SUBSTANCE AND THE
CONCEPT OF PERSONAL IDENTITY

JENS KIPPER
University of Cologne

In this paper, I identify and discuss the following feature of our judgments about hypothetical scenarios concerning the identity of persons: with respect to the vast majority of scenarios, both members of a pair of logically complementary propositions about personal identity are conceivable. I consider a number of explanations of this feature that draw on the metaphysics and the epistemology of personal identity, none of which prove to be satisfactory. I then argue that in order to give an adequate explanation, one needs to recognize an important characteristic of our concept of personal identity: it is such that if there are mental substances (or the like), they constitute personal identity. At the same time, there can still be persons if there are no such substances. Since this finding casts doubts on the way that thought experiments about personal identity are usually set up, I end by outlining its potential consequences for the debate over the identity of persons.

1. Introduction

The debate over the nature of personal identity is as replete with thought experiments as many other philosophical debates. But there is a nagging suspicion that thought experiments work differently when the identity of persons is concerned. For example, it is natural to think that our judgment that the person in Harry Frankfurt’s (1969) famous case is responsible for what he did is based on our conceptual competence. When John Locke argues, however, that the person in the body of a cobbler who has the memories of a prince is that prince (1694/1975: 340), he does not seem to be stating a conceptual truth.

In what follows, I first flesh out the suspicion that thought experiments work differently in cases of the identity of persons by identifying the following remarkable feature of our intuitions about hypothetical scenarios concerning the

Contact: Jens Kipper <jens.kipper@uni-koeln.de>

http://dx.doi.org/10.3998/ergo.12405314.0003.001
identity of persons: for most such cases, both members of a pair of logically complementary propositions are conceivable. I then discuss a number of potential explanations of this feature, all of which appeal to metaphysical considerations concerning facts about identity or to considerations concerning our epistemic access to these facts. Since none of these explanations turns out to be satisfactory, I propose an explanation of a different type that appeals to the nature of the concept of personal identity. This proposal involves the claim that if there are mental substances, souls, or something of this kind, then these constitute the identity of persons.1 I argue that this account yields a compelling explanation of the conceivability of logical complements. My account also implies, however, that the way thought experiments concerning personal identity are usually set up is questionable. I therefore propose a better way of constructing the relevant hypothetical scenarios, and I end by outlining the potential impact of this new approach on the debate concerning personal identity.

2. Specifying the Explanandum

2.1. Williams and the $100,000 Torture

A well-known thought experiment formulated by Bernard Williams (1970) comes in two versions. In the first version, we are asked to consider two people, A and B, who are told the following: All of A’s psychological characteristics will soon be transplanted into B’s body and vice versa. After this procedure, one of them will be tortured while the other will receive $100,000. Both of them are asked whom they wish to receive the money and whom the torture. Intuitively, it seems obvious that if they base their decisions on purely egoistic considerations, A will want the person with B’s body to receive the money, while B will want the person with B’s body to be tortured. This is because the procedure appears to involve A and B swapping their bodies. Now consider the second version of the thought experiment, in which A is told that the following will happen to him: he will lose all of his memories and instead be equipped with different ones, and all of his other psychological characteristics will similarly be replaced. Finally, after this procedure, he will be tortured. It is to be expected that A will not like these prospects at all, and, as Williams argues, it will hardly console him if he learns that someone else will get his memories and other psychological traits and, on top of this, be handed $100,000. If you think about it, however, this

---

1. A note about terminology: throughout the paper, by ‘constitution’ I mean the relation, whichever it is, that people are after when they try to develop an account of personal identity. This relation should be at least as strong as supervenience. Whether it needs to be stronger, and if so, what exactly that relation might be, will be left open.
is surprising. The same procedure that A desired on the basis of the first description is now feared by him.

It is not obvious what lessons are to be drawn from Williams’s thought experiment. One of its features that I would like to highlight, however, is that with respect to the same scenario neither A’s survival in B’s body nor his survival in A’s body involves a conceptual or a priori contradiction. Following common usage, I call propositions that share this feature ‘conceivable’. Accordingly, a proposition is conceivable if it cannot be ruled out on conceptual or a priori grounds and can thus be consistently entertained. That both propositions at hand are conceivable in this sense can be shown as follows. Let us call the description of the first version of the case D. Our judgment about D was that A goes with his mind. The second description can be phrased as ‘D and A stays with his body’. Since ‘A goes with his mind’ and ‘A stays with his body’ are mutually exclusive, this leads to the following dilemma: if D were to conceptually or a priori necessitate that A goes with his mind, then the description of the second version of the case would be inconsistent. If, on the other hand, D were to conceptually or a priori necessitate that A stays with his body, then our judgment about the first version of the case would be inconsistent. To avoid this dilemma, we have to accept that both of the propositions in question are conceivable.

2.2. The Conceivability of Logical Complements

Williams’ scenario is surely a curious one. I would like to argue, however, that the feature just identified is a general feature of our judgments about possible (including actual) cases involving personal identity. Assume, e.g., that you believe you would survive teletransportation (see Parfit 1984: 200ff.). Even then, I claim, it is nevertheless conceivable for you that you would not. Likewise, it would not be inconsistent to hold that the person in the body of a cobbler in Locke’s thought experiment is the cobbler. Now take a much more extreme case: suppose that, as is likely, there is no psychological\textsuperscript{2} or physical continuity or connectedness whatsoever between me and Hatshepsut. Nevertheless, there is no inconsistency in the idea that I was Hatshepsut in a former life, i.e., that she and I are the same person. Similar claims about cases of rebirth have in fact often been made. While they may seem incredible to many, it is not apparent that they are inconsistent.\textsuperscript{3} Thus, as far-fetched as the thought that I am identical to

\textsuperscript{2} By ‘psychological’, I mean features of the kind mentioned above: memories, propositional attitudes, phenomenal experiences, etc., but not involving (putative) mental substances or anything similar. The latter are discussed below.

