Rationality and Synchronic Identity

Brian Hedden

Abstract

Many requirements of rationality rely for their application on facts about identity at a time. Take the requirement not to have contradictory beliefs. It is irrational if a single agent believes P and believes $\neg P$, but it is not irrational if one agent believes P and another believes $\neg P$. There are puzzle cases, however, in which it is unclear whether we have a single agent, or instead two or more. I consider and reject possible criteria of identity at a time before proposing a pluralist alternative on which there are vastly more agents than we might have thought. This pluralist thesis is analogous to mereological universalism, on which there are all sorts of strange disconnected objects we don't usually take note of. I conclude by giving a pragmatic account of which of these rational agents it makes sense to attend to by appealing the purposes we have in employing rationally evaluative language.

1 Introduction

If a single person believes both P and $\neg P$, then that person is irrational. But if one person believes P while another believes $\neg P$, then neither need be irrational. In this way, the norm against having contradictory beliefs relies on a notion of personal identity at a time, or synchronic identity, to distinguish between cases where we have a single person at a given time from ones where we have two or more.

Many other requirements of rationality are like this. The requirement that one's beliefs be proportioned to one's evidence presupposes that there is a fact of the matter whether a given

set of beliefs and a given body of evidence belong to the same person. The requirement that one's preferences be transitive presupposes that there is a fact of the matter when distinct preferences belong to the same person. And so on.

But just as there are puzzle cases for personal identity over time, so there are cases where the facts about identity at a time are far from clear. How should epistemology, and more generally the theory of rationality, deal with these puzzle cases?

In setting up the issue, I used the term 'person.' But there are other closely related terms like 'agent,' 'self,' 'mind,' and 'subject of experience.' I am not sure about the exact relations between them, and so I use the more neutral term 'rational subject' to refer to the kind of entity that is subject to rational requirements and is the proper target of rational evaluation. Our question, then, is how to individuate rational subjects.

2 Disunified Minds

Let's start with some examples. These have been discussed by philosophers of mind, especially in the debate over the 'unity of consciousness.' But they have mostly been ignored by epistemologists and decision theorists.

2.1 Multiple Personality

Real cases of multiple personality are varied, and their description is vexed. To avoid these complications, consider a fictional case from Olson (2003, 330):

> Imagine, then, a human being with two personalities, one of which is "out" or active on even days, and the other of which is active on odd days. Each day at midnight, like clockwork, one personality goes dormant and the other takes over. Call the being who thinks and acts on even days *Even* and the one who thinks and acts on odd days *Odd*...Suppose that Even cannot recall, on even days, anything that Odd thinks or does on odd days, and that Odd likewise cannot recall on odd days anything that Even does or thinks on even days...Even (on even days) is cheerful and gregarious, while Odd (on odd days) is sullen and withdrawn. Even enjoys smoking on even days; Odd detests it on odd days. We may want to add

> that the even-day and the odd-day thoughts could no more *become* integrated, through psychotherapy or the like, than your thoughts could become integrated with mine.

In the case of Even and Odd, do we have a single rational subject, or two (or three, if both Even and Odd and the whole human being qualify)? Is there any irrationality in virtue of the lack of coherence between the mental states associated with Even and those associated with Odd?

2.2 Split-Brain Patients

In split-brain cases, the corpus collosum is severed, largely separating the two hemispheres of the brain. Under experimental conditions, split-brain patients exhibit behavior indicative of mental disunity. In a standard sort of experiment described in Nagel (1971), one word (e.g., 'toothbrush') is flashed to the patient's left visual field (processed by the right hemisphere) while another word (e.g., 'pencil') is flashed to the right visual field (processed by the left hemisphere). When the patient is told to retrieve the relevant object, the left hand (controlled by the right hemisphere) will retrieve the toothbrush, while the right hand (controlled by the left hemisphere) will retrieve the pencil. This and related experiments suggest some mental disunity in the patient, with mental states corresponding to one hemisphere unable to causally interact in the usual way with mental states corresponding to the other.

This disunity shows up in daily life as well. Some subjects report 'intermanual conflict,' in which the two hands seem to be working toward conflicting goals, and some have 'ongoing difficulty successfully and efficiently carrying out activities that involve complex or embedded sequencing of actions' and experience that making 'decisions on the basis of multiple and potentially competing considerations can be difficult' (Schechter 2012, 578).

But it's not all disunity. In most contexts, split-brain patients behave in unified ways; they are able to 'play the piano, button their shirts, swim, and perform well in other activities requiring bilateral coordination' (Nagel 1971, 401). Thus it is not clear whether a split-brain

patient is a single rational subject with a striking degree of mental disunity, or whether instead such cases involve two rational subjects, one corresponding to each hemisphere.

2.3 Craniopagus Conjoined Twins

Krista and Tatiana Hogan are craniopagus twins, meaning that they are conjoined at the head. They have a unique neural anatomy, with their brains connected by a structure that their neurosurgeon, Douglas Cochrane, calls a 'thalamic bridge.' It connects the thalamus of Krista with the thalamus of Tatiana. The thalamus plays a role in the relaying of sensory and motor information to the celebral cortex and is implicated in consciousness. This suggests that the twins might share some sensations and other phenomenal experiences.

