized it cannot be attended, noticed, or reflected upon: at most it blends with the inarticulate background of consciousness.

The process of making particular experiences properly, usefully conscious is a process of aligning them with other experiences so as to coalesce from the blur, and this process will happen by degrees over a period of hundreds of milliseconds. At certain points there will be no objective fact about whether a given “draft” is integrated enough to be called “an experience of the whole” in the ordinary, everyday sense. And if two subsystems have conflicting contents (e.g., “taps all along **(p.225)** the forearm” and “taps only at the wrist and elbow”) then, until one or the other is revised, the total stream will include both, representing that it is indeterminate whether there are or are not taps along the forearm.

This proposal does make a phenomenological prediction: it predicts that over short periods of time, the best description of what we are seeing (or hearing, or feeling, etc.) may simply be “something.” When surprising things happen quickly, the best way to characterize how things seem to us is first as “something is happening,” and swiftly thereafter some more determinate kind of content. This seems to me a correct phenomenological prediction (with all the caveats about the unreliability of such declarations); when I am startled, for instance, it often feels as though I am not aware of what, exactly, I felt, until a half-second or so after I am aware that I am feeling something (cf. Gross and Flombaum 2017). To put it another way, the arising of new experiences often feels more like a coalescing, a rapid progression from blurriness to determinacy, than like the appearance of fully formed images out of nothing.

Of course, it is hard to be confident about any of this, for a reason that both the multiple drafts theory and functionalist combinationism predict: the more we pay attention to a particular perception, the more swiftly and fixedly it becomes integrated and consistent. The period of indeterminacy is hard to detect because it vanishes as soon as we attend to it. My aim is not to commit combinationists to anything resembling the multiple drafts theory, but merely to show functionalist combinationism’s flexibility, and its surprising affinity for accounts which are usually regarded as antirealist about consciousness, and to that extent sharply opposed to some of the views, like CRP, which would motivate combinationism.

6.5. Conclusions

My four case studies are only a small selection from the range of cases that I believe would benefit from, or call out for, a combinationist treatment. I cannot here attempt a full analysis of, for example, the possible consciousness of ant colonies (though I suspect they have something in common with the Nation-Brain), or the relationship between consciousness in an octopus’s central brain and consciousness in its arms (though I suspect that case is analogous to the

split-brain phenomenon in many respects). But I hope that my discussion of these four examples has illustrated how functionalist combinationism can cover and illuminate a wide variety of compositional relationships between intelligent subjects. **(p.226)**

Notes:

(1) If functionalist-panpsychist combinationism is true, then there is no contradiction in a nonconscious, “zombified” version of the Nation-Brain. But even then, given that the psychophysical laws of the actual world make human intelligent subjects, those laws will likewise make some versions of the Nation-Brain intelligent subjects.

(2) A Turing-machine table is a sort of very abstract program defined by a set of inputs, a set of outputs, and a set of internal states, along with exhaustive rules for what outputs should be generated, and what state moved to, when a given input is received while in a given state.

(3) Strictly, of course, the system also includes various radio devices, but to pose our question most sharply we should keep their information-processing role to a minimum: the citizens must choose the right buttons to press, not allow any automatic programming in the radios to do it for them.

(4) It will say yes when we have given the avatar a radio device of its own, but this is no more relevant to the metaphysics of its consciousness than giving a normal human some brain matter to feel with their fingers.

(5) I will focus on nations for simplicity, but any large group whose members interact frequently would do as well (corporations, provinces, cities, or even arbitrary chunks like “everyone in Southern England but not in London”).

(6) Of course, each individual neuron’s interests are inextricably bound up with the functioning of the whole; they do not have the option of living independently and cannot procure food, oxygen, or anything else beneficial to them except by procuring them for the whole organism. This just illustrates the point being made: human nations do not act like human brains, because humans do not act like neurons, because they *are* capable of (more or less) independent existence, or at least of ditching one set of cooperators for another. The establishment of such identity of interests at the *genetic* level seems to play a crucial role in the establishment of large-scale biological systems, such as multicellular organisms, eusocial insect colonies, and even the eukaryotic cell (see Godfrey-Smith 2012).

