Erkenntnis
https://doi.org/10.1007/s10670-023-00722-2

ORIGINAL RESEARCH

Some Problems for the Phenomenal Approach to Personal Identity

Ivar Labukt[1]

Received: 21 October 2022 / Accepted: 3 July 2023

Abstract
I present some problems for phenomenal (i.e. consciousness-based) accounts of personal identity and egoistic concern. These accounts typically rely on continuity in the capacity for consciousness to explain how we survive ordinary periods of unconsciousness such as dreamless sleep. I offer some thought experiments where continuity in the capacity for consciousness does not seem sufficient for survival and some where it does not seem necessary. There are ways of modifying the standard phenomenal approach so as to avoid these difficulties, but I argue that they all lead to other problems that are no less serious.

According to the phenomenal approach to personal identity, the crucial ingredient in survival it is that your conscious life goes on. As long as there are future experiences that are, in the right way, a continuation of your present experiences, you will be there, too. This idea has considerable initial plausibility. It is difficult to imagine ceasing to exist in the middle of a continuous conscious experience (cf. Dainton, 2008, ch. 1.5; Duncan, 2015/2020). It is also difficult to imagine continuing to exist after having permanently lost all capacity for consciousness (cf. McMahan, 2002, part 5, ch. 1; Dainton and Bayne, 2005, p. 558).

My aim in this paper is to show that after this promising start, the phenomenal approach runs into some problems. In order to explain how we survive ordinary periods of unconsciousness such as dreamless sleep or general anesthesia, phenomenal accounts typically rely on continuity in the capacity for consciousness. But while continuity of actual consciousness arguably seems to guarantee your continued existence, the idea that a mere capacity for consciousness is sufficient for survival turns out to be significantly less appealing. This makes it harder for supporters of the phe-

✉ Ivar Labukt
ivar.r.labukt@uit.no

[1] Department of Philosophy and First Semester Studies, UiT The Arctic University of Norway, Tromsø, Norway

Published online: 17 August 2023

nomenal approach to defend the survivability of radical breaks in non-phenomenal continuity, such as receiving a brand new body or brand new memories. It also makes it harder to identify a suitable notion of phenomenal continuity; in particular, I will argue that any policy towards continuity in the physical basis for consciousness is bound to be at least somewhat counterintuitive. Finally, while we might be happy to say that a permanent loss of the capacity for consciousness ends a person's existence, it is much less tempting to hold that all temporary losses of the capacity have this effect. We do, after all, take reversible comas to be survivable.

These problems can be avoided by versions of the phenomenal approach that do not rely on continuity in the capacity for consciousness. However, it turns out that these views all lead to other problems that are at least as serious. There are, in other words, some unpleasant choices facing the phenomenal approach.

Why does this matter? Most obviously, perhaps, it could be a reason to reject the approach. I suspect, however, that quite a few of its defenders would claim that their view can withstand the difficulties since the considerations in favor of it are so strong (cf. Duncan, 2020), or since the non-phenomenal alternatives are even worse. For these committed supporters, the unpleasant choices will be significant in a different way: they actually have to be made. Depending on one's starting point, then, the arguments in this paper can be taken primarily as an objection to the phenomenal approach, or as an exploration of it. I will remain neutral between these alternatives, but I will revisit the issue at the end of the paper.

I explain in more detail what the phenomenal approach is and why it appears promising in Sect. 2. I discuss the problems in Sects. 3-5. First, however, I will say a few words about the philosophical issues we are concerned with.

1 Preliminaries

The phenomenal approach has primarily been presented as an approach to personal identity (Foster, 1991, ch. 8, McMahan, 2002, part 1, ch. 5.1, Dainton and Bayne, 2005, Wright, 2006, Dainton, 2008, Duncan, 2020). Somewhat more precisely, it is taken as an attempt to specify when persons are *numerically* identical – that is, one and the same person. The issue is most pressing when we consider persons at different times. What, for instance, does it take for a person existing in 10 years to be me? If we can uncover the correct identity criteria for such cases, we know what it takes for persons to *persist* or *survive* or *continue to exist*.

There are also accounts of the *object of egoistic concern* that are based on consciousness (Crisp, 2006, ch. 5.1, McMahan, 2002, part 1, ch. 5.2). Accounts of this kind aim to explain which future person(s) you should care about when thinking egoistically.

In this paper, I will take the various views I discuss to be views about both personal identity and the object of egoistic concern. I will usually write as if the question "which future person do I have an egoistic reason to care about" is just another way of framing the question "which future person is me?" It may seem obvious that it is. But as has become clear from the philosophical discussion (in particular Parfit, 1984), there are actually good reasons for thinking that personal identity and rational egois-

tic concern can, at least to some extent, come apart. One could still take the two issues to be very closely connected, so that if consciousness is the key to resolving one of them, it is also going to be the key to resolving the other (cf. Dainton, 2008, p. xxiii). Treating the issues as equivalent is then a safe simplification for the purposes of this paper. But perhaps there will be readers who only find the phenomenal approach suitable for handling one of them (cf. McMahan, 2002, part 1, ch. 5; Gottlieb, 2021). Such readers should feel free to translate between talk about personal identity and talk about egoistic concern as they see fit.[1]

By consciousness, I mean phenomenal consciousness, and when I talk about experiences, I mean phenomenally conscious experiences. As philosophers like to say, there is something it is like to experience a phenomenal state. Beyond that, the phenomenal is not exactly a well understood aspect of reality. Except in Sect. 5.2, I will assume that consciousness is in some sense produced by or grounded in physical processes in brain, but I will remain agnostic about the exact relationship between the phenomenal and the underlying physical processes.

2 The Phenomenal Approach

Since the phenomenal approach has not been discussed as extensively as some of its rivals, I will spend a few pages explaining what it is and why it merits scrutiny – that is, why it has, at least initially, considerable plausibility.