A *New York Times Magazine* article (Dominus 2011), provocatively titled 'Could Conjoined Twins Share a Mind?,' reports anecdotal evidence for this hypothesis. One twin has been able to report where the other is being tickled, or what object is being shown to the other, despite looking in another direction. An early video showed one twin starting to cry just as the other was pricked for a blood test. And Cochrane performed a (unpublished) study in which Krista's occipital lobe, implicated in visual processing, emitted an electric response when a strobe light was flashed in Tatiana's eyes, and vice versa.

The twins also seem to have an ambiguous self-conception, which is unusual given their age (four years old in 2011). Dominus reports that they rarely use 'we,' and sometimes use 'I,' to refer to the two of them together, as when Krista said 'I have two pieces of paper' in a case where each twin had one piece.

2.4 The Octopus

Octopuses are intelligent. They have large brains, can learn to navigate mazes and unscrew jars, and seem to recognize particular humans and adjust behavior accordingly. They may also engage in 'play' and various forms of social behavior (Godfrey-Smith 2016).

But their nervous systems are more distributed than chordate ones. The octopus's central

brain contains only around half as many neurons as the arms (ibid, 67), and each arm's nervous system is connected to the brain by 'a relatively small number of nerve fibers' (Sumbre et al 2001). This, as well as behavioral studies, have suggested to some that the arms are reasonably independent, with Hanlon and Messenger (1996, 15) describing the arms as 'curiously divorced' from the brain and stating that 'many of the arms' actions are performed without reference to the brain.'[1]

Based on more recent research, Godfrey-Smith (2016, 68) suggests that octopuses in fact have an odd mix of disunity and unity, of 'localized and top-down control':

> A 2011 paper by Tamar Gutnick, Ruth Byrne, and Michael Kuba, along with [Binyamin] Hochner, described a very clever experiment. They asked whether an octopus could learn to guide a single arm along a mazelike path to a specific place in order to obtain food. The task was set up in such a way that the arm's own chemical sensors would not suffice to guide it to the food; the arm would have to leave the water at one point to reach the target location. But the maze walls were transparent, so the target location could be seen. The octopus would have to guide an arm through the maze with its eyes.
>
> It took a long while for the octopuses to learn to do this, but in the end, nearly all the octopuses that were tested succeeded. The eyes *can* guide the arms. At the same time, the paper also noted that when octopuses are doing well with this task, the arm that is finding the food appears to do its own local exploration as it goes, crawling and feeling around. So it seems that two forms of control are operating in tandem: there is central control of the arm's overall path, via the eyes, combined with a fine-tuning of the search by the arm itself.

Should we then think of the octopus as a single rational subject, or a collection of distinct but largely cooperative rational subjects, or something in between?

2.5 Fragmentation

Mental disunity is not confined to rare surgical procedures, mental illness, and highly alien nervous systems. There is a long history of thinking of ordinary people as somewhat disunified, with minds composed of interacting subunits. We have Plato's city-soul analogy, Freud's id,

[1] Sumbre et al (2001) show that some arm actions can be generated even when the nerves connecting the arms to the brain have been severed.

ego, and super-ego, Fodor's (1983) modular conception of the mind, Kahneman and Tversky's System 1 and System 2 (see Kahneman 2011), and Ainslie's (1992) picoeconomics.

Here I'll discuss fragmentation, in which an entity's mental life is modeled not by a single set of attitudes but by a set of fragments, each containing a set of mental states, which are operative in different situations (Davidson 1982, Lewis 1982, Stalnaker 1984). Fragmentation has been proposed as a way of modeling creatures which seem to have contradictory beliefs, failures of information access (Elga and Rayo ms), failures of self-knowledge (Greco 2014), and knowlege-that without knowledge-how (Elga and Rayo ms).[2]

In a now-famous example, Lewis describes finding himself with contradictory beliefs about the layout of Princeton: (i) that Nassau Street runs roughly east-west, (ii) that the railroad runs roughly north-south, and (iii) that Nassau street and the railroad are roughly parallel. He argues, however, that his state of mind was not adequately modeled as a single set of inconsistent beliefs. Instead, his ' system of beliefs was broken into (overlapping) fragments. Different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once. The first and second sentences in the inconsistent triple belonged to—were true according to—different fragments; the third belonged to both' (436).

Why model Lewis's state of mind using a set of internally consistent bodies of belief rather than a single inconsistent one? Fragmentation, its advocates claim, makes better sense of Lewis's behavior. We expect that people will by and large perform the actions which would best satisfy their desires, were their beliefs true. And as charitable interpreters, we attribute to people the beliefs and desires that best rationalize their actions. But it is difficult to see what actions would best satisfy one's desires if an inconsistent set of propositions were true. It is difficult to make sense of what actions would best satisfy Lewis's desires if (*per impossibile*) Princeton had an east-west street and a north-south railroad running parallel to each other. And so we should be loath to attribute to him an unfragmented state of mind involving a single, inconsistent, belief state. Instead, we should attribute to him the above fragmented

[2]Elga and Rayo (p. 10) also propose fragmentation as a way of modeling split-brain patients, and it would of course be natural to use it to model cases of multiple personality.

state of mind, thereby making sense of the fact that for some purposes, he acts as if the street and railroad both run roughly east-west, while for others he acts as if they both run roughly north-south (cf. Egan 2008).

