Scientific Research Publishing

Open Journal of Philosophy, 2025, 15(1), 126-158
https://www.scirp.org/journal/ojpp
ISSN Online: 2163-9442
ISSN Print: 2163-9434

Parfit, Self and Unity

Przemysław Paleczny

Institute of History, Faculty of Social Sciences, Opole University, Opole, Poland
Email: paleczny@protonmail.com

How to cite this paper: Paleczny, P. (2025). Parfit, Self and Unity. *Open Journal of Philosophy, 15,* 126-158.
https://doi.org/10.4236/ojpp.2025.151008

Received: November 24, 2024
Accepted: February 9, 2025
Published: February 12, 2025

Abstract

The paper presents an effort to connect two ideas that are underrated in Derek Parfit's analysis of David Wiggins' split-brain thought experiment. Firstly, it argues that the original person, whose brain is divided, survives as only one of the two human beings obtained by the division. For Parfit, it seems hopelessly hard to determine which one of the two is the initial person. The second idea is the unity of consciousness. For Parfit, the best-known approach to the idea is Cartesianism, where the Cartesian Ego provides the unity in question. Therefore, he poses a dilemma between the Cartesian Ego concept and his view that consciousness is like a river and unity plays no role in survival. The paper explores the notion that, firstly, only one person could survive the split-brain experiment proposed by Wiggins; secondly, that the unity of consciousness plays a fundamental role in personal identity over time; and thirdly, that it does not engage with the Cartesian Ego concept, but instead, with the concept of the epiphenomenal self.

Keywords

Epiphenomenalism, Personal Identity, Self, Split Brain, Unity of Consciousness

1. The Trilemma of the Split-Brain Thought Experiment

The brain of patient P_{Zero} is divided into two hemispheres and two symmetrical halves, and then transplanted into two different bodies. Assuming the brain's neuroplasticity, the two halves compensate for each other's functions, giving rise to two distinct personalities, P_1 and P_2. Both are psychological duplicates of P_{Zero}. Does P_{Zero} survive the operation or not? Derek Parfit offers three possible answers to the original split-brain thought experiment, initially proposed by David Wiggins (Wiggins, 1967: p. 50). Parfit's answers, which I have named for clarity, are detailed in his works (Parfit, 1971: p. 5; Parfit, 1984: p. 256).

The negation claim: P_{Zero} does not survive. Although it seems probable at face

value, the negation claim has some absurd consequences. If we do not divide the experimental brain but transplant it whole into a different body, the intuition accepted by the psychological view on personal identity (referred to as "the transplant intuition" in Olson, 1997) is that P_{Zero} goes with the brain. In short, transplantation of the entire brain transplants the person.

The split-brain experiment assumes that both halves take each other's functions and realise personalities identical to the initial one. Therefore, the transplantation of only one-half in the procedure should be like that of the entire brain, at least psychologically. Hence, the same transplantation intuition should lead us to conclude that the initial person of P_{Zero} goes with the half in question. This scenario amounts to the survival of P_{Zero}. However, the negation claim states that the successful transplantation of two halves ends with the death of P_{Zero}. One half is survival, but two halves are death. It seems like the survival of P_{Zero} depends on the unsuccessfulness of the transplantation of the second half. It seems to be absurd. However, in the literature, there are some defences to the negation claim (e.g., Shoemaker, 1970).

The conjunction claim: P_{Zero} survives as both persons, P_1 and P_2. In contrast to the previous answer, the conjunction claim seems absurd at face value. As Parfit writes (Parfit, 1971: p. 8, n. 8), if P_1 and P_2 fought with each other and P_1 killed P_2, how should we understand the situation? Could the act be a murder and a suicide at the same time? Or could it be a suicide that succeeds and does not succeed as well? The conjunction claim demands a deeply revisionary explanation because it is highly counterintuitive. However, some sophisticated defences of the claim are also in the literature (e.g., Lewis, 1976 or Noonan, 1989).

The disjunction claim: P_{Zero} survives as only one person, e.g., P_1 or P_2. The split-brain experiment assumes that P_1 and P_2 are psychologically indistinguishable from and continuous with P_{Zero}. Therefore, taking the psychological criterion, deciding which person should be identical to the initial one, looks impossible. It seems that both persons satisfy the criterion in the same ideal way. This line of thought is considered unpromising from the outset.

The principal aim of this work is to argue that Parfit underrates the disjunction claim and that there is a way to defend the claim within the psychological approach to personal identity. Therefore, for the sake of my reasoning, I will limit my inquiry to psychological accounts of the self.

I would like to sketch a view that could enable a defence of the disjunction claim and demonstrate that this neglected strategy hides valuable insights that could enrich the debate. Therefore, I will focus less on the discussion with proponents of the two other claims and more on presenting the view that could support the disjunction claim. However, in the last section, I will briefly outline how the inquiry's results relate to the former two answers.

2. The Split-Brain Paradox

The paradox presented below grapples with doubts about the disjunction claim.

It consists of three statements.

The first statement is a general thesis of the psychological approach to personal identity: the criterion of personal identity is psychological continuity. A person P_{Zero} in time t_{Zero} and P_{One} in t_{One} are identical if and only if there is a psychological continuity between the total mental state of P_{Zero} in time t_{Zero} and the total mental state of P_{One} in time t_{One}. The total mental state is the state of a whole mental system, including all the system's mental states.

According to Parfit (1984: p. 206): "Psychological connectedness is the holding of particular direct psychological connections." Meanwhile: "Psychological continuity is the holding of overlapping chains of strong connectedness." The crucial difference is that continuity is transitive while connectedness is not. Personal identity, like other identities, is also transitive. In Parfit's view, personal identity takes place when relation R ("psychological connectedness and/or continuity") does, and the relation is "holding *uniquely*—holding between one present person and *only one* future person" (Parfit, 1984: p. 263; original emphasis). In short: "the fact of personal identity just consists in the holding of relation R, when it takes a non-branching form" (Parfit, 1984: p. 263; therefore, the split-brain thought experiment concludes with the survival of P_{Zero} but without preserving P_{Zero}'s identity; this is because personal identity does not matter for survival in Parfit's view).

Psychological connectedness is a concept introduced only to explain personal identity over time when it is noncontroversial, specifically when the identity takes the standard non-branching form. If the disjunction claim is accepted and P_{Zero} is identical to P_1 and not to P_2, P_{Zero} is psychologically connected only with P_1. However, psychological continuity may be enough to claim the fact of identity, specifically if the disjunction claim is at work. Therefore, connectedness could be a redundant concept. Not only connectedness but also any non-branching condition would probably be needless. The disjunction claim posits that branching is no threat to personal identity since the initial person survives as only one of the resulting individuals.

The case of Methuzalem—an "everlasting body" that "gradually changes in appearance" (Parfit, 1971: p. 23)—could be helpful here. After thousands of years, Methuzalem's total mental state in t_2 has nothing in common with the state in t_1. There is psychological continuity but no connectedness. Since Parfit proposes to use the word "I" for "the greatest degree of psychological connectedness" (Parfit, 1971: p. 25), Methuzalem can think about his "past self" (as Parfit proposes in Parfit, 1971: p. 22 or Parfit, 1984: p. 304). But the proposition must mean that Methuzalem is a different person in t_1 than in t_2. Therefore, if he were offered money in t_1 for severe torture in t_2 (as in Bernard Williams's experiment (Williams, 1970)), it would be money for torturing someone else. In t_1, Methuzalem does not need to be afraid of pain in t_2 because the states have no psychological connectedness. Maybe Methuzalem could have reasons for caring for the person in t_2 (after all, there is survival between the states, at least in Parfit's terms), but it would be caring for another person. There must be no fact of personal identity if

the interpretation of Parfit's work is correct. In turn, if psychological continuity is enough, and the disjunction claim permits this assumption, then given the continuity between Methuzalem's states in t_1 and t_2, he should indeed be terrified of severe pain in t_2. He would be the same person.

The second statement of the paradox is that P_1 and P_2 are psychologically indistinguishable. However, it is possible that both P_1 and P_2 are indistinguishable in their psychological profiles, but only one of them is identical to P_{Zero}. Moreover, they could be psychologically continuous with the initial person, and the disjunction claim could still be valid. We assert the fact of identity within the verification procedure. It is necessary to differentiate the objective fact of personal identity from a conclusion taken from the examination (see Swinburne, 1974 for an argument of a different kind that supports such a claim). Therefore, the idea of the psychological continuity test is introduced here.

It is an idealised test used to verify whether two persons at two points in time are psychologically continuous or not. A computer conducts the test by comparing psychological profiles of persons at a given time, calculates the possibility of psychological continuity, and based on that, provides a percentage verdict of the probability of personal identity. Consequently, the claim that P_1 and P_2 are psychologically indistinguishable means no more than that P_1 and P_2 have passed the test with P_{Zero} equally well, say nearly 100% compatibility.

The test may seem to make Methuzalem's case much worse if the claim is that Methuzalem in t_2 is the same person after thousands of years since t_1 because his psychological continuity is enough for it. His psychology in t_2 has nothing in common with that in t_1. It has to mean that the test should provide a near 0% probability verdict. How could he be the same person in t_2 as I claimed?

The test should not be treated as one taken from God's perspective; verification of the fact does not equal the fact itself. The thought experiment *assumes* that Methuzalem in t_2 is psychologically continuous with him in t_1. It accepts continuity as a fact from the start. But we know why the test must fail. Namely, the computer has no data about the mental life of Methuzalem in the stages between t_1 and t_2. If it monitored Methuzalem's lifespan the whole time, it would claim a probability near 100% because it would have no problem recognising psychological continuity. The ideal test introduced here is infallible, provided that crucial data are accessible.

The third claim of the paradox is the disjunction claim stating that P_{Zero} survives as only one person, namely, P_1 or P_2. The Methuzalem case presented before shows that we could be wrong in our conclusion based on the psychological identity test alone. But we can also know why we would be wrong. The same could be the case here. Even though P_1 and P_2 pass the test equally well, something could be missing here. The concepts of the self and the unity of experience presented below are supposed to fill the gap.

In sum, the split-brain paradox that assumes the disjunction claim looks like this:

1) The criterion of personal identity is psychological continuity. If someone in t_2 passes the psychological continuity test with someone in t_1, s/he is considered the same person.