\textsuperscript{3} One might argue that such claims only seem to be consistent and involve some deep inconsistency. This option is discussed in Section 3.2.1.
Hatshepsut may be, it is nevertheless conceivable. I believe that with respect to all descriptions of cases concerning the diachronic identity of persons that do not explicitly state the facts about the identity of those involved, it is conceivable both that identity obtains and that it does not. It is admittedly hard to establish such a general claim. But the fact that it applies even to cases where there is no physical or psychological continuity or connectedness at all lends at least some support to the claim. Below, I show that there are also general reasons for supposing that this thesis is true. The precise version of the thesis regarding the conceivability of complements that will concern me in the remainder of this paper is arguably slightly weaker. It is as follows:

(CC) For the vast majority of hypothetical cases about personal identity, logically complementary propositions about the identity of those involved are conceivable.5

CC is strongly suggested by the discussion above. On the face of it, CC is also rather striking, since it clashes with a number of popular theses concerning our judgments about hypothetical cases. Thought experiments often only aim to establish a possibility claim, e.g., the claim that it is possible that there is physical continuity without personal identity. They try to achieve this by presenting a possible case that is supposed to have the features in question. But now consider our judgments about such cases, e.g., the judgment that the scenario in question represents a case of physical continuity without personal identity. Many philosophers have argued that these judgments, and, more generally, our ‘rational intuitions’ bear some intimate relation to necessity. For example, George Bealer (1998), Joel Pust (2000: 46) and Jonathan Ichikawa and Benjamin Jarvis (2009) say that such judgments are associated with a ‘feeling of necessity’; as Bealer puts it, they “present themselves as necessary” (1998: 207). Others argue that they are

4. It might be claimed that it is only conceivable because the scenario invokes the first-person perspective. (A number of philosophers claim that similar phenomena are connected to first-person thought, e.g., Blackburn 1997: 131; Nagel 1986: 33; Ninan 2009; Williams 1970.) I do not deny that our judgments may be susceptible to influence depending on whether a scenario is presented from the first-person or the third-person perspective. (However, see Nichols & Bruno 2010 for some empirical evidence to the contrary.) But the example at hand does not depend on the first-person perspective. Consider, e.g., the question of whether Hatshepsut and Boris Becker, who, let us assume, have no psychological or physiological continuity or connectedness, are the same person. This seems just as conceivable. Hence, the conceivability of complements is independent of the perspective from which one considers a scenario.

5. Could an analogous thesis concerning thought experiments about material objects be held as well? While such a claim does not seem obviously false, I do think that the case regarding personal identity is much stronger. I also believe that if the analogous thesis turns out to be correct, it can be explained in a similar way as CC, namely by appeal to the idea that material objects might have hidden non-physical essences that are picked out by our relevant concepts, if they exist.
necessity judgments or are at least about necessary truths (e.g., BonJour 1998; Ichikawa & Jarvis 2009; Sosa 2007). It is difficult to square these views with the fact that the negation of the intuitive judgment is also conceivable. First of all, if a proposition presents itself as necessary, then one would expect its negation to somehow present itself as impossible. (Compare: ‘It must be that p but it might be that non-p’ is at least infelicitous.) Further, if the judgments in question are necessarily true or are even explicitly judgments about necessity, one faces a problem if one also believes, as many do, that conceivability is evidence for possibility (e.g., Yablo 1993) or even entails possibility (e.g., Chalmers 2002). For if about half of the propositions about hypothetical scenarios concerning the identity of persons that we find conceivable are in fact impossible, then conceivability can hardly count as evidence for possibility in this domain. It is in any case natural to think that rational intuitions about necessary truths should also be conceptually or (broadly) epistemically necessary. As I suggested above and will argue in more detail in Section 3.1.2 below, our judgments in other philosophical areas, e.g., concerning an analysis of causation, knowledge, or free will, are much more in line with such a picture. CC thus confirms the suspicion expressed at the outset, namely, that judgments about scenarios concerning the identity of persons do not reveal conceptual truths, and calls for an explanation.

Not everyone accepts the account of thought experiments that I assumed above. To someone who believes that thought experiments are not supposed to reveal conceptual truths and that judgments about them bear no intimate relation to necessity, CC might not seem all that surprising. A philosopher of this type should expect it to be trivial to explain CC. However, the following discussion will demonstrate that no easy explanation of CC is available.

3. Explanations from the Metaphysics and the Epistemology of Personal Identity

It is natural to try to explain CC by appealing either to features of the facts about identity (in the hypothetical scenarios or in general) or to the nature of our epistemic access to these facts, or to a combination of these. The explanations I consider in this section are all of this broad kind. Let me divide them into two categories, according to an underlying metaphysical premise:

1. The case descriptions do not metaphysically determine the facts about personal identity.
2. The case descriptions do metaphysically determine the facts about personal identity.
Let me add a few words about how these claims are to be understood. The explanations in category 2 do not merely assume that identity facts supervene on the facts given by the descriptions. That would be compatible with there being no (determinate) facts about personal identity. Rather, the underlying assumption is that there are facts to be determined and that they are, moreover, determined by the descriptions. The explanations of category 1 thus involve either the claim that there are no determinate facts about personal identity in the described scenarios or the claim that these facts are not determined by the descriptions. A proponent of one of these explanations can therefore hold that complementary propositions about the identity of those involved in a hypothetical case are both possible, which can then account for their conceivability. These explanations thus deny that judgments about hypothetical scenarios concerning personal identity express metaphysically necessary truths. By contrast, explanations of category 2 entail that one of the two complementary propositions must be impossible. Accordingly, they reject the connection between metaphysical and a priori or conceptual necessity.

3.1. The Case Descriptions Do Not Metaphysically Determine the Facts about Personal Identity

3.1.1. The Cases Are Semantically Indeterminate

It is natural that at least some of the hypothetical scenarios discussed in the debate over personal identity represent cases of semantic indeterminacy (see, e.g., Quine 1972: 490). Nearly all of our concepts have vague application conditions, so it is reasonable to assume that our concept of personal identity is no exception. Moreover, if a case is semantically indeterminate, then neither the presence nor the absence of personal identity is conceptually impossible. This fact would then explain why both options are conceivable. However, there are a number of reasons to doubt that the appeal to vagueness yields a promising explanation of CC.

For a start, it is difficult to make sense of the idea that there are indeterminate cases concerning personal identity. The claim that, for example, it is indeterminate whether I still exist after teletransportation seems bizarre. Accordingly, a number

---

6. As I explain below, it will be useful to weaken this assumption slightly to the claim that descriptions of hypothetical scenarios usually determine the facts about personal identity.

7. I leave out the most radical views in either category, i.e., anti-realism about personal identity and the view that facts about the identity of persons are unknowable. However, I discuss anti-realism in Section 4.2.