In the usual fragmentation story, the beliefs are all attributed to Lewis himself but indexed to different fragments. But we can also consider the possibility that each fragment corresponds to a distinct rational subject. Indeed, one tempting proposal, considered below, yields just this result.

3 Criteria of Rational Subjecthood

In each of the above puzzle cases, where multiple sets of mental states aren't integrated with each other, do we have a case involving one rational subject which violates rational requirements, or a case involving multiple rational subjects but no irrationality? What we need is a criterion of (synchronic) rational subjecthood, by which I mean a theory stating the conditions under which two mental states at a given time count as belonging to the same rational subject. Note that a criterion of rational subjecthood, so understood, is not a theory about the metaphysical nature of rational subjects, which would involve some commitment to facts about rational subjects' essences, persistence conditions, and the like.

Below I consider three natural criteria of rational subjecthood: a bodily criterion, a psychological criterion, and a criterion that appeals to entity's own self-conceptions. I argue that each is problematic before proposing a pluralist alternative.

3.1 The Organismic Criterion

One natural criterion of rational subjecthood gives pride of place to bodies. It says that two mental states (had at the same time) belong to the same rational subject if and only if they are realised[3] by (parts of) the same body. However, some authors (e.g., Olson 1997) hold

[3]The concept of realisation is a new one, which Kim (2010) traces to Putnam (1960). While there is no consensus characterisation of realisation, on one popular view it has to do with the determinable-determinate

that we have no good conception of what bodies are, and so it would be better to speak of organisms. Following their lead yields a criterion that states that two mental states belong to the same rational subject if and only if they are realised by (parts of) the same organism. (As noted above, a criterion of rational subjecthood is not a theory about the metaphysical nature of rational subjects, and so this criterion does not entail that rational subjects *are* organisms, as opposed to, say, some other entities coincident with organisms.)

Thus in the various puzzle cases above (except perhaps that of the Hogan twins; see below), we have a single rational subject with a disunified mind, and this means that (at least where the mental disunity involves conflicting attitudes) the subject is irrational. This doesn't mean that the subject is blameworthy for its irrationality. Insofar as the cognitive architecture of the split-brain patient, for instance, prevents her from noticing and resolving conflicts between thoughts associated with different hemispheres, this would constitute an excusing factor. But the patient would still count as violating the rational ideal.

Why endorse the organismic criterion? First, it has an affinity with animalism about personal identity (van Inwagen 1990, Olson 1997, Snowdon 2014)[4], according to which persons are biological organisms. Second, organisms must act in unified ways in order to be successful. Korsgaard (1989, 115) defends something like the organismic (or bodily) criterion on such pragmatic grounds:

> Pragmatic necessity can be overwhelming. When a group of human beings occupy the same territory, for instance, they have an imperative need to form a unified state. And when a group of psychological functions occupy the same human body, they have an even more imperative need to become a unified person. This is why

relation (Yablo 1992). A mental property is realised by a physical property just in case the former is a determinable of the latter. Note that here realisation is a relation between properties, whereas I have talked of mental properties (or states) being realised by physical entities. This should be regarded as loose talk; really it is the physical properties of those entities that realise the mental states.

[4]Why just an affinity, rather than being equivalent to animalism? First, some animalists, including Olson, do not claim that necessarily, all persons are animals or biological organisms. Rather, they claim only that *we* are human animals, leaving open the possibility of other persons who are not animals. Second, animalism is a thesis about the metaphysical nature of (human) persons. It says that such persons *are* animals. This means, *inter alia*, that they have the persistence conditions of animals. By contrast, the organismic criterion of rational subject is not a theory about the nature of rational subjects, but rather a theory of when two mental states belong to a single rational subject, as opposed to distinct ones.

> the human body must be conceived as a unified agent. As things stand, it is the basic kind of agent.

Third, the organismic criterion is particularly attractive in the case of fragmentation. Fragmentation is supposed to be a way of modeling certain sorts of irrationality. But if the mental states indexed to different fragments count as belonging to different rational subjects, we get the unintended result that there is in fact no irrationality. The organismic criterion, by contrast, has the intuitive result that the mental states indexed to different fragments belong to the same rational subject.