2) P_1 and P_2 are psychologically indistinguishable. This means that P_1 and P_2 in t_2 pass the test with P_{Zero} in t_1 equally well.

3) P_{Zero} survives as only one person, e.g., P_1 or P_2. Only one of them is considered identical to P_{Zero}.

Assuming the psychological criterion of personal identity in the split-brain experiment, P_1 and P_2 are psychologically indistinguishable—they have passed the test equally well. The conclusion should be that they are both psychologically continuous with P_{Zero}. Because continuity is the criterion, they should be considered identical to the initial person. This seems to stand in contradiction to the claim that only one of them, P_1 or P_2, is, in fact, identical to P_{Zero}. Nevertheless, this work aims to show that the three statements could be non-contradictory.

3. The Physics Exam Experiment

Parfit often seems to treat the so-called stream of consciousness as something like a river. Water flows inertly in a riverbed, and if the riverbed branches, the water easily adapts to it. He even writes: "We can come to believe that a person's mental history need not be like a canal, with only one channel, but could be like a river, occasionally having separate streams" (Parfit, 1984: p. 247). Maybe this is why he imagines a division of consciousness as a simple experiment (Parfit, 1984: p. 288).

During a physics exam, Parfit divides his mind so that each part works with a different calculation method. Then, he reunites them and writes down the better result. To avoid problems with the physical realisation of the experiment, Parfit assumes that he is among the minority of people whose hemispheres had not specialised while growing up, and both can perform the same functions equally well. He also has a science-fiction device to divide his brain's work and reunite it whenever he wants.

Parfit dismisses the unity of consciousness, citing the case of patients who have undergone the commissurotomy operation as his rationale. The latter is sometimes done to prevent epileptic seizures by severing the corpus callosum, the fibres connecting both brain hemispheres. In one such experiment (Sperry, 1974: p. 9), the word "key ring" was shown to participants by projecting the word "key" onto the right sight of both retinas and the word "ring" onto the left sight. The latter feeds to the left hemisphere's cortex, while the former goes to the right hemisphere's cortex. The left hemisphere controls speech, while the right controls the left hand. Patients were asked to point with their left hand to the thing referred to by the flashed word. They pointed to a key, not a ring. When asked what the thing was, they said it was a ring and denied that it was a key. It was as if they saw different things with different hemispheres.

In view of such experiments, Parfit takes it as a fact that it is possible to divide the stream of consciousness. He writes:

> "It might be objected that my description [of the physics exam experiment] ignores 'the necessary unity of consciousness'. But I have not ignored this alleged necessity. I have denied it. What is a fact must be possible. And it is a fact that people with disconnected hemispheres have two separate streams of consciousness—two series of thoughts and experiences, in having each of which they are unaware of having the other" (Parfit, 1984: p. 247).

Nevertheless, the physics exam experiment is unimaginable, mainly because it engages the first-person perspective. The crucial thought in Parfit's reasoning is that the two branches are unaware of each other. If any mental connection between the streams is preserved, the argument fails because it makes the experiment something different than the commissurotomy case. If there were any connection, the experiment would deal with something like an overdeveloped (thanks to the fictional device) attention shifting (or division)—not the expected division of consciousness. But if one branch is unaware of the other, they can't decide on the reunification rationally and consciously.

Since the first-person perspective is necessary here, I could imagine that I am Parfit in this experiment. When I decided to divide my consciousness stream, I could no longer be aware of both branches. Such awareness would establish the connection that makes the experiment different from the commissurotomy case. I could go with one of the streams and do the calculations with one of the methods (I do not know how I could decide on which stream; maybe I could choose it in advance with my science fiction device). Since I am within one of the branches, I am unaware of not only what the other branch is doing; I have to be unaware that there is even another branch. The commissurotomy patients are unaware that they give different answers, and Parfit wants me to be like them. Since I am unaware that another branch exists, how could I know I should do the reunification? I do not even know yet that there is anything to unify. Therefore, I cannot imagine consistently that I reunify my divided stream of consciousness.

Nevertheless, the physics exam is imaginable from the third-person perspective. It seems logically possible to divide the mental machinery, force the two parts to work separately and reunify it later. However, the person who decides must be someone else rather than the experimental agent. The one deciding about the fusion must know that there are two streams. If the two streams are unaware of each other, only the external observer can grasp the situation correctly. Therefore, I can imagine from a third-person perspective that I observe what Parfit is doing in his exam. But I have the remote control at my disposal. By pushing the button, I divide Parfit's mind and make it possible for him to do the calculations with different methods simultaneously. I can make Parfit like one of the commissurotomy patients for a few minutes. Next, in the same way, I reunify them. I do what Parfit cannot do in this situation by himself.

Given the third-person possibility, it may be tempting to ignore the first-person worries. But the latter depends on the former. The objectively considered mental system makes it impossible to imagine the situation subjectively. In the absence of

accounting for this impossibility, the question is ignored rather than denied, as Parfit has it. I cannot imagine simultaneously being like a commissurotomy patient and reunifying my divided stream of consciousness in Parfit's experiment. The impossibility shows that the stream is not like a river. It is not a chaos of loosely appearing mental states that can be freely arranged in one or two streams. In contrast, I shall explore the idea of the unity of consciousness.

4. The Unity of Experience

Unity is a relation between elements that makes them interdependent. When one element changes, it causes changes in other elements and/or the structure of the elements bound by the relation. However, such a definition may, at best, be an empty form without content. In the case of the unity of consciousness, it is extremely difficult to determine what content should fill this form. There is a broad discussion about what these elements are, whether there are any elements at all, how they are unified, what the structure of the unity is, what logical properties the unity has, and many other related questions. Since this paper does not have space to discuss even the main issues in detail, I will present a few assumptions here rather than arguments.

The unity of consciousness has many aspects that are differently classified in the literature. Concepts of unity are often based on one aspect that is considered more crucial than others (e.g., its representational aspect or the aspect of belonging to a subject; see Bayne, 2010: Ch. 1.3). The aspect I am the most interested in here is the one that presents the greatest difficulties in defining it: phenomenal unity.

The concept of phenomenal consciousness is derived from Thomas Nagel (1974). He explained that, for phenomenally conscious beings, there is the "what it is like" quality of their experience. The key feature of the quality is its exclusive first-person character. People know what it is like to sit with a friend in a pleasant pub while listening to their favourite music. On the contrary, as Nagel argues, people do not know what it is like to navigate by echolocation while flying through a cave in complete darkness, which is a common experience for bats. People lack that "what it is like" quality because we cannot take the first-person view of a bat.

The challenge of defining phenomenal consciousness and, consequently phenomenal unity, lies in their entirely subjective and qualitative character. As Ned Block (1995: p. 166) puts it: "I cannot define [phenomenal consciousness] in any remotely noncircular way. (…) But the best one can do for [phenomenal consciousness] is in some respects worse than for many other things because really all one can do is *point* to the phenomenon" (Block, 1995: p. 166). Properties that could be pointed out are experiential ones: "what it is like" to feel, smell, or hear.

At the same time, it seems there is no consciousness without the phenomenal one. As David J. Chalmers writes: "a mental state is conscious if there is something it is like to be in that mental state. To put it another way, we can say that a mental state is conscious if it has a *qualitative feel*—an associated quality of experience.

(…) The problem of explaining these phenomenal qualities is just the problem of explaining consciousness. This is the really hard part of the mind-body problem" (1996: p. 5). The problem of "why cognitive functioning is accompanied by conscious [phenomenal] experience" (Chalmers, 1996: p. 26) is not the focus of this paper. For my purposes, I will simply assume that the phenomenal properties of consciousness exist and will not engage here in the vast debate on their nature.

The notion of phenomenal unity is derived from the notion of phenomenal consciousness. Two states are phenomenally unified if there is a single "what it is like" quality that encompasses both states together. Taken separately, there could be a "what it is like" quality of seeing a friend, a different one of sitting in a pub, and a different one of listening to music. However, in one experience, there is one unified "what it is like" quality of sitting with a friend in a pub while listening to music. This means that unity makes the three elements interdependent. Therefore, "what it is like" to see a friend on the way to work is not the same quality as "what it is like" to see a friend while listening to music in a pleasant pub. The three elements are joined into one phenomenal experience with a single "what it is like" quality.

Block also famously distinguished phenomenal consciousness from access consciousness. Mental states are access conscious if available "in reasoning, reporting, and rationally guiding action" (Block, 1995: p. 160). The notion of access unity is derived from the notion of access consciousness. Two states are access unified if they are available together in "reasoning, reporting, and rationally guiding action". If the experiences of a friend and a piece of music are access unified for me, I can think that my friend does not like the music, so I can ask if they would like me to change the music in a jukebox or switch to another pub, and then do it, thereby rationally guiding my action. The elements are interdependent, since what I think about the music depends on who sits with me at the table, influencing my decision to change the music.

Block's distinction can help illustrate a crucial difference between phenomenal and non-phenomenal aspects of experience. Non-phenomenal aspects involve what psychology or cognitive science explores: properties that could be defined in terms of their functional and causal roles. Therefore, whether the commissurotomy patient meets access unity demands is experimentally verifiable. The patient reports that they see a ring and denies that they see a key, but point to a key, not a ring. There is no unity in reporting and guiding action.

On the other hand, whether the patient experiences phenomenal unity remains a mystery (it may be that they have a single phenomenally unified experience that is not cognitively available as one experience; see Bayne & Chalmers, 2003). There is no way to verify anything about the patient's phenomenal consciousness. In other words, we cannot know "what it is like" to be the patient in the laboratory, just as we cannot know "what it is like" to be a bat.

There is another assumption that I must make to maintain coherence throughout my inquiries: everything mental is phenomenal. I mean precisely everything,

including unconscious mental states (which affect the phenomenology of conscious ones, e.g., we can today like something that we did not like yesterday, without knowing why) or thoughts (e.g., we do not like to think some thoughts). Bayne (2010: p. 7) proposes a somewhat less radical version: "I will take 'phenomenal consciousness' to be pleonastic: *all* consciousness is phenomenal consciousness." Other proponents of this stance include Flanagan (1992) or Strawson (1994), while opponents include Carruthers (2005) or Tye & Wright (2011).