8. This way of putting things presupposes a semantic account of vagueness. For a proponent of the epistemic theory of vagueness (e.g., Williamson 1994; Sorensen 2001), an explanation invoking vagueness would fall into category 2.
of philosophers have argued that questions about the persistence of persons can never be indeterminate (e.g., Butler 1844; Chisholm 1969; Reid 1785/2002; Williams 1970: 174–175). Even Parfit, who believes that such questions sometimes have no determinate answer, concedes that this idea is highly counterintuitive (1984: 214). This is not a decisive objection against the explanation at hand. We may eventually be forced to accept that there can be indeterminacy concerning personal identity, counterintuitive as it may seem. We should, nevertheless, take this intuition seriously—and, as I show below, the intuition that the persistence of a person admits of no indeterminacy can be explained within the account I propose. Furthermore, when it is indeterminate whether p, we do not usually find both p and not-p to be conceivable. If confronted with such a case, we usually feel that it makes no sense to continue asking whether, say, a specific person is a child or whether a particular arrangement of grains of sand is a heap. But in judging that both the obtaining and the non-obtaining of personal identity are conceivable, we seem to assume that there is an open question to be answered. In response, one might claim that the fact that we continue to treat such questions as genuine even in indeterminate cases attests only to the fact that we intuitively believe that there are no such cases. Since I already put forth this intuition above, we should be careful not to double-count this objection against the envisaged explanation from semantic indeterminacy. In any case, however, the final and most straightforward objection stems from the observation that the conceivability of complementary propositions applies to most, if not all, cases concerning the identity of persons. Plainly, not all of these cases can be indeterminate if there are to be any conditions under which the concept of personal identity is correctly applied. Therefore, CC cannot be explained by appealing to vagueness.

3.1.2. The Case Descriptions Leave out Relevant Physical or Psychological Facts

The descriptions of hypothetical scenarios never, or almost never, refer to a unique possible world. In this and the following section, I consider accounts that try to explain CC by appealing to this kind of incompleteness of the descriptions. These accounts thus agree with the account considered in the previous section in saying that there is no fact of the matter concerning the identity of the persons involved in the scenarios. However, the two explanations that we now consider do not locate the source of this indeterminacy in our concept of personal identity, but rather in the scenario descriptions.9

---

9. Kathleen Wilkes’s critique of the use of thought experiments in the debate about personal identity is based on the same idea: She argues that the descriptions of many hypothetical cases that have been discussed leave out background conditions that are relevant for evaluating them (Wilkes 1988: 8–48).
Thought experiments concerning personal identity usually mention physical and psychological facts. As in Williams’s thought experiment, they typically describe what happens to people’s bodies, including their brains, and to their psychology—their memories, traits of character, etc. The thesis to be considered in this section is that the descriptions of the scenarios leave out relevant facts of these kinds and that, as a result, the set of worlds picked out by such a description may contain both worlds where identity obtains and worlds where it does not. In this case, the descriptions do not necessitate the facts about personal identity, and CC can be explained straightforwardly: if a scenario description does not metaphysically determine whether identity obtains or not, then both options are possible. It is thus unsurprising that they are both conceivable, as is stated by CC. Note that the viability of this proposal crucially depends on the following two assumptions: firstly, that thought experiments concerning the identity of persons leave out relevant physical or psychological information, and secondly, that this is what is responsible for the conceivability of complementary propositions. However, there are good reasons to contest both of these assumptions. It is true that what is literally said in the descriptions of hypothetical scenarios often leaves out features that are potentially relevant for the evaluation of the cases. Indeed, it is reasonable to assume that this applies to thought experiments across many or all philosophical subject matters. However, as will become apparent in what follows, this problem goes deeper in the debate on personal identity than elsewhere.

In The Philosophy of Philosophy, Timothy Williamson claims that typical descriptions of Gettier cases do not necessitate that there is justified true belief without knowledge (2007: 185). He supports this claim by pointing out that the descriptions of Gettier cases allow for deviant—i.e., unintended—interpretations. Take the first of Edmund Gettier’s original cases (Gettier 1963). Suppose that as Smith deduces ‘The man who will get the job has ten coins in his pocket’ from ‘Jones will get the job’ and ‘Jones has ten coins in his pocket’, he learns that his company’s president—who told him that Jones would get the job—likes to joke about such matters. Just a few weeks ago, for example, he had told Brown that he would never be promoted, only to promote him two days later. This is consistent with Gettier’s description of the case. On this reading, however, the case fails to serve as a counterexample to the classical analysis of knowledge. Since Smith acquires defeating evidence against his belief that Jones will get the job, he is no longer justified in believing it. Hence, he is not justified in believing that the man who will get the job has ten coins in his pocket, either. Notice, however, that if one interprets the scenario in a certain way (i.e., the intended way), then it is plausibly necessary that it is a case of justified true belief without knowledge.11

10. This is a modified version of a case invoked by Williamson (2007: 185).
11. Ichikawa and Jarvis’s (2009) proposed solution to the problem raised by Williamson rests on this idea.
Presumably, most people who are confronted with the case do not even realize that such unintended readings exist and automatically interpret it as intended by Gettier. Moreover, such deviant interpretations can be excluded by supplementing the description. In the present case, for example, one can stipulate that Smith at no point acquires evidence that undermines one of the relevant beliefs. Williamson seems to think that any list of additions to scenario descriptions to exclude deviant readings is bound to be incomplete. However, this claim has been disputed (Grundmann & Horvath 2013). And even if Williamson is correct, it is nevertheless plausible that we know, for any given deviant interpretation, what would have to be added in order to exclude it.

Now consider, by contrast, thought experiments on personal identity. For a start, it is not credible that the conceivability of complementary propositions would vanish if further physical or psychological information were to be added. Take Williams’s thought experiment, for example. The scenario is certainly not described in complete detail. But it is not obvious that if further physical and psychological details were to be specified about the case, one of the versions would become inconsistent. The same applies to Locke’s scenario involving the prince and the cobbler, to Parfit’s teletransportation case, and so on. It is implausible that if one were to add further physical and psychological information to the descriptions of the scenarios, they would conceptually necessitate particular judgments about the identity of those involved. Next, consider my Hatshepsut case. I stipulated that there is no physical or psychological connectedness or continuity between her and me at all. It is unclear what kind of additional physical or psychological information there could even be that would render it inconceivable that we are the same person.