I will begin with the observation that continuous experience, of the kind had in an ordinary stream of consciousness, seems sufficient for personal persistence. As Barry Dainton puts it, it not exactly easy to imagine that your current stream of consciousness goes on but fails to take you with it (Dainton, 2008, p. 26; see also Wright, Sect. 7). To illustrate, let's say you have a piece of chocolate and close your eyes. You know that for the next ten seconds, whatever mechanism in your brain that is producing your current sensation and enjoyment of the taste of chocolate will produce another ten seconds of this experience. That seems to guarantee that you will exist for at least ten more seconds. There are, to be sure, of all sorts of non-phenomenal continuities that could be lost during this time. Aliens might painlessly replace all your memories or give you a brand new body. But it is still difficult to see how *you* could be completely gone by the end of this continuous experience. Say that the person existing towards the end of the chocolate experience will either get to enjoy one more piece of chocolate, or be tortured for hours on end. Mustn't there be at least some egoistic reason for preferring the former alternative?

According to Matt Duncan, the idea that continuous experience is sufficient for survival is not merely intuitively plausible in the way that many claims about personal identity can be. He argues that we can know with certainty in some special and immediate first-personal way that we persist during brief but nevertheless temporally

[1] There are also other practical concerns that can seem closely related to personal identity. For instance, you might expect that if you continue to exist, you will retain moral responsibility for your current actions. I will not discuss these further practical concerns. For an argument that phenomenal continuity is irrelevant for moral responsibility, see Gottlieb 2021.

extended experiences, such as thinking that 2+2=4 (Duncan, 2015). If we take this argument to be sound, we are more or less forced to accept a phenomenal outlook on personal identity (Duncan, 2020, Sect. 2).[2] But even if we take our responses to cases like the ones presented above to be ordinary fallible intuitions, they are still powerful enough to clearly suggest that the phenomenal approach is on to something.

Some form of continuation of conscious life can also easily seem necessary for personal persistence. Barry Dainton and Tim Bayne provide a vivid illustration of this. They ask us to consider the prospect of zombification: you completely lose your capacity for phenomenal consciousness. The zombie can do all the things you can do and is outwardly indistinguishable from a person with phenomenal consciousness. Still, Dainton and Bayne "suspect that many will respond to this fate in much the way that they respond to the prospect of imminent death: one aspect of my mind might survive, but I won't!" (Dainton & Bayne, 2005, p. 558; see also Dainton, 2008, 79, 190-191; McMahan, 2002, part 5, ch. 1). This is certainly my intuitive response. I do not merely have the impression that I would miss out on some experiences that it would be nice to have. The problem seems to go deeper. After irreversible zombification, I would not be deprived; I would be gone.

It is also worth mentioning a less fanciful reason for favoring the phenomenal approach, at least in comparison with the view on personal identity that has long been dominant among philosophers. This is the view that personal persistence involves continuities in certain non-phenomenal psychological entities such as memories, plans and character traits (see e.g. Parfit, 1984). In severe cases of dementia, these continuities are broken: the capacity for engaging in long-term projects disappears along with almost all episodic memory, and there can be dramatic changes in character. It would thus seem that on the standard, non-phenomenal psychological views, *you* could never come to suffer from severe dementia. There is, accordingly, no (direct) egoistic reason for making sure that if you develop, say, Alzheimer's disease, the human being existing in the final stages of the disease will be taken well care of. The phenomenal approach can avoid these implications, since the capacity for phenomenal consciousness seems to remain intact even in the most seriously affected patients. We can thus accept that we may end up as such patients, and that we should take this into account in our egoistic deliberation. This seems more in line with how we ordinarily think about dementia (see McMahan, 2002, part 1, ch. 4).

There are of course lots of reasonable things supporters of non-phenomenal accounts could say in response to the claims made here. This section is not supposed to establish that the phenomenal approach is more promising than all the alterna-

[2] When evaluating the argument, it is useful to distinguish between two claims: (1) There is a temporally extended experience of thinking that 2+2=4, and (2) all of this temporally extended experience is had by one and the same person, i.e. you. Perhaps you can never be absolutely certain that a claim like (1) is correct. After all, it would seem that it is always possible that the universe was created just this instant. But aside from radical skeptical worries of this kind, I am inclined to accept that we are directly acquainted with temporally extended experience. I think it is more doubtful whether (2) is knowable in the same way. Immediate access to continuous experience does not necessarily entail immediate access to the significance of continuous experience to personal identity. A proper discussion of these issues is, however, beyond the scope of this paper.

tives. I am content to claim that it is sufficiently promising to make examining it worthwhile.[3]

There is a long way to go from the considerations above to a complete and satisfactory theory of personal identity. An obvious issue to deal with is the fact that ordinary human lives contain frequent periods of unconsciousness; we undergo dreamless sleep almost every night, as well as occasional fainting and general anesthesia. Unsurprisingly, most defenders of the phenomenal approach have wanted to say that we survive such periods of unconsciousness. They have typically done this by requiring continuity in the *capacity* for consciousness rather than continuity in consciousness itself. For now, I will assume that this is the way to go when developing the phenomenal approach. I will explain the idea in more detail in Sect. 4 and consider alternatives to it in Sect. 5.

We also need to be provided with a more detailed explanation of what it takes for consciousness, or the capacity of consciousness, to go on into the future. A natural thought might be that what unites the different elements of a stream of consciousness is simply the fact that they are experienced by one and the same person. Similarly, you might think that what relates your future capacity for consciousness to your present is the fact that both capacities belong to you. This would of course be an objectionably circular explanation in a phenomenal account of personal identity. Such accounts presuppose that there is some sort of more basic and direct link between phenomenal entities that can be spelled out without a prior understanding of what makes persons numerically identical. For the purposes of illustration, I will briefly mention two possibilities.

One possibility, favored by Duncan, is to define continuity of consciousness in terms of the physical structures involved. More precisely, Duncan's suggestion is that two capacities for consciousness at adjacent times are identical if and only if they share physical parts that would be capable of producing consciousness on their own. Capacities at non-adjacent times are identical if and only if they belong to an unbroken chain of capacities that are connected at adjacent times in the way just specified. Since none of these criteria presuppose criteria for personal identity; we can now informatively say that two persons are identical if and only if they have the same capacity for consciousness (Duncan, 2020, p. 2044-2046).

Dainton (2008) instead relies on an irreducibly phenomenal form of continuity. The idea is that experiences can be directly phenomenally connected, i.e. experienced together in some primitive way that does not presuppose personal identity. He refers to such experiences as *co-conscious*. Persons are identical if they have co-conscious experiences, or if they have capacities for consciousness that, had they been active, would have produced co-conscious experiences. As with Duncan's view, there can be chains of experiences or capacities that are connected in the most basic and direct way, and persons will be identical if and only if there is such an unbroken chain between them.[4]

[3] More detailed and dialectically ambitious comparisons with non-phenomenal views can be found in McMahan 2002, Dainton & Bayne 2005, Wright 2006 and Dainton 2008.