Despite these attractions, I find the organismic criterion untenable. First, it gets some cases wrong. It is natural to treat the case of the Hogan twins as one where there are two different rational subjects, but it is not clear that they count as distinct organisms (or as having distinct bodies) (cf. Campbell and McMahan 2010). They are physically connected and cannot be separated without (a high probability of) death or serious impairment. They do, I presume, have different DNA, but we can easily imagine a case of identical craniopagus conjoined twins with the same DNA. (Note that this case also puts pressure on a brain-based criterion of rational subjecthood.) Even if we can come up with a way of individuating organisms such that the Hogan twins count as distinct organisms (see Liao 2006, Olson 2014), imaginary cases put further pressure on the organismic criterion. van Inwagen (1990)[5] imagines a creature, Cerberus, which has two heads, each with its own brain, where these brains are identical in physical structure and control ratiocination and voluntary motion. It is natural to treat the Cerberus case as one where there are two rational subjects, *contra* the organismic criterion of rational subjecthood.

Second, the organismic criterion faces an explanatory challenge. Why is are organisms (or bodies) so important for epistemology and the theory of rationality? (The concept of a rational subject, recall, is connected to epistemological concerns; as introduced above, the term 'rational subject' refers to the kind of entity that is subject to rational evaluation and is the proper target of rational evaluation.) After all, epistemology seems to be about mental

[5]See Bayne (2010, 271-5) for further discussion of the Cerberus case.

states. We can even imagine rational requirements applying to disembodied souls. The above quote from Korsgaard goes part of the way toward meeting this challenge. An organism is the site of action and must act in a unified way in order to be successful. But this is not altogether adequate. Unity within an organism is not always necessary for organismic success. Octopuses are amazingly successful despite, or perhaps because of, their highly distributed nervous systems. There are also reported cases of multiple personality in which the patient is healthier while 'suffering' from multiple personality disorder than when cured (Wilkes 1981, 344). And Egan (2008) argues that fragmentation may be beneficial for creatures with certain sorts of perceptual systems. Similar comments on the merits of unity and disunity apply when we consider groups of organisms. A group of humans will sometimes better achieve its goals if its members are unanimous in their beliefs and desires, but in other cases it is better if each brings to the table different perspectives and different problem-solving strategies. Of course, a defender of the organismic criterion could respond by arguing that only whole organisms can have mental states. Then, there would be no metaphysically acceptable option except to attribute all the mental states realised by (parts of) a single organism to the same rational subject. However, I know of no compelling argument that only organisms can be the bearers of mental states.

3.2 The Psychological Criterion

More promising, to my mind, is a psychological criterion of rational subjecthood. Start with a version of the psychological criterion cashed out in terms of minds (as opposed to mental systems). The problem is that it is not clear what exactly a mind is supposed to be (assuming the falsity of Cartesian dualism). Even if we allow ourselves to run with the notion of a mind, we have no principled way of counting or individuating minds. Indeed, this is the conclusion that many draw from the cases discussed above, especially split-brain cases (Schechter 2015). For instance, Wilkes (1988, 161) writes that 'There is no theoretically coherent account of what minds are, and no reason to suppose that the term 'mind' refers to anything that it is

necessary to individuate, nor to anything that we can count.'

Might the so-called 'unity of consciousness' help? Bayne (2010) offers a powerful defense of the unity thesis, which says that necessarily, for any conscious subject of experience and any time, the simultaneous conscious states that the subject has at that time will be subsumed by a single conscious state, namely the subject's total conscious state. If the unity thesis is true, then in all these puzzle cases in which there is mental disunity of various kinds, there will still be a fact of the matter about how many unified 'phenomenal fields' there are.[6]

Bayne goes on to defend roughly the view that the self is a stream of phenomenal consciousness, which at any one time will be constituted by a single unified phenomenal field. Analogously, we can consider a psychological criterion of rational subjecthood on which rational subjects are individuated by phenomenal fields, with two mental states counting as belonging to the same rational subject if and only if they somehow belong to the same phenomenal field.

But this proposal is problematic. First, phenomenal consciousness isn't all the matters as far as epistemology is concerned. While some theorists (e.g., Smithies (ms)) do maintain that phenomenal consciousness plays a central role in nearly all aspects of epistemology, most do not. Rationality has much to do with relations among functional mental states like beliefs and desires that may not show up in phenomenal consciousness. Even phenomenal zombies, which lack any phenomenally conscious states whatsoever, can be evaluated for rationality. For this reason, the unity of consciousness is largely irrelevant to the question of how to individual rational subjects. Second, it is unclear how to make sense of the notion of a mental state 'belonging' to a phenomenal field. Of course, if all mental states were potentially phenomenally conscious, we might say that a belief belongs to a phenomenal field just in case

[6]Bayne (2010, 11) introduces the notion of a phenomenal field as follows: 'Phenomenal unity is often in the background in discussions of the 'stream' or 'field' of consciousness. The stream metaphor is perhaps most naturally associated with the flow of consciousness—its unity through time—whereas the field metaphor more accurate captures the structure of consciousness at a time. We can say that what it is for a pair of experiences to occur within a single phenomenal field *just is* for them to enjoy a conjoint phenomenality—for there to be something it is like for the subject in question not only to have both experiences but to have them *together*.'

it could arise in phenomenal consciousness in that field, but I doubt whether all mental states are potentially phenomenally conscious. And if they are not, then there is no clear way of pairing up mental states with phenomenal fields, as required by a unity of consciousness-based criterion of rational subjecthood.