I conclude that, firstly, even if physical and psychological information were to be added to the descriptions of hypothetical scenarios concerning personal identity, this would not make the conceivability of complementary propositions go away, and secondly, not all descriptions of cases to which CC applies leave out relevant physical or psychological facts. The proposed explanation considered in this section therefore fails; CC cannot be attributed to the fact that the scenario descriptions leave out relevant physical or psychological facts.

3.1.3. The Facts about Personal Identity Are Not Determined by Physical and Psychological Facts

Although the attempt to explain CC that we have just considered fails, it suggests another explanation. Perhaps the descriptions of hypothetical scenarios about the identity of persons are indeed incomplete, in the sense of not determining the facts about the identity of those involved, but this is not because they leave out relevant physical or psychological features but rather because physical
and psychological features are insufficient in principle to determine facts concerning the identity of persons. This explanation thus involves the claim that facts about our persistence do not supervene on physical or psychological properties. In the following, I refer to any position that is committed to this premise as anti-reductionist and to any view on which identity facts do supervene on the conjunction of physical and psychological properties as reductionist.

The explanation that anti-reductionism provides for CC is similar to the explanation considered in the previous section: if the descriptions of scenarios do not determine the facts of identity, then both members of a pair of complementary propositions about the identity of persons are possible, which suffices to explain the fact that they are conceivable. At the same time, the current proposal is compatible with, and even predicts, the fact that the addition of further physical and psychological information to the scenario descriptions changes nothing about the conceivability of complementary propositions. It thus seems as though the anti-reductionist account provides a compelling explanation of CC. In addition, an anti-reductionist can also explain our intuition that there can be no indeterminate cases concerning personal identity by holding, e.g., that the substances constituting our identity are indivisible and therefore admit of no degrees. However, there are well-known reasons to doubt anti-reductionism. I will not take sides in the debate between reductionists and anti-reductionists here, but the fact that anti-reductionism is rejected by the vast majority of philosophers suffices to make it unattractive as a premise in the explanation I am after. Below I argue that it is no accident that anti-reductionism fits our intuitions about hypothetical cases so well. At the same time, the explanation of CC I offer does not commit me to anti-reductionism. Before I turn to my own account, however, I still need to consider the potential explanations from category 2. They are based on the assumption that the descriptions of hypothetical scenarios determine facts about personal identity.

3.2. The Case Descriptions Do Metaphysically Determine the Facts about Personal Identity

The next accounts to be discussed assume that there are determinate facts about the identity of persons and that these facts supervene on those given by the descriptions of hypothetical scenarios. This entails that facts about personal identity supervene on physical and psychological facts. But this is a thesis that most philosophers are committed to anyway. As it stands, however, the claim shared by the category 2 explanations is too strong in another sense. For a number of reasons, it would be hopeless to argue that the descriptions of hypothetical scenarios always determine facts about personal identity. Therefore, I will take the
explanations I consider to be committed only to holding that these descriptions metaphysically determine facts about personal identity usually, or in most cases.

3.2.1. Cognitive Limitations

According to accounts of the type at hand, the descriptions of hypothetical scenarios necessitate the facts about identity, whatever these may be. Therefore, when we judge complementary propositions about a hypothetical case to be conceivable, one of these propositions is in fact impossible. The current account tries to explain this by appealing to our cognitive limitations. That is to say, it claims that if we were smarter, we would recognize the impossibility of the relevant proposition and it would no longer be (or seem) conceivable.

The appeal to our cognitive limitations does not yield a particularly attractive position, however. The view is admittedly difficult to refute. Maybe there is a hidden contradiction in the thought that I am Hatshepsut and in the thought that I do or that I do not survive teleportation. But on the face of it, such cases do not seem more intricate than hypothetical cases in other areas. So why should we fail to spot the impossibility here, but not there? It is completely unclear why one should think that CC would go away with better reasoning. Let me thus turn to another and, in my view, more promising account.

3.2.2. A Posteriori Necessities

The work of Saul Kripke and Hilary Putnam showed that there are necessary truths that can only be discovered empirically, e.g., ‘Water is H$_2$O’, ‘Hesperus is Phosphorus’, and ‘Light is a stream of photons’. So perhaps truths of the kind ‘If such-and-such were the case (where ‘such-and-such’ stands for a description of a hypothetical scenario), then personal identity would / would not obtain’ are also a posteriori necessities.

A posteriori necessities can arise whenever concepts whose application conditions are not a priori accessible to a subject are involved. It is widely accepted that this applies to natural kind concepts and to proper names. The current proposal is that it applies to the concept ‘same person’ as well. David Wiggins makes a related claim: he argues that personhood is akin to a natural kind, and thus the concept ‘person’ functions just like other natural kind concepts (Wiggins 1976). This is not quite the same claim as the one currently under discussion. However, if, as many believe, every person is necessarily a person, then the semantics of ‘person’ is likely to affect that of ‘same person’ as well. The following passage from *The Meaning of ‘Meaning’* suggests that Putnam also thinks that necessary truths about personal identity are a posteriori:

*Ergo* • vol. 3, no. 1 • 2016
There is a joke about a patient who is on the verge of being discharged from an insane asylum. The doctors have been questioning him for some time, and he has been giving perfectly sane responses. They decide to let him leave, and at the end of the interview one of the doctors inquires casually, ‘What do you want to be when you get out?’ ‘A teakettle.’ The joke would not be intelligible if it were literally inconceivable that a person could be a teakettle. (Putnam 1975: 244)

It would be unremarkable that a person’s becoming a teakettle is conceivable if it were also possible. The fact that Putnam stresses its conceivability thus suggests that he believes that a person cannot be numerically identical with a teakettle. This interpretation is supported by the context of this passage, which is a discussion of his claim that certain words tend to assume a semantics analogous to that of natural kind terms. Accordingly, the conceivability of a person becoming a teakettle is no more informative than that of water being XYZ and merely reflects the fact that it involves no a priori contradiction. Hence, necessary facts about the identity of persons are a posteriori.