[4] This brief summary does not do justice to the impressive richness and precision in Dainton 2008 (see also Dainton & Bayne 2005). For a very similar view, see Wright 2006. For a similar view, see Foster

In the next sections, I will describe some problems facing accounts of personal identity along these lines.

3 Different Intuitions About Active And Inactive Capacities For Consciousness

On the phenomenal approach just sketched, an inactive capacity for consciousness is just as good actual consciousness when it comes to survival. This makes it possible to classify ordinary periods of unconsciousness as survivable. There is, however, a problem lurking here. While we have a strong intuitive propensity to find continuity in actual experience sufficient for personal persistence, the same is not true of continuity in inactive capacities for consciousness. As Peter Unger notes, we are much less inclined to believe that persons survive certain changes if there is no actual consciousness present (Unger, 1990, ch. 2.4-2.5). I will examine two versions of this intuitive discrepancy in this section. They both contribute to weakening the intuitive support of the phenomenal approach.

3.1 Non-phenomenal Continuity

As we have seen, it is tempting to believe that as long your stream of consciousness goes on, you can persist in spite of radical non-phenomenal changes such as receiving new memories or a new body. It is not equally tempting to believe that such changes are survivable when they happen while you are unconscious. Here is a case to illustrate:

The Conspiracy For reasons only they know, aliens anesthetize you and Barack Obama. There is no consciousness during the procedure. The aliens destroy your body and all of your brain except the neural basis for your capacity for consciousness. This neuronal matter is safely stored in a vat. The aliens then remove and destroy the neural basis for consciousness in Barack Obama's brain. They replace it with the neural matter from the vat. After a couple of hours, the anesthetic wears off and a person becomes conscious, reminiscing about being re-elected as president of the United States.

According to the phenomenal approach, the post-surgery person who looks, thinks, talks and acts like Barack Obama will in fact be you. The person does, after all, have the same capacity for consciousness as you do before the procedure. Does this seem right? This is clearly not a case about which there is likely to be universal intuitive agreement. I do, however, think it is safe to say that more people would be skeptical about the prospects for survival in this case than in a corresponding case involving a continuous stream of consciousness running from your pre-surgery experiences to the post-surgery reminiscing.

1991, ch. 8.

Now it may not seem reasonable to take this intuitive discrepancy to be indicative of a genuine difference in outcome. It would certainly be surprising if simply adding an ordinary general anesthetic to a procedure could in and of itself make it non-survivable. Presumably, the loss of non-phenomenal continuity either undermines personal identity in both the conscious and the unconscious versions of such cases, or in none of them.

This makes the present problem a bit more manageable for the phenomenal approach. Its defenders can argue that as long as we are convinced that continuity of actual consciousness guarantees persistence, we should also be prepared to say that continuity in the capacity for consciousness does so – even if this latter claim on its own is far from obviously correct. But critics could of course also make the opposite argument: when we realize that loss of non-phenomenal continuity undermines personal identity during periods of unconsciousness, we should also admit that it does so when it happens during a stream of consciousness – even if the presence of experiential continuity might incline us to think otherwise. This would be in line with the more general lesson drawn by Unger after considering some broadly similar cases, namely that actual consciousness serves as a distraction to our intuitive judgments about personal identity, and that it is our judgments about cases without consciousness we should trust (Unger, 1990, ch. 2.6-2.7).[5]

The idea that one can survive a complete break in non-phenomenal continuity is, in other words, less plausible than one might think if one only considers the cases involving actual consciousness relied upon by defenders of the phenomenal approach.

3.2 Continuity in The Physical Basis for Consciousness

Focusing on inactive capacities of consciousness not only makes it possible to reveal intuitions about non-phenomenal forms of continuity that tell against the phenomenal approach. It also helps uncover a conflict in our intuitions about a phenomenal form of continuity. More precisely, I will argue that there is no attractive way for the phenomenal approach to deal with continuity in the physical basis for consciousness.

We can begin by noting that sudden and complete replacement of the physical basis for consciousness is clearly not survivable on the phenomenal approach. If someone were to destroy the consciousness-producing mechanism in your brain and afterwards replace it with a perfect copy, you would also be destroyed in the process. On Duncan's proposed criterion, the problem would be the lack of shared parts between the mechanisms. On Dainton's view, the problem would be that the experiences produced by the new mechanism could not be co-conscious with the experiences produced by the old one. In this kind of case, it makes no difference to our intuitions whether the procedure takes place with or without anesthetics, as there would not in any case be a continuous experience running through it.

[5] When defending this view, Unger's aim is not to discredit the phenomenal approach. Indeed, he later argues that that the capacity for consciousness is part of our core psychology, the continuous physical realization of which underlies personal identity (Unger 1990). However, he nevertheless cannot be considered a (wholehearted) supporter of the phenomenal approach, since he is inclined to believe that the capacity for consciousness is not an essential part of the core psychology, and that it is therefore possible to survive without it (Unger 1990, ch. 5.2).

It is less obvious how the phenomenal approach should handle gradual replacement of the physical basis for consciousness. Now it should, presumably, allow replacing just a few of the consciousness-producing neurons with perfect copies. This would certainly leave enough shared parts for Duncan's criterion to be fulfilled, and the kind of phenomenal connection posited by Dainton could still be sustained. But how often can you do this? On Duncan's and Dainton's views, there are no restrictions. The next instant, a few more neurons could be replaced. And so on. These views thus allow for a complete gradual replacement of the material basis for consciousness during a very short period of time – say, for instance, 10 s (cf. Dainton, 2008, p. 165-167). Jeff McMahan's version of the phenomenal approach, on the other hand, does not. On this view, complete gradual replacement of the physical basis for consciousness is only compatible with personal persistence if it happens much more slowly (McMahan, 2002, part 1, ch. 5).

When it comes to fast gradual replacement, it does make a difference to our intuitions whether it happens with or without consciousness. Let us begin with a case of the latter kind:

Phenomenal Replacement 1 Aliens anaesthetize you. They destroy all the neurons in the consciousness-generating mechanism of your brain and replace them with perfect copies. The replacement is done gradually, one neuron at a time, but takes only 10 s. After a couple of hours, the anesthetic wears off and a person becomes conscious.