More promising is a psychological criterion cashed out in terms of 'mental systems,' where a mental system is a system of causally interconnected mental states. This criterion says that two mental states belong to the same rational subject if and only if they are part of the same mental system (Shoemaker 1984).

This criterion is attractive but faces two major problems. First, it is difficult to see how the notion of a mental system can be made sufficiently precise to yield verdicts on the puzzle cases without also yielding problematic implications. Mental systems are supposed to be systems of causally interconnected mental states. But what sorts of causal connections between mental states matter? And how much interconnectedness is required? I'm not sure how to go about trying to answer these questions, and I worry that it may be impossible to do so without getting the either the result that the Hogan twins correspond to a single rational subject or the result that the different fragments in a Lewis-style fragmentation case correspond to distinct rational subjects, each of which is at least counterintuitive. If a mental system is one that is closed under 'direct' mental connections, then the Hogan twins may well count as sharing a single mental system, since a belief which we would intuitively class as Krista's may be directly connected to a sensation which is shared by them both, which in turn is directly connected to a belief which we would intuitively class as Tatiana's. On the other hand, if we require that mental systems consist of mental states that cohere with each other in some more robust way and which combine to produce behavior, then it is difficult to avoid the implication that each fragment in a Lewis-style fragmentation case counts as a distinct mental system.

Second, a mental systems-based version of the psychological criterion threatens to yield unacceptable results about what rational subjects are. I have noted that a criterion of rational

subjecthood is not a theory about the metaphysical nature of rational subjects. But the two are not entirely independent. Consider the following argument from Olson (2003, 336). Suppose two distinct mental systems are realised in the very same physical matter. This is likely true in cases of multiple personalities and fragmentation. There is no reason to believe that the different personalities, or different fragments, are realised in distinct parts of the brain.[7] Then, according to the mental systems-based criterion, we have two distinct rational subjects. But then the rational subjects cannot both *be* the organism, or the body, or the brain with which their mental systems are associated, for they are distinct from each other. And since the case is symmetric it would implausible to suppose that one of them is the brain while the other is the organism, say. It is then difficult to see what those rational subjects could actually be other than the mental systems themselves. But the problem is that mental systems don't themselves believe or desire anything. Mental systems may have beliefs and desires in the sense of having them as parts (just as a set of beliefs can have beliefs in the sense of having them as members) but not in the sense of being believers and desirers. We might bite the bullet and hold that the targets of rational evaluation are, first and foremost, systems of mental states rather than the bearers of those states. But I will not pursue this revisionary thesis here.

3.3 The Self-Conception Criterion

The third kind of criterion of rational subjecthood appeals to an entity's own way of conceiving of itself. (This could be classed as a version of a psychological criterion but is sufficiently different to warrant separate consideration.)

Here are two ways of fleshing out the idea that an entity's self-conception determines rational subjecthood. First, Carol Rovane (1998) argues that a person is an entity with

[7]Olson (2003, 342) notes if the brain is like a computer, then we could perhaps defend the claim that the different personalities, or different fragments, are realised by distint parts of the brain, just like two programs running on the same computer are realised in 'discrete sets of on-off switches.' But, he claims, we don't know that the brain works like that. We might add that even if brains do work like that, we can imagine other beings whose thought-producing organisms don't work that way.

a commitment to some unifying projects, which brings with it a commitment to rational unity—a commitment to having consistent beliefs, transitive preferences, and so on. While Rovane's is a theory of what it is for an entity to be a *person*, it could also be thought of as a criterion of rational subjecthood. Two mental states belong to the same rational subject if and only if they are realised by an entity (or parts thereof) with a commitment to unifying projects and to rational unity.

Second, we might appeal to an entity's self-referential thoughts (cf. Davis 1997), holding that two mental states belong to the same rational subject if and only if they are realised in an entity which contains token 'I' thoughts referring to that entity. So in a split-brain case, whether we have one rational subject or multiple depends on whether 'I' thoughts tokened in a single hemisphere (if there are any) refer to the patient as a whole or instead to some entity which contains or is partially composed of only that hemisphere.

I have two concerns about these approaches. The first worry is one of circularity. Suppose we have an entity, within which we find a token of a thought expressing commitment to rational unity within that entity. That is insufficient for rational subjecthood, unless that token thought belongs to the entity itself. To illustrate, suppose that we find in my brain a token of a thought expressing commitment to achieving rational unity within the mereological sum of you and I (assuming such an entity exists). This fact alone does not mean that all mental states realised by your brain and mine belong to a single rational subject (and not because group agents are impossible; Rovane is explicit that one of the upshots of her criterion is the possibility of genuine group agency). Of course, if it is our mereological sum itself that is committed to rational unity despite the corresponding token thought being realised in my brain, then that might well suffice for all the mental states realised by your brain and mine to belong to a single rational subject. But it is not clear whether we can determine to whom to attribute the commitment thought without first determining which entities count as rational subjects. Similarly, suppose some 'I' thought is tokened in one hemisphere of a split-brain patient. Offhand, it seems that whether the 'I' refers to the hemisphere or to the whole

patient depends on whether the hemisphere and/or the whole patient counts as a rational subject.