5. The Self as the Unifier of Experience

Why does the unity of experience occur? There is a standard view that Parfit expresses this way: "On this view, we should explain the unity of a person's consciousness, at any time, by ascribing different experiences to this person, or 'subject of experiences'. What unites these different experiences is that they are being had by the same person" (Parfit, 1984: p. 249). A similar view of the concept of the self is presented in the literature. Tim Bayne, trying to grasp a few vital features of the self that any account should include, writes: "The self or subject of experience provides one form of unity to be found in consciousness. My conscious states possess a certain kind of unity insofar as they are all mine" (Bayne, 2010: p. 9). There is an idea that it is the self that provides the unity of experience. I shall take a closer look at this idea.

A minimal but substantial assumption I employ here is that the self, however understood, makes something a person. I do not mean that there is a causal relation between a self and something, where the self causes the thing to become a person. I mean only that if the thing has a self, it is a person. Since the psychological criterion of personal identity is at work, the self should also be considered a mental phenomenon. Psychological continuity could depend on physical features, entities or events. Still, as long as the psychological view of personal identity is at work, personal identity should be explained in psychological terms. Therefore, in the view discussed here, the mental self is considered the unifier of experience.

The idea that the self provides mental unity makes any split-brain case even more puzzling than it is usually conceived. Nagel writes:

> "[It is] difficult to conceive what it is like to *be* one of these people [who underwent the commissurotomy]. (…) lack of interaction in the domain of visual experience and conscious intention threatens assumptions about the unity of consciousness, which are basic to our understanding of another individual as a person" (Nagel, 1971: p. 407).

It seems that the quote presupposes that it is the self that puts everything together. Having experiences of seeing e_1 simultaneously with e_2 by a single subject is supposed to constitute one unified $e_1 + e_2$ experience. But the experimental patient, on one occasion, admits to seeing e_1 but not seeing e_2, while on the other hand, s/he admits to seeing e_2 but not seeing e_1. There seems to be nothing like $e_1 + e_2$. However, the two experiences are separately unified since such a patient con-

sistently answers the experimenter's questions about e_1 and e_2 if taken independently (e.g., they do not say that they see the key and then that they do not see it; what one hemisphere "experiences" remains consistent). The S_1 self must provide the unity of the e_1 experience, while the S_2 self must provide the unity of the e_2 experience. According to the claim about the unity-providing self, if S_1 and S_2 were the same self, there should be one $e_1 + e_2$ experience, but this is not the case. Therefore, two selves seem responsible for two unifications and the two divergent answers. In Nagel's analysis, the suspicion that the patient has two selves also appears to be a serious possibility. We think we are dealing with one person while two seem to be in a single body.

Here is another example. Parfit holds that his physics exam experiment proves that the concept of a person and that of a subject of experience do not coincide. In his case, one person, Parfit himself, but two subjects of experience calculate with different methods. He writes:

> "We must now abandon the claim that 'the subject of experiences' is the person. On our view, I am a subject of experiences. These [two branches] are not the same subject of experiences, so they cannot both be me. Since it is unlikely that I am one of the two, given the similarity of my two streams of consciousness, we should probably conclude that I am neither of these two subjects of experiences" (Parfit, 1984: p. 249).

In the same way, the commissurotomy patient may seem to be a single person but a double subject of experience. Such an interpretation is impossible if the self is what makes something a person and, at the same time, what unifies experience. Parfit seems to double the subject in his interpretation because of two unified streams of consciousness. He probably unintentionally takes the two subjects as unifiers of the two streams. But if the self is supposed to be the unifier of experience, the two subjects enjoy two different selves. Therefore, if the self is also understood as what makes something a person, there must be two persons. But this is at odds with Parfit's view that there is only one person, while there are two subjects of experience. This is not a straightforward refutation of his interpretation. Rather, this proves that his interpretation is incompatible with the assumption that the self makes something a person and unifies consciousness at the same time.

Further problems appear for all of the three claims of the split-brain trilemma formulated before.

The negation claim: P_{Zero} does not survive the experiment. It has to mean that P_{Zero}'s self does not survive as well. Otherwise, P_{Zero} would survive, contrary to the negation claim, for the self is what makes P_{Zero} a person. Thus, the person's self must cease to exist. If this is the case, nothing could unify any experience, and there is no experience to be unified. This way or another, there is no psychology. But even if there could be some remnant psychology of P_{Zero} saved somehow in the two halves of the brain, the selves of P_1 and P_2 are also supposed to be the

unifiers. How could they arise from something they should unify, namely, the remnant psychology? It looks like something akin to Russell's paradox: the unifier would be one of the elements that it unifies. The self of P_{Zero} must cease to exist to make him or her dead. But there must be two selves of P_1 and P_2. Because they must be the unifiers, they cannot be created from remnant psychology. It would make the paradox. Since P_{Zero}'s self disappears, the other two seem to come from nothing. It appears to be an odd conclusion.

The conjunction claim: P_{Zero} survives as both P_1 and P_2. It may have at least two different interpretations. The first is that the single self of P_{Zero} is somehow doubled, and not divided, not derived one from another, but doubled. As in the previous case, the selves of P_1 and P_2 cannot be created from the remnant psychology of P_{Zero} because the unifier created from what it should unify creates the paradox. Therefore, since the survival of P_{Zero} depends on the survival of the self, and the latter cannot arise from the remnant psychology, both selves, P_1 and P_2, must come from the original self alone, independently of the rest of P_{Zero}'s psychology. But P_{Zero}'s self, understood as a mental phenomenon independent of the rest of psychology, cannot be just divided because it is hardly conceivable that this or any other mental phenomenon (like pain, desire or fear) could be divided. According to the conjunction claim, P_{Zero} survives as both persons. It must mean that the self of P_1 is the self of P_{Zero}, but the self of P_2 is also the self of P_{Zero}. If P_{Zero} had not been both persons before the operation, s/he was doubled. One person miraculously became two. This begs an explanation as to how the miracle is possible.

Another possible interpretation of the conjunction claim could be that P_{Zero} has two selves from the start. However, the experiment assumes P_{Zero} is a single person before the operation. The assumption could be a mistake, but if so, it should be justifiable before the operation, while such a theory allows it only after the surgery. It looks like history is being rewritten through surgical operation (David Lewis, 1976 even accepts something like this). If it were at all possible, any normal human being before undergoing the supposed operation would be like Schrödinger's cat: a person could have two selves, but nobody—not even the individual in question—would know it before the surgery. If this scenario were not absurd enough, one might mistakenly believe that the operation creates two selves, when in fact, it would actually be doing the patient a favour by separating two selves that were conjoined since birth. The situation seems crazy (however, there is a way to think that we are all double subjects; see Schechter, 2018).

The disjunction claim: P_{Zero} survives as only one person, P_1 or P_2. Since only one could be identical to P_{Zero}, only one must have the original self, while the other does not. However, the previously discussed problem reappears. Since the self is the unifier of experience, it is hardly believable that the unifier is created from what it unifies; the self, which is not identical to the initial one, comes from nothing. And we do not know how to recognise which self is the original one.

To evade the Russell-like paradox, the self, as a unifier of experience, must be independent of what it unifies. If the self is a psychological phenomenon, it has a

unique and special status among other mental phenomena. It can unify them, while nothing else in the psychological domain can do it. In this respect, the self appears as an entity separated from the whole psychological domain. It must be outside the total mental state because the latter is a set of all mental phenomena unified into the total mental state. The unification is what makes a mental state a part of the set. It is supposed to be what makes them "my mental states". However, if the self is what unifies them, it must not be among other states that it unifies, so the self must not be within the total mental state.

Consequently, the ideal test for psychological identity introduced before would find no self of the kind described above in the history of a person. Instead, the test tracks only psychological continuity as defined before, which is the continuity between one total mental state at one time and another total mental state at another time. The self would not manifest in the psychological identity test. Therefore, we could not determine if P_1 or P_2 is identical to P_{Zero}.

Such a result appears to align closely with first-person phenomenology accounts. Nothing could be recognised as the unifier of experience in our phenomenological field. We have a unified experience, and that is all. The popular argument, starting at least with David Hume, is that nothing in our experience could be called the self, which may be true if the self is, by default, understood as the entity that unifies experience. Nothing in the content of the phenomenological field could be recognised as the entity being the source of the unity.

Such an entity is a suspicious one. If understood as a psychological phenomenon, it appears above and beyond any other mental state. Considering "the view that psychological unity is explained by ownership" (Parfit, 1984: p. 249), Parfit concludes that "on the best known version of this view, we are Cartesian Egos" (Parfit, 1984: p. 252). He makes us confront the dilemma: to accept the Cartesian Ego concept or that consciousness is like a river. Since there could be one person but two subjects of experience with two unified streams of consciousness—all in one body—the unity seems to play no role in personal identity.

To sum up, there are two problems associated with all the answers, and the idea of the self as the unifier of experience is the source of the troubles. One problem is the paradox of the unifier being what it unifies. If we want to accept the claim that the self is a purely psychological phenomenon or entity, it belongs to the mental domain of a person. But the problem appears if we also want to claim that the self provides the unity of the domain. Since the self is supposed to be a part of the mental domain, it should also be unified. But the self is also what provides the unity. As such, the self cannot be anything unified because the unifier cannot belong to what is unified. If everything in the mental domain of a person is unified under the person's self, and the self cannot be one of the mental states it unifies, the self cannot belong to this mental domain. The latter contradicts the initial claim that the self is a purely psychological phenomenon and, therefore, it belongs to the mental domain of a person. How could it be that the self belongs and does not belong to the mental domain at the same time?

Another problem is that the self cannot come from nothing. Since the self provides the unity of consciousness or experience, any mental state derived from the set unified under a single self—like in the split-brain thought experiment—is non-unified. Only the self could unify them, but the unifier—the self—cannot arise from what it unifies, namely, mental states. Therefore, if the self makes something a person, and there is another person to whom the derived states belong, there must be another self. However, the latter cannot come from the derived psychology. Therefore, the second self seems to come from nothing. Two selves come from nothing if P_{Zero} dies (the negation claim), or there is one person that comes from nothing if P_{Zero} survives as only one person after the division (the disjunction claim). Alternatively, the self of P_{Zero} is miraculously doubled if P_{Zero} survives as both persons after the division (the conjunction claim). In the last case, the initial self cannot be just divided because it is hard to understand how the mental phenomenon could be divided, especially if it cannot be created from the remnant psychology that the self is supposed to unify.