To the extent that I find a phenomenal approach to personal identity convincing, I find this procedure worrisome. It does not seem that the post-surgery experiences will be produced by the same thing that is producing my pre-surgery experiences in any important sense. Rather, the procedure seems disturbingly similar to this alternative procedure, which, if the phenomenal approach is correct, clearly would end my existence:

Phenomenal Replacement 2 Aliens anaesthetize you. They remove and destroy the neural basis for your capacity for consciousness. 10 s later, they replace it with a perfect copy. After a couple of hours, the anesthetic wears off and a person becomes conscious.

I find it hard to believe that the fact that the new neurons spend a few moments connected to the old ones in one Phenomenal Replacement 1 makes a crucial difference to the outcome. The mere possibility of consciousness during the replacement seems too flimsy a basis for persistence. From an egoistic point of view, it does not seem matter at all whether I undergo Phenomenal Replacement 1 or 2.

More generally, it appears that if I am to survive a period of unconsciousness in virtue of an inactive capacity for consciousness, there must be some fairly robust continuity in the physical stuff that has the capacity during this period. The capacity cannot, as it were, surf around on an ever-changing stream of material bases before consciousness resumes.

As long as we stick to scenarios without consciousness, then, McMahan's attitude towards replacement of the physical basis for consciousness seems preferable. Let us

now consider the same kind of replacement with actual experiences running through it. Here is a case of this kind:

Phenomenal Replacement 3 You have a piece of chocolate and close your eyes. For the next 10 s, there is continuous enjoyment of the taste of chocolate. During this time, aliens destroy all the neurons in the consciousness-generating mechanism of your brain and replace them with perfect copies. The replacement is done gradually, one neuron at a time, and does not at any point affect the nature of the experience.

Though I won't go as far as saying that this procedure seems innocuous, there is a very strong temptation to think that I will go with the flow of experience. From an egoistic point of view, it seems to matter greatly what happens after the replacement is complete. If you are unsure about what to make of the case, perhaps it might help to consider the fact that this kind of replacement could be happening to you right now. Would that really entail that you will be gone in a few seconds?

When it comes to cases with streams of actual consciousness, then, McMahan's requirement of physical continuity in the basis for consciousness suddenly seems otiose. Now I do not want to claim that every single person thinking about how to best develop the phenomenal approach will end up in this predicament, but I do think there will be a significant tendency for this to happen.[6] As with the losses of non-phenomenal continuity discussed above, it does not seem like an attractive option to take the intuitive divergence at face value. There is something deeply implausible, or at least revisionist, in the idea that your survival could actually depend on your capacity for consciousness being active – that, for instance, falling asleep during a period of rapid gradual physical replacement would end your existence.

The present problem is thus not just that it is difficult to ascertain precisely how much physical continuity to require. This is likely to be the case on most approaches to personal identity, and can be seen primarily as a puzzle for the supporters to work out, rather than a threat to the approach itself. What we have encountered in this section is a *conflict* in the attitudes that can be relied upon when developing phenomenal accounts of personal identity. A form of phenomenal continuity seems necessary for personal persistence in one scenario and not another, and at the same time we feel that there shouldn't be a difference of this kind between the scenarios. This is a more serious problem than uncertainty about which exact version of an approach to defend. For one thing, any way of resolving the conflict will yield a view with some aspects that are positively counterintuitive (and not just aspects that do not have particularly strong intuitive support). Second, and perhaps more importantly, the conflict may to some extent undermine the general credibility of the intuitions in favor of the

[6] For a defense of the claim that most people find rapid neural replacement survivable when actual consciousness is present and not survivable when it is not, see Unger 1990, p. 53–54. Dainton agrees with Unger that there is an intuitive discrepancy of this kind, but seems to think that it goes away as long as we stipulate that there is a continuous capacity for consciousness throughout the replacements (Dainton 2008, p. 165–167). I doubt that this is generally true, but it is certainly not a plausible diagnosis of the specific case from Unger Dainton addresses. Unger explicitly says that the unconscious version of the case is just like the conscious one except for the presence of an anesthetic (Unger 1990. p. 53–54), which means that it does contain a continuous capacity for consciousness in Dainton's sense.

approach. If it clear that we have to reject at least some of them, this makes the job easier for those who want to argue that we should reject all of them.

4 Temporary Losses of The Capacity for Consciousness

According to the (standard) phenomenal approach, continuity in the capacity for consciousness is not only sufficient for personal persistence; it is also necessary. We have seen that the former idea has questionable intuitive support. What about the latter?

As illustrated by the prospect of zombification, it does seem that for me to exist in the future, I must have some sort of conscious life in the future. For this to happen, I must have some sort of capacity for consciousness in the future. In this sense, then, a capacity for consciousness seems necessary for personal persistence. However, this must not be confused with support for the stronger claim that a *continuous* capacity for consciousness is necessary for persistence, i.e. that even the briefest absence of such a capacity necessarily and irrevocably ends your existence. This, I will argue in this section, is not an attractive view.

Before making this argument, I need to specify more precisely what is meant by a capacity for consciousness in the present context. Following Dainton's and Duncan's terminology, let us say that something has a capacity for consciousness if and only if it will produce consciousness when *triggered* (see Dainton, 2008, ch. 4 and 10.7) or *appropriately stimulated* (see Duncan, 2020, p. 2042-2043). This of course invites the question of what sort of causal influence counts as triggering or appropriate stimulation. There clearly must be some restrictions if continuity in the capacity for consciousness is to form the basis of a plausible account of personal identity. For instance, it is, presumably, possible to rearrange the atoms in my desk so as to make them produce phenomenal experience. We would not want to say that persisting desks are therefore instances of personal persistence. Making desks conscious should count as creating a new capacity for consciousness rather triggering an already existing one.