As the preceding remarks already indicate, the thesis I am now considering can explain CC straightforwardly. Complementary judgments about a hypothetical case are jointly conceivable because neither is conceptually impossible. Nevertheless, one of these propositions is metaphysically necessary and the other is thus impossible. The downside of this proposal is that the premise that there are such a posteriori necessities is quite bold. Let me give just one reason to doubt it. Typically, a posteriori necessities can be derived a priori from some piece of contingent information. To give a simple example, from the assumption that streams of photons are the dominant cause of light sensations, one can conclude a priori that light is a stream of photons.12 In the case of personal identity, this contingent information could be physical or psychological, or it could be another kind of information. If it is the latter, then the account at hand is just a version of anti-reductionism, which was discussed above. If it is the former, then facts about personal identity would have to be derivable a priori from complete physical and psychological information. This entails that if we had such complete information, it would no longer be conceivable that I am (or maybe that I am not) Hatshepsut. But, as I argued above, this seems implausible.

What I said in my discussion of anti-reductionism applies here as well: My aim is not to refute the position at hand. But the mere fact that it rests on problematic metaphysical assumptions likely to be rejected by most makes it unsat-

---

12. Particularly hard-nosed semantic externalists might reject this claim. However, note that the conclusion follows from Kripke’s assumption that ‘the dominant cause of light sensations’ is the reference-fixer associated with ‘light’ (Kripke 1980: 131).
isfactory as an explanation. Let me note again that the explanation I propose below avoids any such assumptions.

4. Conceptual Considerations Concerning Substance and Identity

None of the potential explanations of CC considered in Section 3 proved to be satisfactory. Among these, the anti-reductionist account and the one invoking a posteriori necessities seem most in line with our judgments about the hypothetical scenarios. Both of these accounts postulate that personal identity is constituted by some kind of hidden essence. But they also thereby share their weakest point, for it is questionable whether such an essence exists. In my view, there is nevertheless something right about these accounts. However, as I argue in what follows, in order to draw the correct conclusions from CC, it is necessary to recognize that our intuitions primarily inform us about the concept of personal identity. It is therefore a mistake to move from the characteristics of these intuitions directly to metaphysical conclusions. Such a mistake has occasionally been made by anti-reductionists. A number of them have pointed out that their view fits our intuitions about personal identity particularly well (e.g. Johnston 1987: 70; Nida-Rümelin 2010; Swinburne 1973). Martine Nida-Rümelin (2012) even bases her case for an anti-reductionist account on a number of principles quite similar to CC. As I argued above, however, the inference from CC to the existence of irreducible identity facts is insufficiently founded. Generally speaking, it is hard to believe that we have intuitive access to the existence of mental substances or similar stuff. In contrast, opponents of anti-reductionism often grant that we have such intuitions but usually dismiss them completely, precisely because there are powerful theoretical considerations in favor of reductionism. It is my belief that this view is also not the correct one.

4.1. CC Explained

Suppose that there are mental substances of the kind postulated, for example, by Descartes. Suppose further that these substances bear intimate nomic, yet metaphysically contingent, relations to the psychological states of human beings. It is intuitively plausible that if such a scenario were actual, these substances would constitute personal identity. A number of anti-reductionists are committed to this claim. Reductionists can also easily accept it. They only need to oppose the view that there are mental substances, not that they would constitute personal
identity if they existed. I consider this latter fact as one about our concept of personal identity, but this is not crucial for the account I want to defend. One could also hold that it is a metasemantic fact about the reference of the concept of personal identity in hypothetical circumstances. A friend of the idea of reference-magnetism could argue that such fundamental substances, if they exist, are excellent referential candidates. The only point that matters here is that the conditional ‘If there are mental substances, then they constitute personal identity’, or some other conditional in its vicinity, is plausibly true.

I think that the scenario I just outlined is one in which anti-reductionism is true. But since intuitions may vary, I should note that the explanation of CC I am about to offer does not depend on this particular scenario. There are countless other ways in which mental substances, souls, or other kinds of things that are neither physical nor psychological might have constituted facts about personal identity. Alternatively, there may be scenarios in which these facts are primitive. All my argument requires is that there is some conceivable way in which personal identity could have been constituted by something other than physical and psychological properties. To deny this, one would have to argue that anti-reductionism is not only false but inconsistent. In light of the pretheoretical attraction of anti-reductionist views and their theoretically informed endorsement by numerous philosophers, this would be a very bold claim to defend. It seems much more likely that among the many potential anti-reductionist accounts, one will find at least one that might have conceivably been correct. In the following, I show that this assumption is already sufficient to explain CC. The argument is based on the scenario described above, in which there are mental substances. But as I just mentioned, any other kinds of non-psychological and non-physical stuff or properties that might constitute personal identity would do as well. So when I speak of mental substances, this should be considered as a place-holder for anything that could potentially serve as a truth-maker of an anti-reductionist account.

Furthermore, in the scenario sketched above, I assumed that there are close connections between mental substances and the psychological states of human beings. I do not have a settled view on whether such connections are required for mental substances to constitute personal identity, and if so, how strong they

---

13. Some reductionists have argued that even if there were mental substances, they would not constitute personal identity (Locke 1694/1975: 332; Parfit 1984: 228). But in the scenarios they consider, there is no relation between those substances and our mental lives.

14. It certainly cannot be denied that anti-reductionism has seemed conceivable to a great number of people even on a considerable amount of reflection. Even if one believes that anti-reductionism is incoherent, its apparent conceivability is sufficient to explain the apparent conceivability of logically complementary propositions about personal identity. (See also Footnote 12.)
would have to be. Moreover, these issues are not essential to my argument. So when I speak of the existence of mental substances, this should be read as implicitly including their being connected to our mental lives, insofar as this is required. The argument will go through on any such reading, as long as there is one conceivable way in which mental substances constitute personal identity.

Let me first state the bare argument and then walk you through its steps.

(1) If there are mental substances, then they constitute personal identity. [Premise]

(2) Conceivably, there are mental substances. [Premise]

(3) If personal identity is constituted by something other than physical and psychological facts, then the totality of physical and psychological facts is compossible with complementary facts about personal identity. [Premise]

(4) Conceivably, the totality of physical and psychological facts is compossible with complementary facts about personal identity. [(1), (2), (3)]

Now let i and not-i be two arbitrary conceivable facts about personal identity.