The circularity worry is not decisive. It may be that broader interpretative considerations can help settle to whom the commitment belongs and to whom the 'I' refers. For instance, whether to attribute the commitment thought tokened in my brain to me or to the mereological sum of you and I might be settled by how well that commitment thought meshes with other thoughts we could plausibly attribute to me, and with other thoughts (if there are any) that we could attribute to our mereological sum (Rovane 1998, 201). And it may be that whether an 'I' thought tokened in one hemisphere refers to the whole patient or to something containing just the one hemisphere can be partly settled by which choice of reference makes the majority of such 'I' thoughts come out true.

But second, and more seriously, I think that rational requirements can apply to entities that lack the specific kinds of concepts and mental states appealed to by these approaches. An entity may lack the capacity for indexical thoughts or for commitment to unifying projects but still instantiate sufficiently sophisticated mental states to make it sensible to evaluate it for rationality. If such entities can be appropriate targets of rational evaluation, then a self-conception criterion of rational subjecthood is untenable.

4 A Pluralist Proposal

We have seen that many purported rational requirements, such as those requiring coherence among attitudes held at a single time, presuppose a notion of rational subjecthood. And while I cannot argue the point here, I am skeptical of the prospects for somehow reformulating rational requirements to avoid reference to rational subjects. But I am also not satisfied by any of the criteria of rational subjecthood we have canvassed. So we face an impasse.

In my view, the proper response is to accept that the theory of rationality requires the concept of rational subjecthood, but to adopt a radically pluralist view of rational subjects. On this view, there are often many more rational subjects in a given case than we might have

thought. With a couple of important exceptions (below), for any two mental states there is a rational subject that has both mental states, as well as rational subjects that have one but not the other. We might put this by saying that parts of organisms, whole organisms, mereological sums of whole organisms, and mereological sums of parts of organisms can all be rational subjects (not to mention the possibility of inorganic rational subjects).[8]

This pluralist theory of rational subjecthood can be illustrated by analogy with mereological universalism. Mereological universalism says that for any set of objects, there is an object composed of all and only those objects. One motivation for mereological universalism is the seeming inability to come up with a satisfactory but more restrictive (though non-nihilistic) criterion of when some objects jointly compose another (van Cleve 2008), Mereological universalists thus believe in the existence of all kinds of strange objects, such as the object composed of the Eiffel Tower and the tip of my nose, to use a standard example. But they will of course concede that this and many other objects are best ignored, giving some pragmatic account of which objects are worth paying attention to (in the sense of giving names to them, tracking their movements, and so forth).

Similarly, the pluralist theory of rational subjecthood says that (subject to some metaphysical constraints, mentioned below), for any set of mental states had at a given time, there is a rational subject which has all and only those mental states. This means there can be all kinds of strange rational subjects which we wouldn't ordinarily notice. But the pluralist can then give a pragmatic account of which rational subjects are worth paying attention to.

Let me start with the metaphysical constraints, and then sketch the pragmatic account. First, if two mental states are realised by the very same physical matter, then there can be

[8]It is also worth considering a more deflationary view. On this view, there is no (or at least no metaphysically deep or fundamental) true answer to the question, 'When do two mental states belong to the same rational subject?' Instead, we should switch to the question, 'When does it make sense to treat two mental states as belonging to the same rational subject?' In my view, the answer to this question is that whether it makes sense to treat two mental states as belonging to the same rational subject depends on the goals and purposes we have in employing rationally evaluative language, as discussed below. I am unsure whether to prefer this deflationary view to the pluralist, realist one, which depends on various metaontological issues, and indeed I worry about how to draw the precise distinction between the two, at least if the deflationary view is combined with a minimalist account of truth (Dreier 2004).

no rational subject that has one but not the other. This was the upshot of the argument at the end of Section 3.2. Second, an entity must exist in order to be a rational subject. This means that what rational subjects there are depends on the metaphysics of composition. I just illustrated my pluralist proposal about rational subjecthood by analogy with mereological universalism. But while I am sympathetic to mereological universalism, pluralism does not require that universalism be true. The pluralist thesis is rather that for any set of mental states, there is a rational subject that has all and only those mental states *provided that there exists an object partly composed of the matter that realises all of those mental states* (and provided that for any mental state in the set, all mental states realised by the same physical matter are also included in the set; this was the first constraint). So, if there are no such entities as (non-atomic) parts of organisms, then there can be no rational subject that has all the mental states realised by one hemisphere of a split brain but not those realised by the other. Similarly, if there are no such entities as mereological sums of organisms, then there can be no rational subject that has all the mental states realised by one human brain and all those realised by another.

Supposing mereological universalism (or some other quite liberal ontology) is true, this means that there will nonetheless be vastly more, and stranger, rational subjects that we might have thought. But the pluralist can concede that most of these rational subjects are not worth paying much attention to. The pluralist can give a pragmatic account of which rational subjects we should focus on: which rational subjects it makes sense to attend to depends on the purposes we have in employing rationally evaluative language.