Understandably, Dainton and Duncan do not attempt to draw a completely precise line between stimulating and creating a capacity for consciousness. But it is clear that they have a fairly restrictive idea of what counts as stimulation. This is how Duncan puts it:

> Appropriate stimulation may include any intervention that does not add to or subtract from the sum of the token parts directly responsible for one's consciousness, and does not alter the structural arrangement of those parts except for minor alterations that are mere bi-products of stimulation (such as small changes in neuronal arrangement that inevitably result from any stimulation). (Duncan, 2020, p. 2043)

Even on this fairly restrictive definition, there is no doubt that our capacity for consciousness remains intact during ordinary periods of unconsciousness. Waking up from dreamless sleep or from general anesthesia does not require direct manipulation of the structural arrangements of the brain parts that generate consciousness.

4.1 Intact Capacities for Consciousness

I will begin with the most straightforward version of the idea that continuity in the capacity for consciousness is necessary for personal identity. This is the view, defended by Duncan (2020, Sect. 3), that an *intact* capacity for consciousness must be present at all times. Only breaks in consciousness that are due to lack of appropriate stimulation are allowed. If the capacity for consciousness itself is malfunctioning, even just barely and for an instant, the person is destroyed. This view is, I will argue, too demanding.

Unlike in the previous section, we are not forced to rely exclusively on hypothetical scenarios to illustrate the problem. We can think of reversible comas. Now some comas, such as those induced by drugs, presumably do not involve loss of continuity in the capacity for consciousness. They merely prevent an intact capacity from being triggered. But it seems likely that some reversible comas involve structural damage to parts of the brain that generate consciousness and that there is a period when the brain cannot produce consciousness no matter how much appropriate stimulation it receives. On Duncan's view, it is impossible to survive this kind of coma. When the damage is repaired, a new person is created.

This clearly conflicts with our attitudes towards comas. At least as long as all normal psychological functioning is restored, we do not have any doubt that someone waking up from a coma is the same person as the one who fell into it. We do not find it odd to suppose that it matters greatly from an egoistic point of view whether a coma will be reversed or not. Admittedly, given our ignorance about precisely where and how consciousness is generated in the brain, we cannot rule out the possibility that none of the actual cases to which we have this attitude have involved a temporally damaged capacity for consciousness. Perhaps the brain lacks the ability to repair this capacity, and that comas are always reversed because appropriate stimulation resumes. But this possibility does not really help the present view. It is clear that our attitude to comas is not conditional on it being actual. We do not take reversible comas to be survivable because we make assumptions about which parts of the brain that are temporally malfunctioning.

To bring out even more clearly the implausibility of requiring a continuously intact capacity for consciousness, let us conduct a thought experiment as well:

Tonsillectomy 1 You are in hospital to have your tonsils removed. You are anesthetized; there is no consciousness throughout the procedure. The surgeon then inadvertently gives you an additional drug that causes some relatively small but functionally important structural changes in the parts of your brain directly responsible for consciousness. With its current structure, your brain could not produce consciousness even after the anesthetic has been metabolized. However, the brain in response releases a chemical that causes the structural changes to be completely reversed. After a couple of hours, a person wakes up at exactly the time you would otherwise have woken up, with a brain composed of exactly the same matter organized in exactly the same way.

If a continuous series of intact capacities for consciousness is necessary for personal identity, you could not survive this procedure. But though the surgeon clearly deserves some sort of criticism afterwards, it does not seem to me that he can be reasonably accused of killing you. I find it hard to believe that his mistake could be extraordinarily lucky for the post-surgery person by allowing this person to come into existence and, as it were, take residence in what was previously your brain.

The demand that an intact capacity for consciousness be present at all times thus seems implausibly strict. If this is the best the phenomenal approach can do, I think we are entitled to say that it has a pretty serious problem with temporary breaks in the capacity for consciousness. The only defenders of the approach to discuss this issue in any detail agree; Dainton and Bayne find classifying a similar case as non-survivable "highly implausible" (Dainton & Bayne, 2005, p. 569; see also Dainton, 2008, p. 328).[7]

It is, however, possible to avoid these implications without giving up on the idea that a continuous capacity for consciousness is necessary for survival. This can be accomplished by introducing new kinds of capacities for consciousness. In the following, I will examine two versions of this strategy.

4.2 Second-Order Capacities for Consciousness

Let us begin with Dainton and Bayne's proposed remedy, which is to allow that persistence can be grounded by a *second-order* capacity for consciousness as well as a first-order capacity. A thing has a second-order capacity for consciousness "if it doesn't have a first-order capacity, but will acquire one if left to its own devices, under normal background conditions." (Dainton & Bayne, 2005, p. 569; see also Dainton, 2008, p. 319-321) This proposal ensures that the tonsillectomy case, and more generally all (currently) reversible comas, come out as survivable.

The problem with this proposal is that it puts an enormous weight on the distinction between internal and external causes of reacquisition of the capacity of consciousness. You continue to exist as long as your brain can repair itself, but cease to exist if any assistance beyond normal background conditions is required from the outside. This does not seem like the right way to draw the line between life and death. Consider the prospect of falling into a coma due to structural damage in the brain parts directly responsible for your consciousness. Imagine that there is a way of rearranging the parts so that they will regain their former, fully functional structure. From an egoistic point of view, does it really matter how the parts are returned to their previous configuration? Is there no reason to hope that the capacity for consciousness will be restored in the event that external intervention, no matter how subtle, is required? Or consider this variation of the tonsillectomy case:

[7] Duncan only addresses the issue briefly, in a footnote. He admits that his view has "(apparently) counterintuitive" implications about comas. He does not provide any defense of these implications except to say that he embraces them (though he does indicate how his view could be modified so as to avoid them) (Duncan 2020, p. 2046, note 16).

Tonsillectomy 2 You are in hospital to have your tonsils removed. You are anesthetized; there is no consciousness throughout the procedure. The surgeon then inadvertently gives you an additional drug that causes some relatively small but functionally important structural changes in the parts of your brain that have already been subdued by the anesthetic. With its current structure, your brain could not produce consciousness even after the anesthetic has been metabolized. There is no way for your brain to fix the problem. Realizing his mistake, the surgeon quickly administers a chemical that causes the structural changes to be completely reversed. A person wakes up at exactly the time you would otherwise have woken up, with a brain composed of exactly the same matter organized in exactly the same way.

In this case, the second-order capacity for consciousness is also disrupted, since the chemical is administered by the surgeon and not released by your brain. But claiming that you will exist after the first tonsillectomy case but not after this one is, I think, even less attractive than saying that you will stop existing in both cases. More generally, a small and temporary damage to the second-order capacity for consciousness seems no more metaphysically or normatively significant than a small and temporary damage to the first-order capacity for consciousness.