Eventually, there is an alternative way for the psychological view of personal identity. Namely, the thesis that the self is a brain (see Nagel, 1986: Ch. III and Unger, 1990). Such a unifier is not a mental phenomenon but a physical entity. Consequently, the paradox of the unifier being something that it unifies is avoided. By the same token, the paradox of the self coming from nothing is avoided as well since a divided brain is also a divided self. However, such an approach does not solve the split-brain trilemma. The latter still requires an explanation in psychological terms. Since it is supposed that after the division, both halves are alive and support personalities, it would be hard to explain why P_{Zero} dies (the negation claim) or survives with only one half (the disjunction claim). I will elaborate further on this account in the next section.

Anyway, this work aims to explore the possibility of the disjunction solution. Parfit's dilemma of the Cartesian Ego and the river-like consciousness views neglect a solution discussed below. The latter rejects the idea that the self provides the unity of consciousness. Therefore, it leaves the Cartesian Ego concept. However, it accepts that mental unity is fundamental for personal identity. Consequently, it also rejects the river-like consciousness idea.

6. The Self as the Epiphenomenon

Now, it should be clear that I blame the concept of the self as the unifier of experience for all the puzzles that arise around the split-brain experiment. Let me introduce the opposite concept: the self that is not the unifier.

As explained before, in his analysis of the physics exam experiment, Parfit separates the concepts of a person from that of a subject of experience. However, when he discusses Nagel and his "brain equals self" thesis, Parfit writes:

> "The existence of a person just involves the existence of his brain and body, and the doing of his deeds, and the occurrence of his mental states and events. But though they are not separately existing entities, persons exist. And a per-

> son is an entity that is distinct from his brain or body, and his various experiences. A person is an entity that *has* a brain and body, and *has* different experiences. My use of the word 'I' refers to myself, a particular person, or subject of experiences. And I am not my brain" (Parfit, 1984: p. 471).

If the subject of experience is identical to a person, then to say that for any mental state, there is a subject that has the state means that any mental state belongs to a person (this strongly suggests that animals can be persons, too). Consequently, if the commissurotomy patient could be understood as a single person, any mental state would belong to the same person, and the latter would be the only subject of experience. If a person is the subject and the self does not provide the unity of experience (as Parfit accepts in his radical "consciousness like a river" thesis), could there be any reason to interpret the commissurotomy case or the physics exam experiment as involving one person but two subjects of experience as Parfit does?

Tim Bayne (2010: pp. 269-270) indicates that having a perspective, apart from being the unifier of experience, as noticed before, is one of the fundamental features that any concept of the self must include. A possible reason to interpret a commissurotomy patient as one person but a double subject could be the thought that if there is a perspective on the world, there must also be a subject or a self because only subjects or selves have perspectives. And the commissurotomy patient appears to have a double perspective. Following Schechter (2018: Ch. 2), it could be explained that access unity and awareness unity are necessary to enjoy a perspective on the world. The patient cannot access elements of experiential content, delivered by the left and right hemispheres, that could be, in Block's terms introduced before, available together for "reasoning, reporting, or rationally guiding action." Such a patient also cannot make the content from the two hemispheres "the common object of a single state or act of awareness"; two elements of the content cannot be "jointly made the object of a single higher-order thought or state of awareness" (Schechter, 2018: p. 26; the notion of awareness unity is derived from the notion of the higher-order awareness; see Rosenthal, 1986 and Rosenthal, 2005). There must be two perspectives because there is no access and awareness unity. If, for any perspective, there must always be a subject that has the perspective, then even if there is a single split-brain person, there must be two subjects of experience (Schechter, 2018: Ch. 2, systematically defends this as her thesis on the commissurotomy case, while Parfit presents it much more casually in his discussion of the physics exam experiment, as indicated above).

Such an argument seems to assume that the unity of experience still demands something else to keep the unity. And the subject takes the same function as the Cartesian Ego; it becomes the unifier. In such a view, it seems unlikely that a single subject of experience has two different perspectives because whatever belongs to the same subject should be unified in the same stream of consciousness. Consequently, it seems improbable that one subject could unify two separate streams of consciousness. If it could, Parfit in the physics exam experiment would be inter-

preted as a single person and a single subject of experience who does the calculations in two separate streams. Despite his "consciousness like a river" thesis, Parfit seems to assume that the stream of consciousness is unified and that the two subjects in the two streams in his physics exam experiment are there to maintain the two unities. If this supposition is incorrect, I cannot find any other reason why Parfit posits two subjects of experience in this case.

My first thesis is that nothing beyond the stream of consciousness provides unity. The parts of our bodies, from cells to organs, including neurons in the brain, are united and constitute the organismic system. Nothing external is needed to maintain this unity. The same must be true for psychological unity: it must be immanent in the mental domain. Mental states unite because of their innate nature, thereby forming the system known as the mind. If something has the mind, the thing is able to experience. Therefore, it becomes the subject of experience. Persons have brains and bodies, as Parfit writes in the passage quoted above. Persons also have minds, which justifies calling them subjects of experience. If a single person has a single mind, that person is a single subject.

An important objection could be put here. If we accept the psychological criterion of personal identity, how could it be possible that a single person has divided psychology and still is a single person? An answer to the question should be a concept of the self since the self makes a person. Speaking about the mind is not speaking about the self yet. The mind is a mental system that enables mental abilities. While persons have such abilities, so do animals. Animals also have minds that make them subjects of experience. However, explaining a mind is not necessarily enough to explain a person. Frogs and dogs have minds since they have mental abilities, but whether they have selves is disputable. I believe they do, but proving they have minds is not sufficient to prove they have selves that make them persons.

The self must be a psychological phenomenon to serve as something useful in the psychological approach to personal identity. Because of the critique presented before, the self cannot be something that provides the unity of experience. It also cannot be merely a single mental state among other mental states since it seems inadequate to account for our phenomenal experience, as indicated before, and this inadequacy was the crux of David Hume's famous argument. It must be something different. I propose the second thesis that the self is an epiphenomenon.

It could be expected that the thesis should be supported by a positive argument proving the existence of an entity called the self, which possesses properties that make it epiphenomenal. However, such an expectation would be rather inadequate for this thesis.

In some respects, an epiphenomenon is like an illusion; indeed, the former could even be conceived as a type of the latter. A mirage in a desert appears to be a physical entity, but nothing of the sort exists in reality. Part of the truth is that it cannot directly cause anything, at least when conceived beyond psychological terms. The only thing that exists is the illusory experience. An epiphenomenon

appears to be an entity, but no entity can be called the self. By entity, I mean something that participates in causal relations. The only thing existing is the "illusory" experience, which cannot cause anything, even in psychological terms. However, like any other experience, it has a mental character.

We would have to follow certain steps to prove that something is an illusion. Firstly, we would need to identify the illusory experience in question. An illusion is inherently phenomenal and, therefore, subjective in character. As a result, we would have to find it in our own subjective experience. Secondly, we would have to prove that something does *not* exist. The illusory experience is pointed out in the first step, but the thing that the illusory experience seems to represent does not exist. Thirdly, we would have to explain how the experience appears.

To argue for the epiphenomenalist thesis, I will follow similar steps. First, I will identify an experience that leads us to believe that there is an entity that could be called a self. Second, I will prove that the entity does *not* exist. By entity, again, I mean something that plays a causal role. Third, I will explain how the "illusory" experience of the alleged entity appears.

1) It is not entirely true that the self does not appear in our phenomenology. Something in our experience could be identified in the first step. Nagel writes about it rather suggestively:

> "The concept of the self seems suspiciously pure—too pure—when we look at it from inside. The self is the ultimate private object, apparently lacking logical connections to anything else, mental or physical. When I consider my own individual life from inside, it seems that my existence in the future or the past—the existence of the same 'I' as this one—depends on nothing but itself" (Nagel, 1986: pp. 32-33).

Similarly to the case of an illusion, I assume that what Nagel describes is something we all recognise from our own experience. This could be called the phenomenal self. I cannot directly prove that we all have this experience because it has an entirely subjective character, like an illusion. As a starting point of my argument, I merely assume that everyone can recognise it. Moreover, I assume that only persons can have such experiences. Therefore, the phenomenal self, whatever it is, qualifies as the self because, according to my initial assumption, if something has a self, it is a person. I aim to argue that the phenomenal self is actually epiphenomenal.

Since the Cartesian Ego concept fits Nagel's description rather well, an intuitive feel of the self or the so-called common-sense view is often considered a kind of Cartesian dualism. The Cartesian Ego concept interprets the "illusion" as "reality", representing the entity that could play a causal role, e.g., maintaining the unity of consciousness. In the previous sections, I refuted the view that the self provides mental unity. Therefore, I rejected the Cartesian Ego concept and proposed my first thesis that nothing beyond the stream of consciousness provides the unity of experience. The self becomes the ultimate *effect* of the unity of con-

sciousness. It is not that the self provides the unity; it is the unity that provides the self. If the unity relations are immanent in the mental domain, mental states unite into the mental system—a mind—and the self appears as the ultimate effect of the unity. The self does not provide the unity of consciousness and is not within the set of unified mental states. The self exists *because* mental states are unified, not the other way. However, I consider the intuitive self to be just what Nagel describes: the phenomenal self that anyone can find in one's own experience. Therefore, I do not propose rejecting this intuition but rather reinterpreting the experience in question.

2) There could be a way to support the thesis that the self, understood as a mental phenomenon, can stand in causal relations, even if my first thesis is accepted. Mental states can have relations with themselves that not only constitute the mental system called the mind. The relations could be considered bottom-up and constitute the emergent self that, in turn, enters top-down causal relations. If such a thesis is true, the phenomenal self is the emergent self. In this case, the epiphenomenalist thesis would be false. Therefore, the emergent self, which I must argue, does not exist.