(5) Conceivably, given the totality of physical and psychological facts, possibly i. [(4)]

(6) Conceivably, given the totality of physical and psychological facts, possibly not-i. [(4)]

(7) All the information given in the descriptions of hypothetical cases that is potentially relevant for the identity of persons is physical or psychological. [Premise]

(8) With respect to hypothetical cases about personal identity, logically complementary propositions about the identity of those involved are conceivable. [(5), (6), (7)]

15. Does the fact, discussed above, that it is conceivable that I am Hatshepsut, despite there being no physical or psychological continuity between us, not suggest that no such connections are required? Not necessarily, for the following two reasons: firstly, the Hatshepsut case is compatible with there being a mental substance that is first intimately connected with the psychological states of Hatshepsut and then with those of me; secondly, the case is compatible with all other mental substances in the world being exclusively linked with one body and one personality.

16. The conditionals occurring in the argument are all to be construed as strict conditionals.
Premise (1) was just argued for; (2) seems obvious enough. Premise (3) should also be uncontroversial: if identity facts are not constituted, and thus not determined, by physical and psychological facts, then the totality of physical and psychological facts should also be compossible with either one of a pair of complementary personal identity claims.

Next, the step from (1)–(3) to (4) depends on the assumption that conceivability is closed under entailment. This is a very natural assumption, if one abstracts away from a subject’s cognitive limitations. The type of conceivability involved is thus akin to David Chalmers’s ‘ideal conceivability’ (2002). The step from (4) to (5) and (6) also depends only on the same closure principle. (7) is another plausible premise that has been assumed throughout much of the preceding discussion. Notice that it does not state that only physical and psychological information is relevant for the identity of persons, but only that the descriptions of cases that have been put forth in the debate do not mention any potentially relevant further facts, e.g., about mental substances. This applies to thought experiments presented by proponents of all types of accounts of personal identity, including anti-reductionists, who often use such hypothetical scenarios to reveal problems of reductionist accounts (e.g., Swinburne 1973).

In order to get from (5)–(7) to (8), one once again needs the previously mentioned closure principle. Additionally, one has to move from the conceivability of the possibility of the conjunction of complete physical and psychological information and an identity fact to the conceivability of this conjunction simpliciter. The general principle that would legitimate this step—‘If conceivably, possibly p, then conceivably, p’—is prima facie plausible: if one can conceive of a situation in which p is possible, then why should one not be able to conceive of one where p obtains? It could be objected, however, that a priori contingent falsehoods provide counterexamples to this principle. Assume that ‘Heat is the cause of heat sensations’ is a contingent a priori truth. Then ‘It is possible that heat never causes heat sensations’ is clearly conceivable. But arguably, since ‘Heat never causes heat sensations’ is conceptually impossible, it is inconceivable. Nevertheless, the step from (5)–(7) to (8) is justified, since the scenario in which personal identity is constituted by mental substances can be used to show that the possibility at hand is not a conceptual impossibility. If either ‘The physical and psychological facts are such-and-such and i’ or ‘The physical and psychological facts are such-and-such and not-i’ were a contingent a priori truth, then one would have to be able to tell a priori that its complement does not actually obtain. Now consider the scenario at hand, in which there are mental substances. There is no

---

17 One might object that my considerations at the beginning of this paper were insufficient to establish that with respect to hypothetical cases about personal identity, complementary propositions are ideally conceivable. However, note that the notion involved in CC can be construed as being weaker than ideal conceivability, meaning that (8) entails CC.
way to know a priori that it does not obtain, and within this scenario, there is no way to determine the facts about personal identity on the basis of physical and psychological facts alone. Thus, from the conceivability of there being mental substances, one can infer (8). Now note that (8) entails CC. Therefore, by inferring (8) from the premises given above, we have in effect explained CC.

Let me restate the main point about this explanation of CC: The crucial insight was that if there is something akin to mental substances (and given certain further conditions), then they constitute personal identity. Given this, CC is to be expected, due to the fact that we can conceive of a scenario in which there are mental substances and that furthermore, in such a scenario the facts about personal identity float free of those mentioned in the descriptions of thought experiments.

There is a growing body of research that suggests that humans are prone to ascribe hidden essences to various kinds and individuals (see, e.g., Gelman 2003). The latter thesis is often called psychological essentialism. Since the explanation that was just argued for crucially involves the idea that the identity of persons might be constituted by some kind of hidden essence, it is natural to think that an appeal to psychological essentialism could also yield at least a partial explanation of CC. Let me therefore briefly discuss the bearings of psychological essentialism on the issues at hand. I will argue that the thesis can be taken to support the explanation of CC I am proposing. To the extent that my explanation goes beyond what is suggested by psychological essentialism, I maintain that the extra commitments I take on are well-founded.

Psychological essentialism is based on a substantial amount of evidence suggesting that we have a strong tendency to form (implicit) beliefs ascribing hidden and intrinsic essences to individuals and kinds. This tendency seems to be present already in children at a very young age. The beliefs in question serve a number of practical and theoretical purposes. For example, they are employed in predicting surface features as well as the behavior of individuals. It should be noted that the majority of existing research on psychological essentialism does not directly bear on the question of whether we believe that the identity conditions of individuals depend on hidden essences. For said research, as far as it is concerned with beliefs about the essences of individuals, usually addresses the question of whether subjects tacitly assume that individuals have essences qua being a member of a particular kind (see also Gelman 2003: 12–13). So, for example, when an individual is believed to have an essence that is common to all members of a kind, this belief does not immediately imply that the individual would cease to exist if that essence was somehow altered, but only that it would

---

18. See the preceding footnote.
19. Leslie (2013) provides an overview of the state of the art.
cease to be a member of the relevant kind. Nevertheless, it is natural to think that creatures that have a tendency to ascribe essences to individuals qua being members of a particular kind may also tend to ascribe essences to them qua being a particular individual.

If one applies this latter assumption to the case of personal identity, it may lead to the thesis that CC is to be explained by our implicit belief that the identity of persons is constituted by a hidden essence that goes beyond our physical and psychological makeup. In order for this implicit belief to be consistent, it must be conceptually possible that personal identity is constituted by a non-physical essence. As we saw above, the latter is all that the explanation of CC I offer requires. The only way in which that explanation is stronger than the one based on psychological essentialism is in that it presupposes that the content of said belief is in fact consistent. However, this additional assumption does not come with great costs. For, firstly, one usually needs good reasons to ascribe inconsistent beliefs to subject, and these are not apparent here. Secondly, it is plausible that creatures that tend to ascribe essences to individuals will also tend to form concepts of individuals that are capable of picking out these essences, if they exist. In this way, psychological essentialism supports my thesis concerning the concept of personal identity.