4.3 Defective Capacities for Consciousness

Dainton admits that even breaks in the second-order capacity for consciousness may seem survivable. As an alternative option for the phenomenal approach, he proposes a more liberal view where survival can be secured also by "defective" capacities for consciousness (Dainton, 2008, p. 328-329). You have a defective capacity for consciousness if you are unable to have phenomenal experience but could gain this ability " if subjected to certain alterations of a fairly minor sort." (Dainton, 2008, p. 315) Since there is no requirement that these alterations must be internally initiated, this modification allows us to classify both tonsillectomy 1 and 2 as survivable. And while some reversed comas may have involved the temporary loss of an intact capacity for consciousness, there was presumably a defective capacity there the whole time, which would make these comas survivable on the present proposal.

This modification of the phenomenal approach of course invites the question of whether it is possible to survive a temporary loss of even a defective capacity for consciousness. To help us think about this, here is another tonsillectomy case:

Tonsillectomy 3 You are in hospital to have your tonsils removed. You are anesthetized; there is no consciousness throughout the procedure. The surgeon has forgotten where the tonsils are and mistakenly points his laser towards the part of the brain directly responsible for consciousness. By the time he realizes his mistake, your brain is only somewhat closer to being able to produce consciousness than your liver. However, the hospital has invested in an incredibly advanced machine that scans all brains and, in case of accidents, can put them back together precisely as they were, atom for atom. The surgeon uses the machine. A person wakes up at exactly the time you would otherwise have woken up, with a brain composed of exactly the same matter organized in exactly the same way.

What the machine does to the consciousness-generating part of your brain could hardly be described as "fairly minor alterations." I take it, therefore, that this is, on Dainton's view, a clear case of ceasing to exist (cf. Dainton, 2008, p. 328-329). Now to me, this seems like the wrong result. While my intuitions about the case are, admittedly, not as clear as my intuitions about Tonsillectomy 1 and 2, I do think I would exist after this procedure, too.

The phenomenal approach could of course allow for more seriously defective capacities than Dainton suggests. Even if it seems questionable whether personal identity becomes undermined when Dainton says it does, it could still be clear that it eventually will be as long as we keep inflicting more and more damage to the capacity for consciousness. So let's consider a final tonsillectomy case where the capacity for consciousness is completely dismantled:

Tonsillectomy 4 You are in hospital to have your tonsils removed. You are anesthetized; there is no consciousness throughout the procedure. The surgeon becomes convinced that a demon has taken residence in the consciousness-producing mechanism of your brain. In order to exorcise the demon, he removes this mechanism and dissolves it in a large tank of acid. He then uses the incredibly advanced machine, which down to the very last elementary particle perfectly reconstructs the consciousness-producing mechanism. A person wakes up at exactly the time you would otherwise have woken up, with a brain composed of exactly the same matter organized in exactly the same way.

I do find this scenario somewhat disturbing, but the more I think about it, the more I struggle to see how it could actually be bad for me. From an egoistic point of view, it does not seem to matter how much the particles that produce my experience when I am conscious move about when I am not conscious. If they spend a few minutes apart before joining up again by the time appropriate stimulation resumes, what's the harm?

For me, then, the whole idea that personal identity requires an unbroken chain of capacities for consciousness emerges as intuitively implausible. It seems that zombification does not have to be a problem as long as it is temporary. However, my intuitions about these matters may not be shared by everyone. At least for the sake of argument, I am willing to grant that when we increase the amount of temporary damage to the capacity for consciousness, we do eventually get to a point where most people no longer have any clear intuitive impression that identity is preserved.

This (possible) progress comes at a cost, however. For one thing, it is hard to imagine any remotely principled way of settling when a capacity for consciousness becomes too defective to sustain personal persistence. All we have to work with, it seems, are gradual scales of increasing destruction and disorganization that contain no particularly salient points. Dainton thus admits that we would have to posit have to accept a wide range of borderline cases, and takes this to count against the view (Dainton, 2008, p. 329).

A more serious problem is that it becomes harder to defend the thesis that a continuous capacity for consciousness is *sufficient* for survival. As we saw in Sect. 3,

even an intact capacity for consciousness does not seem intuitively sufficient to the same degree as actual consciousness. Defective capacities fare even worse.

In the conspiracy case, the present view will not only allow the aliens to destroy your body and non-phenomenal brain parts without undermining your persistence. They can also destroy parts of the neural basis of your capacity for consciousness, stopping only when they are approaching the minimum requirements for a defective capacity. They can then use neurons from (what used to be) Barack Obama's brain to repair the capacity. All that will be physically left of the pre-surgery person are some neural structures resembling a mechanism for producing consciousness. There is not even a possible stream of consciousness running from your present experience to the person waking up. This modification makes it even less clear that you will survive the procedure.

Fast gradual replacement of the neural basis for consciousness also seems more troubling if what is present throughout is only a defective capacity. Imagine that the aliens are less careful during the gradual phenomenal replacement, so that what is there during the 10 s is merely a series of structures quite similar to a working capacity for consciousness. There cannot be any experience present even if the anesthetics are omitted. It is now even harder to believe that the procedure is crucially different from sudden replacement of the physical basis for consciousness, which as you may recall will end your existence according to the phenomenal approach. This is of course not a problem for phenomenal accounts that forbid fast gradual replacement, but it will be for those that do not, such as Dainton's own view.

I find, then, that accounts of personal identity that are based on continuity in the capacity for consciousness are subject to conflicting intuitive pressures that are difficult to reconcile. When thinking about whether such continuity is necessary, we are pushed towards a conception of capacities for consciousness that is very lenient, perhaps even more lenient then Dainton's proposal. When thinking about whether such continuity is sufficient, it seems that the stricter we are, the better. Thus, to the extent that allowing capacities for consciousness to be defective helps the phenomenal approach avoid intuitive implausibility in the kind of scenarios discussed in this section, the gains will be offset by increased implausibility in other scenarios.

There are, in other words, significant difficulties with all the proposed versions of the thesis that continuity in the capacity for consciousness is necessary for personal persistence.