Most of the time, our purposes settle which rational subjects it makes sense to pay attention to. They settle, for instance, that it typically makes sense to attend to the rational subjects corresponding to whole human organisms, ignoring those corresponding to only parts of brains and those corresponding to mereological sums of human organisms.

But in the puzzle cases, we often have options. In such cases, it is most useful for some purposes to focus on the entity as a whole as our rational subject, despite its mental dis-

unity, while for other purposes it may be better to pay attention to the rational subjects corresponding to proper parts of that entity and whose mental states are fairly unified and coherent.

Consider the various roles played by rationally evaluative concepts. First, we use them to predict and explain behavior. *Ceteris paribus*, we predict that an entity will behave in the ways that it rationally ought to behave, and we can explain an entity's behavior by pointing out that such behavior was rational in light of the entity's situation. But our predictive and explanatory purposes needn't always point towards a single answer about which rational subjects to focus on in a given case. Take split-brain patients. On a pluralist view, there is a rational subject with all and only the mental states realised by the whole brain, as well as rational subjects with all and only the mental states realised by one or the other of its hemispheres. In ordinary contexts it is generally more useful to focus our attention only on the former rational subject. But in certain laboratory contexts, it can be useful to shift our focus to these single-hemisphere rational subjects. Consider, for example, the experiment sketched above, in which the word 'toothbrush' is flashed to the patient's left visual field (processed by the right hemisphere) while the word 'pencil' is flashed to the right visual field (processed by the left hemisphere). In making predictions about what will happen if we ask the patient to retrieve the object named with the left hand, or instead with the right hand, it is useful to talk about, and treat separately, the rational subject corresponding to the left hemisphere (which believes that the word 'pencil' was shown) and the one corresponding to the right hemisphere (which believes that the word 'toothbrush' was shown). The same goes for explaining why a given object was retrieved by a given hand.

Indeed, it is striking how easily practitioners, as well as laypeople reading about split-brain studies, fall into speaking of the hemispheres as having their own beliefs, desires, and actions. Of course, it is not *necessary* to focus attention on these smaller, single-hemisphere rational subjects. We could instead treat the patient as a whole as our subject while being careful to subscript its beliefs with the hemisphere in which the relevant processing takes place. But

on a pluralist view, we shouldn't think that one of these approaches is right and the other wrong. Insofar as it makes life easier to focus our attention on the rational subjects that have the mental states realised by one hemisphere but not those realised by the other, that is good enough reason for doing so.

Second, we use rationally evaluative language to criticise others, often with the goal of coaxing them to come closer to satisfying rational requirements. Again, while this role will generally favour focusing our attention on ordinary, whole-organism, rational subjects, it won't always do so. For instance, there would perhaps be little point in criticising Cerberus for having a belief instantiated in one head which conflicts with a belief instantiated in the other, as there may be no need and no direct way to resolve this conflict. (Or at least, there may be no more point in criticising Cerberus in this way than there is in criticising distinct humans for having beliefs that conflict.) But it may well make sense to attend to each head or brain separately, as the corresponding rational subject may be able to introspect and directly resolve conflicts among beliefs realised by that brain. Similar comments apply in the case of the Hogan twins. Despite the possible 'direct' mental connections between the brain (or portion of the brain) we call Krista's and the one we call Tatiana's, it typically will not serve our critical purposes to focus on the single rational subject that has all the mental states realised by those brains (or that brain), since, I presume, conflicts among some of those beliefs cannot be resolved except through characteristically interpersonal means. Better to address Krista and Tatiana separately, treating them and not their mereological sum as our rational subjects of interest.

5 Conclusion

Synchronic requirements of rationality govern how a subject ought to be at a single time. Purported synchronic requirements include the requirement that one's beliefs be consistent, that one's credences obey the axioms of probability, and that one's preferences be transitive. Diachronic requirements govern how a subject ought to be over time. Purported diachronic

requirements include the requirement that one's credences be updated by Bayesian conditionalization and that one later carry out the intentions one formed earlier.

Sychronic requirements presuppose for their application a notion of identity at a time, or what I have called 'rational subjecthood.' Diachronic requirements presuppose a notion of identity over time. In earlier work, I argued, in part on the basis of puzzle cases for identity over time and dissatisfaction with all of the solutions on offer, that we should reject all diachronic requirements of rationality in favour of a purely synchronic theory of rationality (Hedden 2015). Such a theory avoids being held hostage to the problematic metaphysics of identity over time.

But things are more difficult in the synchronic case. Just as with identity over time, we have puzzle cases for identity at a time, and good reasons for dissatisfaction with all the standard solutions. But whereas theorists have argued that it is possible to have a respectable theory of rationality which does away with diachronic requirements, it doesn't seem possible to do away with synchronic requirements as well and hence avoid reliance on the notion of a rational subject.