Emergentism typically concerns the mind, not the self. It states that the mind is a mental system that emerges from the physical one, specifically the brain. I do not deal with the mind-body problem in this work. As indicated before, explaining the mind is not enough to explain a person. Therefore, I distinguish the mental system called the mind from its specific feature called the self, which is what makes something a person. This framework establishes the following relation: a creature has a mind, and a mind has a self, which makes the creature a person. Considering this relation, if the self is emergent, it emerges from the mind (consequently, emergentism about the mind is unnecessary for emergentism about the self, and similarly for epiphenomenalism; therefore, arguments for an emergent mind would not be helpful).

Why reject emergentism in favour of epiphenomenalism? If something is an emergent entity, it should be recognised by its higher-order properties that are not reducible to lower-order ones. If we cannot indicate such properties, there is no reason to suppose anything is emergent. What properties does the phenomenal self have that cannot be reduced to those of the mind?

The phenomenal self not only does not have higher-order properties. It does not have any properties at all. The self is "pure" and "lacking logical connections to anything else, mental or physical".

Persons can be tall, heavy, round, colourful or shy. Selves are none of these. The only quality of the self is the phenomenal quality of "what it is like" to be the person. However, such a quality is the ultimate effect of unified phenomenal qualities of all mental states at a given time and, therefore, the phenomenal quality of a total mental state (as I will explain in the third step of the argument). Whatever could be said about such a total quality could be said in terms of mental states within the set (e.g., explaining what it is like to be me would include the phenom-

enal quality of writing this article, not qualities of my self because there is nothing to say about it). Therefore, the total "what it is like" quality is not emergent from qualities of all mental states. It is difficult to understand how something without some substantial properties could cause and make a difference in the effect. Therefore, the supposed top-down causation of the "pure" self is also hard to understand.

Moreover, it would be peculiar to say that a self has caused a mental state or an agent's action, e.g., to say that a self has caused a feeling of disgust. A cause can be a physical state in the world, e.g., a disgusting object, or a mental state of mind, e.g., a disgusting thought. I can say that I did something because I had a particular mental state, e.g., I wanted to achieve something. There is no sense in saying that I did something because of my self or that this self caused me to do it. Whatever could be indicated as a cause of one's action or a mental state is another mental or physical state, not the self.

It may seem a strangely complicated account to suggest that the self emerges from the mind. Why not claim that a mind is just a self? As indicated before, some creatures have minds, but whether they are persons is disputable. Therefore, having a mind is not enough to designate something for a person. There must be something specific about the mind that makes some creatures persons, and this is precisely what I call the self. Some creatures have minds, and because some minds have selves, some creatures are persons.

Another candidate for being a self with causal powers could be a brain, but it should be clear that the "brain equals self" thesis will not provide the expected answers, since I assumed the self makes a person. We can find correlations between mental states and brain states. However, these are correlations between a mind and a brain, not between a self and a brain. The question of why some brain owners are persons is similar to the question of why some mind owners are persons (Nagel further complicates the matter because he tacitly identifies a mind with a self in his works).

This issue could be illustrated through a psychophysical spectrum thought experiment (like those Parfit, 1984 uses for his reasons). We reject a healthy brain neuron by neuron, aiming to reject the organ altogether. Neuroplasticity must have limits, and it may be expected that psychological functions will degrade eventually (functions of the whole brain will not be restored in a few neurons). In terms of mind, it may be clearly explained which functions and abilities the organism loses along the way. The mental system is decomposing. However, if a mind is just a self, and the self is what makes something a person, does this mean that we are decomposing a person? Does this mean that a person without a part of the brain is less a person than people with whole brains? Would a person with only one hemisphere be considered only half of a person? If not, at what point on the spectrum does the person cease to exist? These questions beg for answers, but their answers are about something more than the mind or the brain. They concern the self: what makes something a person?

Being an epiphenomenon is still something rather than nothing. A more radical thesis must also be addressed. Since there are no causal relations that the self could be a part of and no sense in which we could ascribe properties to the self, does this not prove that the self just does not exist? I consider answering this question a purely conceptual choice. If the self does not exist, the phenomenal self is an illusion. Indeed, I have already indicated that an epiphenomenon could be viewed as a kind of illusion. Nonetheless, I would like to explain why I prefer to call it an epiphenomenon.

If something is an illusion, it misleads us, representing something that does not exist. There is an experience, but no actual thing that the experience represents. It could be argued that the phenomenal self represents something like that, namely, a mental object that could be conceptualised as the Cartesian Ego. If the latter does not exist, the phenomenal self represents something that does not exist. It could be justified to call it an illusion.

However, illusions have some experiential and describable properties. Thus, some experiential properties can be indicated as misrepresenting the properties of a thing in the world, e.g., an experiential broken stick in water misrepresents the properties of an actual stick that is not broken. Meanwhile, no properties of the phenomenal self could be considered similarly misrepresentative for at least two reasons. First, there is no way to compare the phenomenal self with something it "represents" and identify the difference. Second, the phenomenal self is not only "pure" but also lacks "logical connections to anything else". Therefore, it is hard to understand how it could represent anything at all. What properties should something have to be what the phenomenal self allegedly represents? To say that something should be "pure" and beyond any "logical connections to anything else" does not provide any meaningful criteria. There would be no way to verify whether something like that exists in the world. Ultimately, it could be a hallucination rather than an illusion: there is the experience, but it represents nothing. However, it seems too odd to suppose that we all hallucinate all the time.

If the self does not provide the unity of experience but is the effect of the mental unity, and there are no top-down causal relations between the self and the mind, the self is an epiphenomenon (of course, there are many other options left, but this paper is limited to psychological concepts of the self, as I indicated at the beginning).

3) According to my first thesis, nothing beyond the stream of consciousness provides its unity. On this basis, I claim that the self is the ultimate effect of mental unity. Since the self in question is phenomenal, how it arises must be understood in terms of phenomenal unity.

For any two experiences, e_1 and e_2, that occur at a specific time t, there is a phenomenal quality of "what it is like" to have a single quality of $e_1 + e_2$. The phenomenal quality of e_1 is never separated but must be understood in relation to the phenomenal qualities of e_2 and other experiences that appear in t. Consequently, a total mental state M, the set of all mental states at t, has its total "what it is like"

quality. The latter appears because all mental states within the set of *M* are interdependent and affect each other. Eventually, the total "what it is like" quality is the phenomenal quality of being the subject of this total mental state. This is akin to the bat's "what it is like to be a bat" quality.

The mental unity of the total mental state has a specific structure organised around an abstract point that Bayne (2010: p. 289) calls "the phenomenal centre of gravity". Everything we experience is somehow related to this centre. Bayne introduces his view this way:

> "*De se* representation isn't the exclusive provenance of explicitly self-conscious thought, but permeates consciousness through and through. (…) the conscious states evoked by the presentations of one's senses are automatically *de se*. In effect, this means that streams of consciousness (…) are constructed 'around' a single intentional object. The cognitive architecture underlying your stream of consciousness represents that stream as had by a single self—the virtual object that is brought into being by *de se* representation" (Bayne, 2010: p. 289).

Because Bayne assumes that one of the fundamental roles of the self is to provide the unity of consciousness, I presume he imagines that the "cognitive architecture" projects an "intentional object" into the phenomenological field as a *de se* representation, or something that "has" experiences and unifies them. Because the object does not just appear in our experience, this *de se* representation "permeates consciousness through and through".

The thesis I propose rejects the necessity of having any "virtual object" that provides the unity of consciousness but does not reject the phenomenal self. The latter is not projected by the "cognitive architecture" for any reason but appears as a side-effect of how the experience is organised, and is thus the structure of phenomenal unity of the total mental state.

Some experiences seem to be closer to the centre than others. These are fundamental for maintaining the self. The experiences that are closest to the centre are probably those mediated by proprioception: a bodily self is the first to develop in a child; even a newborn presents a very organised awareness of his or her body (see Rochat, 2001: Ch. 2, for many examples of psychological experiments). Next, there are likely experiences of our agency and will: when we do something expecting particular results (a child develops a sense of agency between the second and sixth month). These and other close-to-the-centre experiences could be called core ones. This notion can be derived from Peter Unger's notion of core psychology: "capabilities [that] are shared (…) with many humans who are markedly below the psychological norms" (Unger, 1990: p. 68). Other experiences are distinctive ones. This notion can be derived from distinctive psychology: "[abilities that I partly] share with some, but not all, other normal human beings, and the rest I share with none" (Unger, 1990: p. 68). Phenomenal unity (and any other mental unity) demands core experiences much more than distinctive ones. Experiencing our own body while walking is a core experience. Experiencing a lecture on anat-

omy is a distinctive experience. Rejecting the former threatens to break the unity of consciousness to a much greater extent than rejecting the latter. Since the unity of consciousness supports the self, breaking this unity means death (I assumed at the beginning that anything mental is phenomenal; therefore, the lack of phenomenology means the lack of psychology).

Phenomenology is organised around an abstract point that is strangely experienced as the "ultimate private object", one that is "pure" and lacks "logical connections to anything else". The order of the phenomenological field relies on the fact that everything we experience relates to ourselves, namely, the persons we are. Therefore, we can recognise the phenomenal self in our experience not through any specific content of our phenomenological field but through the order of the field. It seems like identifying a cyclone's eye while standing inside. In my experience, I recognise myself as the centre of everything given in the experience. But because *everything* is given to me through experience, nothing is left, and the centre appears to be "pure", as Nagel writes in the quote. Whatever I can indicate in my experience is something organised around the centre, not the centre itself. Therefore, the "what it is like" quality of being me is constituted by the "what it is like" quality of the total mental state, so all my mental states at the time, not by qualities of the phenomenal self.

There is an interesting illusion that the self is the source of everything we do. Ludwig Wittgenstein described it this way:

> "One imagines the willing subject here as something without any mass (without any inertia); as a motor which has no inertia in itself to overcome. And so it is only mover, not moved. That is: One can say 'I will, but my body does not obey me,'—but not: 'My will does not obey me.' " (Wittgenstein, 1958: §618).