5 Alternative Versions of The Phenomenal Approach

In this section, I will look at some ways of developing the phenomenal approach that would allow us to avoid at least some of the problems encountered so far. I will argue that they lead to other problems that are, on the face of it, even more serious. That is not to deny that they are worth exploring, or that that there might turn out to be good ways of dealing with the problems. All I want to show is that they clearly do not offer an easy way out for the phenomenal approach.

5.1 Sameness of Matter in The Physical Basis for Consciousness

A natural response to the problems discussed in the previous section would be to come up with a phenomenal view where even a complete break in the capacity for consciousness, as in Tonsillectomy 4, is survivable. For this strategy to work, there must of course be some other kind of phenomenal continuity that could plausibly be thought to get us through such breaks. The only potential phenomenal basis for the survivability of Tonsillectomy 4 seems to be the fact that the post-surgery experience will be produced by the same physical matter as the pre-surgery experiences. What we end up with, then, is the idea that sameness of matter in the physical basis for consciousness grounds personal identity.[8]

This idea can be developed in different ways. One possibility is to hold that you persist as long as the relevant physical matter persists. On this view, you continuously exist throughout Tonsillectomy 4, albeit at some point completely dissolved in a tank of acid. The notion that you could hang around in such a dissolved state seems more than a little odd, so perhaps it is better to say that the case involves a temporal gap in your existence. The continuity of the physical matter does not ensure that you exist throughout Tonsillectomy 4, but it does ensure that you exist at the end.

If we opt for this alternative, we can still take continuity in the capacity for consciousness to be necessary for persisting *during* a period of consciousness, even if it is not necessary for existing after it. However, it is not clear that we need anything to play this theoretical role. We get a simpler account if we jettison all talk about capacities for consciousness and instead say that there is a gap in your existence whenever the relevant matter temporarily ceases to produce consciousness.

Though this way of thinking about personal persistence might be a bit unorthodox, I do, for what it's worth, find its way of dealing with breaks in consciousness quite appealing. It does not, to be sure, avoid the issues discussed in Sect. 3. Another, and presumably more serious problem, is that it rules out long-term persistence for all actual human persons.

If sameness of matter in the brain structures producing consciousness is necessary for personal identity, any kind of extensive replacement of this matter will undermine it. And while disallowing very quick gradual replacement arguably has some intuitive plausibility, disallowing slow gradual replacement is a different story. This kind of replacement actually takes place in human brains, partly on a neuronal level but mostly on a molecular level. It seems safe to assume that most of the matter underlying your experience of reading this sentence will have left your brain in a few months.[9] The present view thus implies that there is no personal persistence throughout human lives, at least not of the all-or-nothing kind that we usually talk about. Short-term egoistic reasons could remain more or less intact, but long-term egoistic reasons would be undermined, or at least much weaker than we have assumed. I

[8] In the debate on personal identity, philosophers tend to assume that it is clear what it means for the same physical matter to exist over time. But it should be noted that given the close connection between matter and energy, as well as the apparent weirdness and instability of the quantum world, this may not be clear at all.

[9] In fact, most proteins in the brain seem to have lifetimes of no more than a few weeks (see e.g. Fornasiero et al. 2018).

suspect that very few people will think this is a price worth paying for solving the problems we have found with the alternative phenomenal accounts.

It might perhaps seem that we could avoid these radical implications by making an exception for slow gradual replacement. Survival would then consist in the same matter or gradually slowly replaced matter generating consciousness. The problem is that these two alternatives can conflict. Let's say that aliens collect the matter in your present experiential mechanism when it is replaced throughout the next few years. They then put it back together so as to form an experiential mechanism and embed it in an (otherwise) new brain and body. Since the same matter is once more producing consciousness, you will exist in the new body. (You have in fact just undergone a more drawn out version of tonsillectomy 4.) But you'll also exist in your original body in virtue of the experiential mechanism that has undergone slow gradual material replacement. This is, at least on the face of it, an unacceptable result.

5.2 An Indivisible Basis for Consciousness

My objections to the various versions of the phenomenal approach examined so far all rest on the assumption that the basis for consciousness consists of numerable small parts that can be gradually replaced or dismantled and put back together. As long as we take the basis for consciousness to be physical, this assumption seems very safe. But if we are willing to postulate a non-physical basis for consciousness, it is no longer clear that we ought make it. In fact, most substance dualists reject it; they take the phenomenal substance, or soul, to be simple and indivisible (see e.g. Foster, 1991, p. 221-222).

If consciousness is generated by a soul of this kind, there will, first of all, be no puzzles about quick gradual replacement of the basis for consciousness of the kind discussed in Sect. 3.2. Either the soul remains completely intact and you continue to exist or it is completely and suddenly destroyed and you cease to exist.

We can perhaps also reasonably take an indivisible soul to be essentially capable of being conscious, in which case we can safely ignore the questions about temporary breaks in the capacity for consciousness from the previous section. But even if we have to grant that such breaks are possible, we can now offer a straightforward explanation how we can survive them: the consciousness after the break is produced by the same soul as the consciousness before it. Unlike for the corresponding matter-based view discussed above, slow gradual replacement of the basis for consciousness will not be an issue.

Substance dualism is of course not generally considered a particularly promising view, and with the exception of Foster (1991), defenders of the phenomenal approach have shown no inclination to rely on the notion of an indivisible soul. So while I think this possible way out is worth noting, I also think it is clear that it won't be an easy one.

5.3 Continuity of Actual Consciousness

Finally, supporters of the phenomenal approach can avoid all the issues I have raised by simply giving up on the idea that we survive breaks in the stream of consciousness

and instead basing their accounts on continuity of actual experience. They would then, it seems, have to admit that we are even more transient beings than on the view discussed in Sect. 5.1; undergoing an ordinary tonsillectomy, or just a period of dreamless sleep, suffices to end a person's existence. Galen Strawson (2008) defends a phenomenal account of this kind. He provides some intriguing arguments for it, which I will neither present nor evaluate here. For my purposes, it is enough to point out that, as Strawson willingly admits, this is a highly revisionist way of thinking about personal identity.[10]

According to Johan E. Gustafsson, however, sticking to actual consciousness does not have to rule out long-term personal persistence. Gustafsson argues that streams of consciousness can continue across ordinary periods of unconsciousness. More precisely, he thinks that the last experience before consciousness stops and the first experience after it resumes can be co-conscious in Dainton's sense: they are experienced together as if the period of unconsciousness hadn't occurred at all (Gustafsson, 2011).