At the same time, a direct explanation of CC in terms of our implicit beliefs in essences that does not appeal to facts about the concept of personal identity needs to assume that such beliefs survive all kinds of theorizing and even persist in subjects that explicitly deny that there are such essences. While it might be possible to argue for this, it is preferable not to be committed to this assumption. Therefore, while it can be argued that the thesis of psychological essentialism can make a contribution to the explanation of CC, that contribution should be considered to be indirect: our tendency to ascribe essences to individuals contributes to the formation of a concept of personal identity that is such that it picks out a certain kind of hidden essence, if it exists. By the time theoretical reflection sets in, the concept in question is already in place. And as was shown above, the structure of that concept is sufficient to explain CC.

Above I proposed that if there are mental substances or similar things, then they constitute personal identity. At the same time, I have argued that this does not give us reason to believe that something like mental substances exists. So what if, as indeed most philosophers believe, there are no such things? Would our concept of personal identity then be empty? This would imply that there are no persisting persons and thus that a large part of our everyday discourse as

20. For considerations against the idea that the essence in question could be physical, see Section 3.2.2.
21. Compare also the considerations invoked in Section 3.2.1.
well as our social practices rests on a mistake. Fortunately, we need not acquiesce to such consequences. As I argue in the next section, there can still be persisting persons even if there are no mental substances or any other such thing.

4.2. Personal Identity without Substance

Nearly everyone believes that there are persisting tables, trees, and trains. Yet it has rarely been proposed that their persistence is constituted by some kind of hidden essence that goes beyond physical properties. Most people think that the fact that certain physical relations obtain is sufficient for the persistence of such objects through time. But if so, why should our own persistence require something like an underlying substance? Joseph Butler and Roderick Chisholm claim that ‘real’ identity that Butler calls “identity in the strict and philosophical sense” is incompatible with change in any of an individual’s parts (Butler 1844; Chisholm 1969).

They also both hold that personal identity is identity in a strict and philosophical sense, while mere physical objects are identical (through time) only in a “loose and popular” sense. Although I am skeptical about the motivation for introducing such a notion of strict identity, let me assume that Butler and Chisholm are right in claiming that there is such a concept that we believe to apply to personal identity. Even then, however, persons could still persist in just as strong a sense as tables, trees, and trains, in case that belief turns out to be false.

In my view, the idea that personal identity requires the existence of something like mental substances also conflicts with our intuitions. Consider a possible world in which only physical and psychological properties are instantiated. In some cases, consecutive instances of these properties stand in relations of physical and psychological continuity and connectedness. Intuitively, this is a scenario in which there are persisting persons. Notice that the majority of philosophers believe that the actual world is just like the one described, yet very few of them have concluded that there are no persisting persons.

One might object to the view I have just outlined by saying that it construes the concept of personal identity in a peculiar way: in one sense, it works very much like a natural kind concept, in that it purports to refer to a kind of hidden essence that underlies the superficial phenomena we observe. However, so the objection goes, in the case of natural kind concepts, if there is no such hidden essence this just means that the concept is empty. A concept of the kind I postulate would thus be highly exceptional, which casts doubt on my account. In my view,

---

22. For a similar view, see Reid (1785/2002).
23. Strictly speaking, psychological and physical features are not, or need not be, superficial phenomena. They are nevertheless ‘more superficial’—in the sense of being more readily accessible—than putative mental substances or the like.
this objection rests on a popular but nevertheless superficial understanding of the semantics (or metasemantics; see above) of natural kind terms. In the following, I argue that with respect to many such terms, successful reference does not require that there is a natural kind to be picked out.\textsuperscript{24}

Let me start with the example of ‘air’. For many centuries, air was believed to be one of four (or five) ‘classical’ elements. Surely, if there had turned out to be such an element or another type of natural kind, then ‘air’ would pick out this natural kind. Today we know that the stuff we breathe is only a heterogeneous mixture of gases. On the standard understanding of natural kind terms, when this was discovered, people should have concluded that there is no air, but this did not happen. Instead, ‘air’ is now taken to refer to the mixture of gases surrounding us. On the face of it, then, the semantics of ‘air’ is analogous to that of ‘same person’ as construed above: it is such that if there is an underlying natural kind, then ‘air’ refers to this natural kind. If, on the other hand, there is no such kind, then the term is nevertheless not empty.\textsuperscript{25}

Next, consider ‘water’. This term does refer to a natural kind. But suppose that when Lavoisier examined the liquid he called ‘water’ (or rather, ‘eau’), he had found that it is just a heterogeneous mixture of many different substances. Would he have thus discovered that there is no water? Alternatively, suppose that, astoundingly, we now discover that Lavoisier and everyone else after him were mistaken and that, in fact, the liquid in our rivers and lakes is not $\text{H}_2\text{O}$ but rather just such a heterogeneous mixture of substances. Would we thereby discover that there is no water? This seems implausible. Intuitively, what we would discover is that water is a mixture of different substances. If this is so, then our concept of water works similarly to our concept of air and the concept of personal identity according to my proposal. The main difference is that we know that ‘water’ picks out a natural kind.\textsuperscript{26}

I just argued that concepts such as ‘air’ and ‘water’ have the following structure: if there is something like an underlying hidden nature or essence, then it is picked out by these concepts. But if there is no such underlying nature, then they are nevertheless not empty. If this is correct, then our concept of personal identity, as I proposed to construe it, is not that exotic after all.

\textsuperscript{24} For a more comprehensive version of the argument, see Kipper (2012: 187–190).

\textsuperscript{25} Robert Stalnaker defends the same view about ‘air’. He also seems inclined to believe that an analogous claim can be made concerning ‘water’ (Stalnaker 2002).