From the external point of view, a person is somebody who takes an action, and the person's mental state starts the action. Somebody wants something and does something as a consequence. However, from the internal point of view, it seems that the self is the ultimate source. The "illusion", however, stems from the opposite view. The self depends on *everything* we experience because everything we experience supports the self. It is like the eye of a cyclone that seems to be a source of the whole phenomenon while it is the ultimate effect. The eye itself cannot directly cause anything. There is nothing in the eye that could do anything. The atmospheric forces that support the eye stand in causal relations, not the eye itself.

Similarly, the self, the "pure" centre of experience where nothing can be found, does not cause anything. The source of action is our will. While we are almost born with a structured proprioceptive feel of our body, our sense of agency develops around the second to sixth month of infancy. The agency's phenomenology is organised very close to the phenomenological centre. However, desires are states among other mental states within the same mental system of a mind. The same system that supports the self, like atmospheric forces which support the eye, causes what we do, not the self.

The view I have presented explains why the physics exam experiment is unimaginable from the inside. A perspective on the world is an individual way of how our experience is unified. If I imagine a first-person perspective, I imagine one unified first-person perspective that could have only a single self at its centre as the ultimate effect of its unity. If I imagine two perspectives, as Parfit wants, there is no unity between them. Therefore, there is no place for a single self shared by both views. Parfit wants us to do it as if the self would be prior to the unity while the unity is prior to the self. There must be unity first; then, there could be a self. We cannot imagine a self having two or more perspectives because we have to imagine a perspective to have a self within the picture. It is more like a perspective having a self than a self having a perspective, despite the misconception that the self is independent of anything within the phenomenological field. On the contrary, the self depends on everything.

I have already rejected two of the three roles that the concept of the self should include, according to Bayne (2010: pp. 269-270). The first is that the self provides the unity of experience. The second is that the self has a perspective on the world. The third is that "I" refers to the self. I would also like to reject the last claim to place everything in proper conceptual order.

The "I" notion does not refer to the phenomenal self. Whatever I could say about myself does not make sense when speaking about the self. I could be a tall, white-skinned, shy father and citizen. The self is not tall, white-skinned, or shy. Selves are not fathers and citizens. Therefore, Parfit is correct in the passage quoted at the beginning of this section: "the word 'I' refers to myself, a particular person, or subject of experiences". Some creatures have minds that make them subjects of experience. Minds have selves that make the creatures persons. "I" refers to the person, not the mind or the self. The person is who I am. Persons can be tall and shy fathers or citizens. The phenomenal self, being "pure", can be none of these.

7. The Simple Split Experiment

I would like to propose a simplified version of the split-brain thought experiment (inspired by John Conway's *The Game of Life* used in Dennett, 1991). Because of its simplified form, it does not serve to prove anything or draw any further conclusions about the self or consciousness but only to illustrate the essence of the proposed view and to examine its key concepts more closely. I hope to show how the crucial aspects of the view sketched earlier can support the disjunction claim. However, despite trying to remain as neutral as possible, the way I present the experiment may still subtly favour some underlying solutions above others. I will briefly address this issue at the end of the section.

The experiment deals with a simple life form whose total mental state in time t_1 consists of only eight mental states: $\{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$. All the states are unified and, therefore, support a single epiphenomenal self: *S*. For the sake of argument, *S* could be considered a mereological sum of all these eight states. The

more states within the set, the stronger their unity and the stronger the support of *S*.

The sequence within the set is not accidental for two reasons. Firstly, the specific order decides how the subjective experience is constructed. Any content of my experience is so related to me that I am at the centre of this experience. But it is not that the order in question appears *because* I am at the centre. I am at the centre because there is an order. In other words, I recognise the phenomenal self of mine based on these relations that constitute the order in question.

Secondly, mental states differ in their significance for providing support for the self. Core experiences are fundamental, while distinctive ones are less important for maintaining the self. For example, a proprioceptive feel of the body plays a crucial role in almost every waking moment, while a memory of the past may be much less critical. The degree of importance is also represented in the order of the set: e_1 plays a crucial role in supporting *S*, being a core experience, while e_8 does not, being a distinctive one.

The set's order does not represent the feature of being conscious. The experiment assumes that unconscious mental states can support the self as well (according to my initial assumption, anything mental is phenomenal; according to my thesis, being unified is the intrinsic feature of a mental state itself; therefore, I accept a controversial thesis that any mental state is unified; see the unity thesis in Bayne & Chalmers, 2003, and Bayne, 2010: Ch. 1.4). Many unconscious states may even provide a more substantial basis for the unity of experience, and consequently, for the self, than many conscious ones, e.g. my unconscious memories can play a more significant role in how I experience social interactions and behave in a group than my conscious memory of a last meeting; my overall behavioural unity depends on probably more skills that I use unconsciously than what I am conscious while doing something, etc. In short, the simple life form experiment assumes that being conscious or not at least does not exclude any mental state from being a support of the epiphenomenal self.

The experimental life form has a simple neuroplastic brain. However, any division must be a process, so there must always be some in-between states. Therefore, if we cut away a part of the simple brain, the total mental state in time t_2 may look like this: {e_1, e_2, e_3, e_4, e_5}; but in t_3 the state may be like this again: {e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8}. But could we cut away states from e_1 to e_4—the most critical states in supporting the self? If we could, it should be expected to be a much more difficult process to get the lost psychology back. Therefore, it does matter what exactly is cut away. If we had after the operation a total mental state like this: {e_1, e_2}, it would be a much stronger basis for the expected restoration than a state like this: {e_7, e_8}.

How could the split operation be done on such a set given in t_1? To start with, we could not just divide the set into two halves like this: {e_1, e_2, e_3, e_4}, {e_5, e_6, e_7, e_8}. The first set would have a much stronger basis for being a psychological continuation of *S* than the second one because states e_1-e_4 play a much more critical

role in supporting the self. Moreover, if e_1 were the proprioceptive feel of the body, saving the state e_1 would mean the necessity of leaving its physical bearer, so a part of the brain, with its initial body. It seems that the initial self S would stay with states e_1-e_4 and just lose its previously supporting states e_5-e_8. There should be an additional step in the procedure where states e_5-e_8, once unified with states e_1-e_4, unify only with themselves within the new set. This means that the second set may have a brand new self, and therefore, it would be a new person, but the person would only be derived from the original one. It is not the fair trade assumed in the split-brain thought experiment. Both persons should pass the psychological continuity test equally well, while it seems obvious that in the case discussed here, the second one should not pass it at all.

Such a division must be something outside the scope of our interest. Maybe we should consider something less radical first. It could be a cut of the simple corpus callosum. In effect, we could have a set like this: $\{e_1, e_2, e_3, e_4, \{e_5, e_6\}, \{e_7, e_8\}\}$, where the two subsets are the effect of the surgical cut. Maybe in our laboratory circumstances, it would not be possible for the creature to unify the e_1-e_4 states with the two subsets simultaneously. Depending on the laboratory circumstances, the set would look like this: $\{e_1, e_2, e_3, e_4, e_5, e_6, \{e_7, e_8\}\}$, or like this: $\{e_1, e_2, e_3, e_4, \{e_5, e_6\}, e_7, e_8\}$. It would mean that one of the subsets would be excluded from the actual unification and, therefore, from supporting the self. However, it would not mean that there are two selves. A single self is primarily and continuously supported by states e_1-e_4, while the two subsets change in their supportive role depending on the circumstances. Nevertheless, the single primary self is continually supported for the whole time. Consequently, the commissurotomy of our simple living creature gives us no material for any split-brain experiment that would result in two numerically different simple persons.

We could also check if our simple creature could participate in a simple physics exam experiment. As I claimed before, the experiment is unimaginable, and if I am right, it could not start or end with assumed results. Parfit wants us to think that his experiment is analogous to the commissurotomy case. A simple version of the latter was represented this way: $\{e_1, e_2, e_3, e_4, \{e_5\ e_6\}, \{e_7, e_8\}\}$. Our simple living being does its two simple calculations in the two subsets. The question I asked before is: how could the decision about reunification be made?

The state of the decision e_D must appear in the main set, among the e_1-e_4 states: $\{e_1, e_2, e_3, e_4, e_D, \{e_5\ e_6\}, \{e_7, e_8\}\}$. The reason is that to make the decision, our creature should be *aware* that it is doing the calculations two ways. Therefore, maybe it could be something like a higher-level skill of a drummer playing different things with two hands. But it differs from what Parfit looks for because it differs from what commissurotomy patients can do. They seem to have no conscious thought that there are two different "branches" within them. This is why Parfit wants a situation where there are "two separate streams of consciousness—two series of thoughts and experiences, in having each of which they are *unaware of having the other*" (emphasise added). If the creature is unaware that it has two

subsets and decides based on that, the decision e_D cannot appear in the primary set among the e_1-e_4 states. It should appear in one of the two subsets, but it is impossible because e_D demands awareness that there is another branch to unify, while there is no such awareness within the subsets. Such awareness could appear only within the main set, but this is not the situation assumed in Parfit's experiment.

The conclusion is that if our simple creature can consciously decide on reunification, then the situation: {e_1, e_2, e_3, e_4, {e_5 e_6}, {e_7, e_8}} is not precisely like the situation of the commissurotomy patients. The difference is the possibility that among e_1-e_4 are conscious states that make the creature aware of the existence of the two subsets. Otherwise, the creature cannot consciously decide on reunification if the situation is similar to the commissurotomy case. If there is no place for e_D among the e_1-e_4 states, there is no place for it at all because the constitutive feature of states within the two subsets is that they are unaware of the other subset. But such awareness is crucial for having e_D. Parfit wants the impossible: a situation like the commissurotomy case and the reunification decision at the same time.

This illustrates why the whole operation is unimaginable subjectively. We can imagine only one unified perspective of doing calculations in one way. If we were unaware of the other "branch", we could not imagine that we would decide on reunification with the latter because it demands at least minimal awareness of the situation. We cannot imagine that we are aware while being unaware at the same time.

The simple commissurotomy situation gives no ground for the ideal split-brain experiment where we should have two simple persons who are fully psychologically equivalent to the initial person. To explore the question further, let us take a closer look at the experiment on our simple living being and discover what could happen.