I suspect that Gustafsson is right in thinking that it would be possible for this kind of phenomenal continuity not to be broken by dreamless sleep or anesthesia. But is this what actually happens? When arguing that it is, Gustafsson exclusively relies on a very small number of introspective reports (Gustafsson, 2011, p. 292-293). Given that philosophers have generally taken it for granted that streams of consciousness are broken when there is a period of unconsciousness, this is hardly an impressive body of evidence. One might suspect that most people would provide conflicting reports (cf. Strawson, 2009, Sect. 5.4). But even if they would, that may not suffice to discredit Gustafsson's hypothesis. When you look back at a night containing dreamless sleep, what appears to be breaks in the stream of consciousness could just be holes in your memory. Even if you conduct your introspection right after being woken up from dreamless sleep, the explanation for why it seems that the stream of consciousness just started anew could be that you have already forgotten your previous dream, and that the lingering presence of the last part of the dream in the first part of the awake experience is too subtle for you to notice. We are, after all, typically pretty confused when we wake up from dreamless sleep, and accurate introspection can elude us even at the best of times.

The idea that human experiences are phenomenally connected across ordinary periods of unconsciousness is, then, difficult to rule out. On the other hand, it is not clear that there is much positive reason to accept it. In particular, it is not clear that it is more plausible than the various problematic claims we are forced to make on other versions of the phenomenal approach. More importantly, even it is correct, I do not think it allows us to formulate an attractive account of personal identity based on actual consciousness. We are still stuck with the implication that you will cease to exist *if* there is a break in your stream of consciousness. A single experience that is not co-conscious with the next experience produced by your brain would result in the creation of a numerically new person. This is very hard to believe.

[10] Strawson actually takes his view to have even more dramatic implications than suggested here. He thinks that breaks in the stream of consciousness are much more frequent than is typically assumed, and that persons are unlikely to last even for a second (Strawson 2009, ch. 5). (Strictly speaking, Strawson presents his view as an account of *selves*, not of personal identity, but what he aims to capture seems at least closely related to what phenomenal accounts of personal identity aim to capture.)

Perhaps counterintuitive implications aren't quite as damning when they concern hypothetical rather than actual cases. However, accounts of personal identity are routinely, and reasonably, evaluated in large part on how well they deal with counterfactuals. They are not just expected to come up with something that is there throughout ordinary human lives, but something that it seems we could not exist without. Now we should not draw too firm conclusions based on outlandish scenarios about which we have unclear and differing intuitions (which admittedly may include some of those offered in this paper). But here, the scenario is very familiar; it is simply what most people believe actually happens in dreamless sleep. And there is of course overwhelming agreement that dreamless sleep is survivable.

In sum, I think there is no getting around the fact that an account of personal persistence based on actual consciousness would be a very radical solution to the problems facing the phenomenal approach.

6 Conclusion

Although I haven't examined every possible version, I do think we have good reason to believe that accepting the phenomenal approach will at some point force us to make claims that are, at least when considered in isolation, quite implausible. As I see it, the basic problem is that the kind of continuity that primarily elicits intuitions in favor of the approach, i.e. the one found in actual streams of consciousness, is so easily broken. Phenomenal accounts either have to admit that human persons are surprisingly short-lived, or at least surprisingly fragile, or they have to rely forms of continuity that are more robust, but do not on reflection seem particularly relevant to personal identity.

Should this lead us to give up the phenomenal approach and focus on non-phenomenal views instead? The answer depends, first of all, on how much the phenomenal approach has going for it in the first place. Some readers might – not unreasonably, I think – find the considerations in its favor sketched in Sect. 2 to be strong enough to outweigh the considerations against it offered in the later sections. Second, it depends on whether the alternatives are any better. Now non-phenomenal views do not generally suffer from the problem that their paradigmatic form of continuity is frequently broken in ordinary human lives. But more broadly speaking, all views on personal identity, while perhaps initially plausible, tend to creak and arguably break when we push them on precisely what the relevant continuity is and whether it is always impossible to continue to exist without it or to cease to exist with it. It is certainly not obvious that the phenomenal approach does (much) worse in this respect than its rivals.

I will not attempt any overall assessment of the phenomenal approach. For all I have said in this paper, it could still be the best we have. If anything, this would make the difficulties we have encountered even more pressing. We would have to think about them not in order to figure out *whether* personal identity depends on consciousness, but in order to figure out *how* it does.

Acknowledgements I would like to thank Fredrik Nyseth and the anonymous reviewers for helpful comments.

Funding Open access funding provided by UiT The Arctic University of Norway (incl University Hospital of North Norway)

Declarations

Conflict of interest I have no conflicts of interest to disclose.

References

Crisp, R. (2006). *Reasons and the good*. Oxford: Oxford University Press.
Dainton, B. (2008). *The phenomenal self*. Oxford: Oxford University Press.
Dainton, B., & Bayne, T. (2005). Consciousness as a guide to personal persistence. *Australasian Journal of Philosophy*, *83*, 549–571.
Duncan, M. (2015). I think, therefore I persist. *Australasian Journal of Philosophy*, *93*, 740–756.
Duncan, M. (2020). A new argument for the phenomenal approach to personal persistence. *Philosophical Studies*, *177*, 2031–2049.
Fornasiero, E. F., Mandad, S., Wildhagen, H. (2018). "Precisely measured protein lifetimes in the mouse brain reveal differences across tissues and subcellular fractions." *Nature Communications* 9.
Foster, J. (1991). *The immaterial self*. London and New York: Routledge.
Gottlieb, J. (2021). Persistence without Personhood: A New Model? *The Philosophical Quarterly*, *72*, 346–364.
Gustafsson, J. E. (2011). Phenomenal continuity and the Bridge Problem. *Philosophia*, *39*, 289–296.
McMahan, J. (2002). *The Ethics of Killing: Problems at the margins of life*. New York: Oxford University Press.
Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
Strawson, G. (2009). *Selves*. New York: Oxford University Press.
Unger, P. (1990). *Identity, consciousness and value*. Oxford: Oxford University Press.
Wright, J. (2006). Personal identity and consciousness. *Iyyun: The Jerusalem Philosophical Quarterly*, *55*, 235–263.