I have suggested that the way to respond to this puzzle is to adopt a pluralist account of rational subjecthood, along with a pragmatic account of which of these rational subjects it makes sense to focus on in a given context. There are rational subjects corresponding to every possible way of individuating them not ruled out by the minimal metaphysical constraints mentioned above. But whether we should pay attention to a given rational subject depends on our purposes in employing rationally evaluative language. If that is correct, it motivates a rethinking of the debate over diachronic requirements of rationality. Rather than responding to puzzle cases for identity over time by rejecting diachronic requirements of rationality, perhaps we should adopt a similarly pluralist account of rational subjecthood-over-time, on which there are lots of different (but partly overlapping) temporally extended rational subjects, but which ones we should pay attention to depends on our purposes and interests. In this way, consideration of puzzle cases for identity at a time undermines one

motivation for a purely synchronic theory of rationality.

References

Ainslie, G. 1992. *Picoeconomics*. Cambridge: Cambridge University Press.

Bayne, T. 2010. *The Unity of Consciousness*. Oxford: Oxford University Press.

Campbell, T. and McMahan, J. 2010. 'Animalism and the Varieties of Conjoined Twinning.' *Theoretical Medicine and Bioethics* 31: 285-301.

Davidson, D. 1982. 'Two Paradoxes of Irrationality.' In R. Wollheim and J. Hopkins (eds.) *Philosophical Essays on Freud*, Cambridge: Cambridge University Press, 289–305.

Davis, L. 1997. 'Cerebral Hemispheres.' *Philosophical Studies* 87: 201-22.

Dominus, S. 2011. 'Could Conjoined Twins Share a Mind?' *The New York Times Magazine*, May 25, 2011.

Dreier, J. 2004. 'Meta-ethics and the problem of creeping minimalism.' *Philosophical Perspectives* 18 (1):23–44.

Egan, A. 2008. 'Seeing and believing: perception, belief formation and the divided mind.' *Philosophical Studies* 140 (1):47–63.

Elga, A. and Rayo, A. 2015. 'Fragmentation and Information Access.' manuscript. Available: https://www.princeton.edu/ adame/papers/fragment/fragmentation-and-information-access-2015-04-08.pdf

Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. 2016. *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. Farrar, Straus, and Giroux.

Greco, D. 2014. "Iteration and Fragmentation." *Philosophy and Phenomenological Research* 88 (1):656–673.

Hanlon, R. and Messenger, J. 1996. *Cephalopod Behaviour*. Cambridge: Cambridge Univer-

sity Press.

Hedden, B. 2015. *Reasons without Persons: Rationality, Identity, and Time*. Oxford: Oxford University Press.

Kahneman, D. 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kim, J. 2010. 'Thoughts on Sydney Shoemaker's *Physical Realization*.' *Philosophical Studies* 148 (1): 101-12.

Korsgaard, C. 1989. 'Personal identity and the unity of agency: A Kantian response to Parfit.' *Philosophy and Public Affairs* 18 (2):103–31.

Lewis, D. 1982. 'Logic for Equivocators.' *Noûs* 16 (3):431–441.

Liao, S.M. 2006. 'The Organism View Defended.' *The Monist* 89: 334-50.

Nagel, T. 1971. 'Brain bisection and the unity of consciousness.' *Synthese* 22 (May):396-413.

Olson, E. 1997. *The Human Animal*. Oxford: Oxford University Press.

Olson, E. 2003. 'Was Jekyll Hyde?' *Philosophy and Phenomenological Research* 66 (2):328–348.

Olson, E. 2014. 'The Metaphysical Implications of Conjoined Twinning.' *The Southern Journal of Philosophy* 52: 24-40.

Putnam, H. 1960. 'Minds and Machines.' In S. Hook (ed), *Dimensions of Mind*. New York: New York University Press.

Rovane, C. 1998. *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.

Schechter, E. 2012. 'Intentions and Unified Agency: Insights from the Split-brain Phenomenon.' *Mind and Language* 27 (5):570–594.

Schechter, E. 2015. 'The Subject in Neuropsychology: Individuating Minds in the Split-Brain

Case.' *Mind and Language* 30 (5):501–25.

Shoemaker, S. 1984. 'Personal Identity: A Materialist's Account.' In Shoemaker and Swinburne (eds), *Personal Identity*: Oxford: Blackwell Press.

Smithies, D. *The Epistemic Role of Consciousness*. Draft monograph.

Snowdon, P. 2014. *Persons, Animals, Ourselves*. Oxford: Oxford University Press.

Stalnaker, R. 1984. *Inquiry*. Cambridge, MA: MIT Press.

Sumbre, G, Gutfreund, Y, Fiorito, G, Flash, T, and Hochner, B. 2001. 'Control of Octopus Arm Extension by a Peripheral Motor Program.' *Science* 293 (5536): 1845-1848.

van Cleve, J. 2008. 'The Moon and Sixpence: A Defense of Mereological Universalism.' In Hawthorne, Sider, and Zimmerman (eds), *Contemporary Debates in Metaphysics*. Oxford: Blackwell Press, 321-40.

van Inwagen, P. 1990. *Material Beings*. Ithaca, NY: Cornell University Press.

Yablo, S. 1992. 'Mental Causation.' *Philosophical Review* 101: 245-80.