(7) The step from joint action to joint intention is motivated by the fact that the joint/nonjoint distinction seems to apply only to intentional actions (no similar difference seems to exist between, say, falling together and falling individually).

(8) The views of Bratman and Gilbert (1990, 2000, 2002, 2009) receive the most discussion here, though I treat their accounts as accurate descriptions of two slightly different phenomena rather than rival accounts of a single phenomenon. Other approaches include those of Searle (1990), Tuomela (2007), and Tollefsen (2014).

(9) It is not quite true that this is independent of A's intentions: A must still intend to follow B's instructions, and when B says "do X," A must form an intention to do X. The point is that A need not intend, or even know about, the whole structure and ultimate goal of B's plan.

(10) In saying that this would make the nation "behave intelligently" I mean simply that we could interpret it as doing things in order to accomplish a single set of goals, according to a single set of beliefs. Its behavior might nevertheless be ruinously stupid, either because the dictator was an idiot or because, lacking active contributions from any of the other individuals, they could not monitor national events or formulate national plans in adequate detail.

(11) It is rather as though we had two identical twins (the hemisphere subjects) pretending to be a single mind directly controlling two bodies, but in fact using a combination of discreet mobile phones and a well-developed ability to read the other's body language. Cutting the corpus callosum is like taking away their phones, and stimulus-segregating experiments are like putting a screen between them to stop them from watching each other's body language: neither by itself will impair their ability to accurately report the other's experiences, but doing both will. (The analogy is poor, of course, in that the twins must be imagined as knowingly seeking to deceive, while hemisphere subjects seem oblivious to their multiplicity—the reasons for this failure of self-knowledge will need to wait for chapter 7.)

(12) It should be recognized that there are various cases in which split-brain patients have displayed apparently unified thought or action, but later been found to have accomplished this integration by indirect means, such as by having one hemisphere direct the eyes to a certain target based on information it possesses, and the other infer that information from observing where its eyes are pointed. But there are also many displays of unity that remain even after care is taken to prevent such cross-cuing. See Schechter (2015) for fuller discussion.

(13) What I have said here treats "keyring" as a compositional term synonymous with "ring for keys"; in practice our experience of seeing keyrings and hearing them called "keyrings" gives us to some extent a noncompositional grasp on what "keyring" means. To avoid unnecessarily complicating my analysis I have ignored this fact: we may suppose our subject has not previously encountered keyrings.

(14) In most people, the left hemisphere takes over most language-processing, though the right seems to be capable of a degree of comprehension even without production. Thus the left-hemisphere subject probably does experience the whole meaning “keyring,” perhaps without seeing the whole image of the word. For simplicity of presentation I have neglected this asymmetry.

(15) Strictly speaking, the relation in question would be something more than confusion: confusion means that two elements can be thought of only together, not apart; what I am suggesting is that elements can be thought of only together *in a particular relation*, not apart or in some other relation.

(16) Whether there are really five modalities, or some larger number, is disputed; for discussion see Grice 1962; MacPherson 2011a and b; Matthen 2015. Nothing I say here depends on the answer to this question.

(17) Creatures which did not work like this are possible, and perhaps synesthetes are an actual example (or perhaps they are not, and simply [for instance] *see* a color whenever they *hear* a sound). That would simply show that the configuration of modalities in those other subjects was different to the human norm.

(18) Perhaps we can account for the difference between visual shape and tactile shape by observing that some experiences of shape come to us bound with experiences of pressure, texture, etc., but not with experiences of color, while others come to us bound in the opposite way.

(19) This has much in common with Zeki’s (2007) suggestion that multiple brain areas independently generate “microconsciousnesses” for different perceived qualities, which are only secondarily unified into “macroconsciousness.” Note, however, that Zeki’s view is that upon being unified together, the microconsciousnesses cease to exist in their own right (584), while functionalist combinationism says that they continue to exist as components of the macroconsciousness.