It should be clear by now that if we did the cut like this: {{e_1, e_2, e_3, e_4}, {e_5 e_6, e_7, e_8}}, the first subset would have a much stronger basis to support the initial self. It should also be clear that a cut like this: {e_1, e_2, {e_3, e_4, e_5}, {e_6, e_7, e_8}} would not give any better result. There are still states e_1 and e_2. Maybe these two states come from the proprioception. As long as the two states are dependent on the body, the part of the simple brain that stays with them is in the privileged position. The initial self would remain with them.

In the case of people, our body is doubled: we have two legs, two arms, two eyes and also two hemispheres that are connected with specific parts of the body, left or right. Our simple creature could be like that. Therefore, perhaps it could be possible to divide its brain in such a way that two parts work separately, each controlling its specific side of the body, left and right: *L*: {e_1, e_2, e_3, e_4} and *R*: {e_1, e_2, e_3, e_4}. The creature is devastated here because it cannot coordinate any move that engages both body parts (something radically more severe than the callosal disconnection syndrome that the commissurotomy patients suffer for some time directly after the operation). But then, we separate them into different simple bodies

by transplanting the divided brain, and maybe it could make two simple persons from the initial one. But we should remember that there are always in-between states. Even if the last total mental state could be possible to achieve, there will always be an in-between state looking more or less like this: {{e_1, e_2, e_3, e_4}, {e_5 e_6, e_7, e_8}}. Some states are always more privileged than others.

What is unified could be perhaps divided, but the unity itself is not dividable. The only move we could make is to separate a part of the unified set and unify the derived part afresh. But since the initial unity supports the initial self, the other self would be newly created from the derived psychology. It does not generate the Russell-like paradox here because the epiphenomenal self is not the unifier of consciousness. Even in normal conditions of healthy people, according to the epiphenomenal concept of the self, such a self depends on the rest of psychology. Therefore, being created from the remnant psychology of the initial being is not a problem. However, such a self is always new, created from the derived psychology, not the original one.

The question immediately arises: which self is the original one? In the idealised thought experiment, it is hopelessly hard to determine which part of the simple brain is more privileged. The commissurotomy case seems to show that it could depend even on actual accidental circumstances. Depending on the experimenter's question, the patient's mind unifies not only perceptual data but also their behavioural abilities called to action by the question. Our simple split-brain case could be even tougher because the division goes deeper. Which part of the brain plays a more important role during the division process could be highly accidental and dependent on the environment.

Nevertheless, assuming that our simple split experiment has ended successfully, one of the two parts saved the original unity while the other recreated it. The former has the original self and is the original simple person, while the latter is a new one. The only alternative is that the original unity was broken due to our surgical intervention. Two new simple persons appeared from the remnant psychology saved in two hemispheres. However, the broken unity would mean the disintegration of personality and the death of our simple living being.

Beyond all I have said about the simple split-brain experiment, there is an issue that could complicate the discussion. I represented the simple commissurotomy this way: {e_1, e_2, e_3, e_4, {e_5 e_6}, {e_7, e_8}}. This seems to suggest that e_1 is unified with e_5, e_1 is unified with e_7, but e_5 is not unified with e_7. As such, this seems to represent the idea of partial unity (discussed in detail in Hurley, 1998). However, the representation of the patient's state is idealised here, e.g., it does not consider time. Therefore, if the idea of partial unity is correct, it could be that the patient's state at one given point in time is exactly this: {e_1, e_2, e_3, e_4, {e_5 e_6}, {e_7, e_8}}. If this idea is incorrect, it could be that at no point in time does the patient's consciousness look like the above; rather, at one point in time, it looks like this: *L*: {e_1, e_2, e_3, e_4, {e_5, e_6}, e_7, e_8}, and at another point in time, it looks like this: *R*: {e_1, e_2, e_3, e_4, e_5, e_6, {e_7, e_8}}.

Nevertheless, both cases could be interpreted without the partial unity idea. It may be claimed that even if e_5 is not directly unified with e_7, they are still unified indirectly through states e_1-e_4. There could be interhemispheric integration without hemispheres being connected through the corpus callosum. Schechter argues that there are two subjects of experience, even in a healthy brain, but they interact. In the commissurotomy patient's mind, the interaction does not occur in a "psychologically directly" manner (Schechter, 2018: pp. 105-106). Bayne's (2010: Ch. 9.5) view is the second answer. In his model, the patient's consciousness switches from *L* to *R* depending on the circumstances. Both Schechter and Bayne reject the partial unity idea on the grounds of experimental evidence that there is interhemispheric integration (like the experiment where the commissurotomy patient eventually answered what the right hemisphere was seeing: a picture of Hitler; Sperry et al., 1979: pp. 159-160), which could be enough to maintain that the unity of consciousness is preserved (or that "mental duality", in Schechter's view, is not partial unity).

This issue would become more complicated if there were any evidence that some mental states are disunified with the rest of the mental system, thereby leading to a lack of interhemispheric integration. A total mental state would look this way: $\{e_1, e_2, e_3, e_4, e_5\ e_6\}, \{e_7, e_8\}$. States e_7 and e_8 would not be unified with the other six states. However, why should we still call it partial unity? It appears there are just two separately unified sets of mental states. But if the previous situation is still the unity of consciousness, while the present one is no longer any unity, where is the partial unity?

It is hard to understand the concept of unity as being both partial and divided. Unity is not gradable. Therefore, as it is usually expected, unity is transitional. Nevertheless, the idea of partial unity demands much further elaboration than what has been provided in the literature so far.

8. The Solution to the Split-Brain Paradox

The solution can be put in order by referring to all its claims individually.

Firstly, the criterion of personal identity is psychological continuity. The crucial difference between continuity and connectedness lies in transitivity. Connectedness has the feature, while continuity does not. This is why connectedness counts more for identity. But how could we realise that the "particular direct psychological connection" (Parfit, 1984: p. 206) that connectedness is supposed to be is "direct" enough to count as connectedness and make identity possible? How far could the directedness go? Considering Methuzalem's case mentioned in the beginning, there is no connectedness between his state in t_1 and two thousand years later, in t_2. But where in the time spectrum is connectedness with the state in t_1 over? It could not be a matter of degree; connectedness would lose the constitutional feature it shares with identity, namely, transitivity. There should be a point when connectedness would be suddenly and entirely lost. But there is no idea when that could be.

The epiphenomenal self does not need the connectedness idea. The self depends only on unity, and the unity is maintained during the whole lifetime of Methuzalem. In effect, he has a single self that makes him the same person for his entire lifespan. If our simple living being were a simple Methuzalem, it would be like this in t_1: {e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8}; like this in t_2: {e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11}, e_{12}}; and like this in t_3: {e_9, e_{10}, e_{11}, e_{12}, e_{13}, e_{14}, e_{15}, e_{16}}. The mental states support the self change, but the continually supported self is still the original one. Because the self is what makes something a person, the simple Methuzalem is still identical to himself from the past. Continuity is enough because it is enough for mental unity, and the unity supports the self.

Secondly, P_1 and P_2 are psychologically indistinguishable. This means that P_1 and P_2 pass the test on psychological continuity equally well with near 100% compatibility. We knew why Methuzalem in t_2 could not pass the test on continuity with Methuzalem in t_1. The reason was a lack of data from the in-between states of his life. The more data the computer had, the higher the probability of results.

The results would be the opposite and not gradable in the split-brain case. If the computer had enough data, it would give a 100% result for one person, say P_1, but 0% for the other, say P_2. It would have enough data, mainly if it continually tracked the mental life of P_{Zero}. It would have the information on which person's psychology continually supports the same self and which person's psychology was separated and started to realise another self. If the computer did not continually track the mental life, it would give a near 100% result for both P_1 and P_2. The reason is that it could count the effect only on the similarity between the *content* of their psychologies, namely, their psychological profiles at different points in time. It would lack the data on *relations* that support the self. The relations are maintained across time, and any gap in the computer's tracking breaks up monitoring the relations that support the self. In our simple life form experiment, based on states of two halves of the initial brain placed in different bodies, the computer would only conclude the states: *L*: {e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8} and *R*: {e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8}. The *content* shows 100% compatibility with the initial state of the original simple being. The probability that the unity of *L*'s and *R*'s total mental states comes from the unity of the original state must be near 100%. However, the data about *relations* and their continual realisation across time are crucial to assigning the identity. Where there is no break up of relations supporting the self, there is the same simple living being identical to the initial one.

Thirdly, P_{Zero} survives as only one person, e.g., P_1 or P_2. The continual realisation of mental unity is possible only for one person. The unity is not dividable. It could be continually realised or broken up. This means that only one person—say, P_1—could have the original unity and, therefore, the original self. The other one would have a newly created self or nothing: there would not be any P_2. The second possibility is overwhelmingly more likely because the psychological conditions for creating the second person are absurdly complicated, as even the simple split-brain experiment shows. It demands the creation of two persons in one

body and then the separation of them. But the same experiment shows that even if such a crazy experiment could somehow result in two persons in two bodies who could pass the psychological test equally well, it would be only one of them that could still be identical to P_{Zero}. Only one of them could have the original self because the unity of consciousness that supports the self is undividable. The one who is still P_{Zero} is the one whose total mental state gives a more substantial basis for the self.

Fund

The paper was supported by the National Science Centre, Poland, under grant no. 2017/25/N/HS1/00672.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

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

Bayne, T., & Chalmers, D. J. (2003). What Is the Unity of Consciousness? In *The Unity of Consciousness Binding, Integration, and Dissociation* (pp. 23-58). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198508571.003.0002

Block, N. (1995). On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences, 18,* 227-247. https://doi.org/10.1017/s0140525x00038188

Carruthers, P. (2005). Conscious Experience versus Conscious Thought. In *Consciousness* (pp. 134-156). Oxford University Press. https://doi.org/10.1093/0199277362.003.0008

Chalmers, D. J. (1996). The Conscious Mind. In *Search of a Fundamental Theory*. Oxford University Press.

Dennett, D. C. (1991). Real Patterns. *The Journal of Philosophy, 88,* 27-51. https://doi.org/10.2307/2027085

Flanagan, O. (1992). *Consciousness Reconsidered*. The MIT Press. https://doi.org/10.7551/mitpress/2112.001.0001

Hurley, S. (1998). *Consciousness in Action*. Harvard University Press.