\textsuperscript{26} David Braddon-Mitchell (2003) advocates an analogous view about the semantics of phenomenal concepts. He argues that if there are particular kinds of non-physical properties, then phenomenal concepts pick these out. If, however, as he thinks is the case, there are no such properties, then they refer to certain physical properties. Related views have been put forth by John Hawthorne (2002) and Stalnaker (2002). Similarly, Chalmers (2006) argues that there could have been fundamental, ‘perfect’ colors that would have been picked out by our color concepts. If, however, as seems highly likely there are no such things, color concepts refer to physical properties.
In this section, I have argued that the existence of persisting persons does not depend on the existence of mental substances. One may thus wonder what constitutes personal identity if there are indeed no such substances. In the case of ‘air’, it is plausible that the term refers to anything that bears certain superficial features, such as being gaseous and breathable. Something similar would apply to ‘water’ if it did not refer to a natural kind. One can think of these features as the reference-fixers or the stereotypes of ‘water’. So what could the relevant features be in the case of ‘same person’? Answering this question would likely amount to determining what constitutes personal identity, which is beyond the aims of this paper, but there are a great number of philosophers who think that there are no mental substances and who have developed theories of personal identity from this premise. Nothing in the account I have suggested militates against the idea that the correct view is among these. My analysis also leaves the precise modal profile of the concept of personal identity open: if there are mental substances, do they necessarily constitute personal identity or could it still have been constituted by something else? Conversely, if there are no mental substances, could they nevertheless have constituted the identity of persons if they had existed? Since the explanation of CC I have offered does not depend on these issues, I will not try to settle them here.

5. On the Proper Use of Thought Experiments about Personal Identity

According to the account I have just defended, facts about the identity of persons depend on whether something like mental substances exists. Usually, though, the descriptions of hypothetical scenarios do not state this. In some cases, this is appropriate. For example, there are thought experiments invoked against anti-reductionist accounts that would beg the question if it was stipulated that there are no mental substances. But in the majority of cases, there is no reason to omit facts relevant to personal identity. On the contrary, this omission has led to considerable confusion in the evaluation of hypothetical scenarios. My proposal concerning the construction of thought experiments is therefore to specify the relevant facts in the descriptions. So, for instance, for a proponent of a physical or a psychological view, it will usually make sense to consider scenarios in which there are no mental substances or the like. Accordingly, the descriptions of these scenarios should involve a phrase to the effect that these are all the facts relevant to the identity of those involved. In particular, it should be ruled out that the scenario contains mental substances, primitive identity facts, or anything like this. In the following, I apply this procedure to a number of cases in order to demonstrate how it can make a difference in the debate concerning the identity of persons.
First, consider my Hatshepsut example. I stipulated that there is no physical or psychological continuity or connectedness between Hatshepsut and me. Now assume further that these are all the facts relevant to the question of whether we are identical. With this additional premise in place, it seems much more plausible that the assumption that Hatshepsut and I are identical involves a contradiction.

Next, consider again the intuition that when it comes to the identity of persons, there can be no vagueness. According to the view I am defending, anti-reductionism is in a sense a part of our concept of personal identity. But beyond that, it is also plausibly a part of our pre-theoretical conception of personal identity. This is witnessed, e.g., by the widespread belief in rebirth, immortal souls, etc. across many cultures and religions. Further, as I noted above, at least on many versions of anti-reductionism, there can indeed be no indeterminate cases. It is therefore natural to think that our pre-theoretical conception of personal identity drives our intuition that personal identity admits of no indeterminacy.

To test this, consider Parfit’s thought experiment of the combined spectrum. It involves a surgeon who can manipulate your brain in any way she pleases at the cellular level. She can thereby induce any degree of change in your brain physiology and in your psychology and thus bring about any condition from a huge spectrum of cases (Parfit 1984: 236–237). At one end of the spectrum, no changes have been made. At the other end, the person bears no psychological resemblance to you and his or her brain bears no physiological resemblance to yours. Now add to this scenario the explicit assumption that anti-reductionism is false.27 There are no mental substances or anything like this. There are only cognitive systems with certain physical and psychological features standing in certain relations. Given this premise, many of the intermediate cases on the spectrum are plausibly borderline cases. Otherwise, one has to assume that the removal of one cell or one minuscule psychological change marks the line between persistence and death. After all, here there are no hidden features that could determine the identity facts one way or the other. This suggests that our reluctance to accept that there can be borderline cases concerning the identity of persons is due to our pre-theoretical anti-reductionist account of ourselves.28

My final example concerns the dispute between proponents of the physical and of the psychological view of personal identity. Consider again the second version of Williams’s ‘$100,000 torture’. Here, the description of the case explic-
itly states that the person in question is identical with the physically continuous person and not with the psychologically continuous one. From the fact that this is perfectly intelligible, some philosophers, including Williams himself, have concluded that the physical view cannot be so easily dismissed after all (see also Johnston 1987; Sider 2001). I am inclined to contest this conclusion, however. Williams states that each of the subjects' options in the scenario—to have the person with A’s body tortured or to have him given the money—involves a risk (1970: 180). He is right about this point. The reason for this, however, is that the subject does not know whether he is some kind of substance and if so, whether this substance stays with his body or goes with his mind. If one adds to the description of the case that there are no mental substances or anything like this, then it is much less obvious that Williams’s second version of it is really coherent.

6. Conclusion

In this paper, I have discussed a striking feature of our judgments about hypothetical scenarios concerning the identity of persons, expressed by CC: for the vast majority of these scenarios, complementary propositions concerning the identity of persons are conceivable. I first considered a number of potential explanations of CC that appealed to metaphysical considerations about identity facts or to the nature of our epistemic access to them. All of these explanations proved to be unsatisfactory, however. I then proposed an explanation of a different type that is based on a particular understanding of the concept of personal identity. According to this account, the concept is such that if there are mental substances, souls, or similar kinds of hidden essences, then it refers to these essences. If, however, there are no such things, then the concept is nevertheless not empty but may, e.g., pick out collections of psychologically or physically continuous and connected states of cognitive systems. I argued that this account is not only independently plausible, but that it furthermore permits explaining CC without committing oneself to any contentious metaphysical assumptions.

In my view, the fact that the structure of the concept of personal identity has not been properly understood has been the source of much confusion. The main reason for this is that the question of whether there are mental substances or the like is relevant to what constitutes the identity facts; yet people rarely specify whether there are such things when they appeal to intuitions about personal identity. I have thus proposed that in the construction of a hypothetical scenario about the identity of persons, one should explicitly state whether there are mental substances or anything similar. Finally, I have discussed several examples that suggest that this procedure can lead to re-evaluations of a number of crucial questions in the debate over personal identity.
Acknowledgements

I have presented versions of this paper at ANU, in Cologne, Essen, and Luxembourg. I would like to thank the audiences on those occasions for helpful comments. I am especially grateful for comments and discussions to Alma Barn-er, Ryan Cox, Thomas Grundmann, Joachim Horvath, Zeynep Soysal, Marius Thomann, Karsten Witt, and to two anonymous referees for this journal.

References