Lewis, D. (1976). Survival and Identity. In *Identities of Persons* (pp. 17-40). University of California Press. https://doi.org/10.1525/9780520353060-002

McDowell, J. (1997/1998). Reductionism and the First Person. In J. Dancy (Ed.), *Reading Parfit* (pp. 230-250). Harvard University Press.

Nagel, T. (1971). Brain Bisection and the Unity of Consciousness. *Synthese, 22,* 396-413. https://doi.org/10.1007/bf00413435

Nagel, T. (1974). What Is It Like to Be a Bat? *The Philosophical Review, 83,* 435-456. https://doi.org/10.2307/2183914

Nagel, T. (1986). *The View from Nowhere*. Oxford University Press.

Noonan, H. (1989). *Personal Identity*. Routledge.

Olson, E. T. (1997). *The Human Animal. Personal Identity Without Psychology*. Oxford University Press. https://doi.org/10.2307/2653504

Parfit, D. (1971). Personal Identity. *The Philosophical Review, 1,* 199-223.

Parfit, D. (1984). *Reasons and Persons*. Clarendon Press.

Rochat, P. (2001). *The Infant's World*. Harvard University Press.

Rosenthal, D. (2005). *Consciousness and Mind*. Clarendon Press.

Rosenthal, D. M. (1986). Two Concepts of Consciousness. *Philosophical Studies, 49,* 329-359. https://doi.org/10.1007/bf00355521

Schechter, E. (2018). *Self-Consciousness and "Split" Brains. The Mind's I*. Oxford University Press.

Shoemaker, S. (1970). Persons and Their Pasts. *American Philosophical Quarterly, 4,* 269-285.

Sperry, R. W. (1974). Lateral Specialization of Cerebral Function in the Surgically Separated Hemispheres. In *The Psychophysiology of Thinking* (pp. 209-229). Elsevier. https://doi.org/10.1016/b978-0-12-484050-8.50012-1

Sperry, R. W., Zaidel, E., & Zaidel, D. (1979). Self-Recognition and Social Awareness in the Deconnected Minor Hemisphere. *Neuropsychologia, 17,* 153-166. https://doi.org/10.1016/0028-3932(79)90006-x

Strawson, G. (1994). *Mental Reality*. The MIT Press.

Swinburne, R. G. (1974). Personal Identity. *Proceedings of the Aristotelian Society, 74,* 231-241.

Tye, M., & Wright, B. (2011). Is There a Phenomenology of Thought? In *Cognitive Phenomenology* (pp. 326-344). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199579938.003.0014

Unger, P. (1990). *Identity, Consciousness, and Value*. Oxford University Press.

Wiggins, D. (1967). *Identity and Spatio-Temporal Continuity*. Basil Blackwell.

Williams, B. (1970). The Self and the Future. *The Philosophical Review, 79,* 161-180. https://doi.org/10.2307/2183946

Wittgenstein, L. (1958). *Philosophical Investigations*. Basil Blackwell.

Appendix

The aim of this work was not to present a ready-made solution but to explore the less-trodden path of the disjunction claim and to discover possible insights along the way. Therefore, it could be worth mentioning how the conception presented here relates to other solutions in the split-brain trilemma. In the fifth section, I discussed the idea of the self as the unifier of experience for all three solutions. Here, I would like to briefly discuss my proposed counterpoint: the self as the ultimate effect of mental unity (not necessarily in the epiphenomenalist version).

The negation claim: P_{Zero} does not survive the experiment. At a specific moment, t_1, the unity of consciousness is broken up. No self and, consequently, no person exists. However, it could still be argued that there is a moment t_1 when P_1 and P_2 are recreated from the remnant psychology of P_{Zero}. Since the self is not the unifier of experience, there is no threat of the Russell-like paradox discussed before. The self exists as long as it is supported by mental unity. If P_{Zero}'s remnant psychology united itself, it would create a new self. Because it would be created from P_{Zero}'s psychology, P_1 and P_2 could potentially even pass the psychological test equally well. But they would not be identical to P_{Zero}. Maybe such a solution could support a Parfit-like theory. It could be maintained that P_{Zero} survives the operation without preserving P_{Zero}'s identity. There is survival because of psychological continuity, but there is no identity because the unity and the self are destroyed. Maybe it could support Parfit's famous thesis that personal identity does not matter for survival.

Nevertheless, the negation claim, in probably any version, relies on the unintuitive idea that mental states could exist without any self. This must be the case at t_1 to make the emergence of t_2 possible. The two selves of P_1 and P_2 are created from such selfless mental states. However, this demands a far more rigorous examination.

The interpretation could follow at least two directions. The radical interpretation accepts that not only could there be mental states without any self, but they also need not be unified. If this were true, mental states would be self-standing entities. Could it be possible to experience even a simple mental state like pain without phenomenal unity? Feeling pain has a distinct "what it is like" quality q_1. This quality differs for a headache versus a lacerated wound, or for burning and pulsating pain. The proprioceptive awareness of the body also has a "what it is like" quality q_2. It is much different for a young child and an old man. It also differs in a healthy body and an ailing one. Therefore, depending on many circumstances, there are different $q_1 + q_2$ experiences of pain. Feeling pain is a phenomenally unified experience. Also, in physiological terms, feeling pain is a unified reaction of the neurological system, not a self-standing event.

The less radical interpretation accepts that unified mental states could exist without self. This demands an answer to the obvious question: What is the difference between a set of selfless unified mental states and a set of mental states that support the self?

The view I propose accepts that any mental state is unified (which I see as a consequence of the thesis that being unified is an intrinsic feature of a mental state itself), and any set of unified mental states supports the self. This view offers a significant advantage. A simple life form is a creature with a modest total mental state. In such a set, few mental states are unified. However, even on this modest ground, a simple self could be maintained. As the life form becomes more complicated, additional mental states need to be unified. However, the intrinsic nature of mental states is to be unified. Therefore, the more mental states, the firmer the ground that supports the self. The total mental state of a human being, being the richest, provides the human self with the most substantial support. The advantage is that it allows for an interpretation of the development of the human self without gaps on the spectrum. Being a person becomes gradable. Humans are model persons; dogs are more persons than frogs, and frogs are more so than mosquitos.

If selfless unified mental states are possible (contrary to the view I propose but in line with the less radical interpretation of the negation claim), there must be a critical point on the spectrum when the self appears. This generates two problems. First, what is the crucial point? What must change in the total mental state to enable it to support the self? Second, a different view of the self should be proposed. According to my assumption, the self is what makes something a person. This means that if something has a self, it is a person. To have a self, there must be a set of unified mental states. This is sufficient because any set of mental states supports a self. If there could be a set of selfless but still united mental states, then the set of united mental states would not be enough to have a self. There is something that remains unexplained.

The concept of a selfless mental state was introduced in the literature as quasi-memory by Sydney Shoemaker (Shoemaker, 1970, where he also uses the concept to defend the negation claim). Many philosophers, including Parfit, widely exploited the concept. However, it remains a controversial, non-intuitive idea (for a critique, see, e.g., McDowell, 1997/1998).

The conjunction claim: P_{Zero} survives as both P_1 and P_2. How could the conjunction claim be coherently defended if the self is not considered the unifier of experience, but rather the ultimate effect of mental unity? Since we have two-sided bodies (doubled eyes, ears, nostrils, limbs and hemispheres), we could be considered single persons but doubled subjects of experience (Schechter (2018) defends the thesis; however, she does not discuss the thought experiment of Wiggins and Parfit).

If a mind is a mental system of unified mental states, and such a system makes something a subject of experience, then there must be two minds, one in each hemisphere to create two subjects. However, both hemispheres are connected by the corpus callosum, enabling interaction between the subjects. Therefore, it could be maintained that a single self is supported by the two minds, and therefore two subjects of experience, thanks to the interaction. Since the self makes something a person, it could be a single person despite having two minds. Moreover, the same

reasoning could apply to the commissurotomy patient, due to the interhemispheric integration discussed before. One could argue that the interaction does not occur in a "psychologically direct" manner for the patient, as Schechter (2018: pp. 105-106) proposes; however, this does not mean that there is no interaction.

If P_{Zero} is a single person with two subjects of experience, P_1 and P_2 are the two subjects. Splitting P_{Zero}'s brain means severing any interaction between the two hemispheres. Each subject is transplanted to a different head. However, does this really mean that P_{Zero} survived the operation? A single self was claimed to be maintained through interhemispheric interaction. Since the interaction is cut, there may still be two subjects of experience, but what about the single self of P_{Zero}? If the self is preserved in one hemisphere, the disjunction claim is true; if it is preserved in neither, then the negation claim is true; if it is preserved in both hemispheres, the self is somehow doubled. The situation is the same as the case discussed in the fifth section. It remains a mystery how the self, or anything, could be multiplied this way. The trouble expressed by the split-brain thought experiment is not only unresolved, but it also becomes even more pronounced.

It could be proposed that the two minds in the two hemispheres maintain not a single shared self but two selves. Therefore, there are two persons from the start: P_1 and P_2. Putting aside the controversial idea that we are all under the illusion of being single, while in fact all human beings are double persons, such a proposition ignores the whole phenomenology I discussed in this work. Specifically, if two minds support two selves, the two selves are not phenomenal ones because having two streams of consciousness supporting two selves is unimaginable. There is no phenomenology like that. I argued for this in the third section, particularly through the physics exam experiment.

This paper is limited to a psychological approach to personal identity. Therefore, the self is assumed to be a mental phenomenon. However, it could be argued that the self is something different, e.g., a brain, a human animal, or even nothing. Nevertheless, we experience ourselves the way we do. We cannot imagine being phenomenally disunified. Even if the phenomenology is illusory and the reality is entirely different from how we experience ourselves, the illusory experience, like any illusion, still demands an explanation of how it arises. Whatever the self ultimately is, the phenomenal self demands an explanation.

The account proposed in this work does not demand far-reaching conceptual revisions. It does not reject the intuition of the self, which I call the phenomenal self. The latter is merely reinterpreted in a non-Cartesian way. There are living beings in this world that have minds, and it is this that makes them subjects of experience. If the mind has the self, it makes that living being a person. This perspective does not seem to be unintuitive.