Abstract for paper AAP Conference July 2001

The Limitations of Thought Experiments

Thought experiments have long been used as a tool of philosophical inquiry. They are an interesting and experimental way of encouraging our thinking to go beyond the mundane and conventional. But unless they are accompanied by, and indeed, outweighed by more direct and comprehensive forms of investigation, such as empirical research, or other types of philosophical inquiry, they can be misleading and deceptive. In this paper, I argue that this has occurred in the case of the psychological continuity criterion of personal identity.

Many theorists who support this criterion place too much weight on the thought experiments they use, and too little on other, more relevant forms of investigation. In most cases, the thought experiments used omit crucial details, and present obscure and incomplete scenarios. Reasoning is frequently deficient and inconclusive. Consequently, the view of mental states entailed is unreliable and controversial. Finally, these key thought experiments fail to demonstrate that personal identity is a matter of psychological continuity, to the exclusion of other forms of continuity. The conclusion that it is therefore questionable, and consequently cannot be taken for granted.

The Limitations of Thought Experiments

1 Introduction

This paper is about the problem of personal identity, the psychological continuity criterion, and the limitations of thought experiments. The problem of personal identity refers to the question of what makes a person at an earlier time the same person at a later time. The psychological continuity criterion is the answer to that question which states that personal identity over time is grounded in the continuation of a person's psychological states, to the exclusion of other factors, such as body or environment.

According to the Widest version, the cause of psychological continuity could be any cause. While not all proponents of the psychological continuity criterion explicitly state this condition, close scrutiny implies that it is ultimately entailed by the presuppositions on which the theory rests, namely that psychological continuity pertains independently of any other form of continuity.

Thought experiments refers to a form of argument which is frequently used to support the psychological continuity criterion. These issues are now detailed.

According to Derek Parfit and other psychological continuity theorists, it is the continuation of the mind, to the exclusion of the body, that grounds a person's continuity over time. Despite the body's irrelevance to personal identity or continuity, psychological continuity is not a theory of disembodied identity. The psychological continuity criterion involves embodiment, but is not reducible to it. This means that psychological continuity is independent of bodily continuity, but is always accompanied by it in some form or another. One consequence of this view is that minds could theoretically operate in different bodies without personal identity or continuity being compromised.

However, normal experience demonstrates that bodily continuity and psychological continuity occur together. Between birth and death, persons exist in the world where mind and body operate as a unit. Swapping of minds and bodies does not occur as a matter of course. Indeed, minds and bodies do not normally become separated from each other. This means that we cannot take for granted that, were a mind to be given a different embodiment, it would not be significantly affected.

In other words, because minds and bodies always occur together, we cannot know that minds would not be altered by different embodiment. Consequently, for the psychological continuity criterion to be taken seriously, it must establish that psychological continuity remains unaltered where bodily continuity is discontinuous.

Thought experiments are a common method of supporting the psychological continuity criterion. Often used to test theories, thought experiments usually comprise imaginary events, which, although unlikely, are logically possible (Wilkes 1988), p 2. Consideration of hypothetical scenarios is meant to reveal what is obscure in the normal course of events. A possible world is imagined in which extra-ordinary events occur. Different types of thought experiments are possible, those concerning philosophy normally being of scientific, metaphysical, moral, or epistemological interest.

Many personal identity theorists use thought experiments as tools of inquiry, and rely on their results to vindicate their particular view. Thought experiments concerning personal identity are usually scientifically based, and in this world are not generally realised. They present imaginary scenarios relating to aspects of personal identity considered to be problematic. This is particularly the case with supporters of the psychological continuity criterion.

Scenarios supporting the psychological continuity criterion usually involve the duplication of whole or part human beings, or the transfer between persons of parts of human beings, such as brains or parts of brains. The reader is asked to consider such an event, and is then questioned as to in what body or brain, or part thereof, personal identity is retained. The theorist then uses the answers to support a particular view of personal identity.

Intuitions elicited by thought experiments concerning psychological continuity are claimed to favour identity being retained in the person who possesses the appropriate psychological states, regardless of other factors, such as body or environment. These intuitions are then used to support the psychological continuity criterion.

Since its inception, the psychological continuity criterion has relied on certain key thought experiments. Most available versions adhere to common principles concerning the discreteness of the mind-body relation. In many instances, these principles are the primary foundation on which the case for psychological continuity is built.

However, a major problem exists with these key thought experiments, which throws their conclusions into doubt. Many of the scenarios are incomplete and opaque. Significantly crucial details are omitted, leaving relevant issues obscure or uncertain. Moreover, because the situations described rest on many complex and untested assumptions, we do not know whether they even represent genuine possibilities.

Many key features described deviate strongly from the norm. It is typically unclear whether these deviations are relevant to the issue under scrutiny. Consequently, evidence used to elicit outcomes is significantly incomplete. Thus, when used to investigate personal identity, these thought experiments carry little, if any weight. Conclusions drawn from them cannot be assumed to be correct. It follows that unless these conclusions are supported by other forms of inquiry and argument, they are inadequate to justify the psychological continuity criterion.

In defence of these points, several key thought experiments commonly used to support the psychological continuity criterion are investigated. Two purposes are served by this investigation. The first is to reveal areas in which the thought experiments are deficient. The second is to reveal issues relating to personal identity which need further inquiry. Based on this investigation, I claim that these key thought experiments fail to adequately justify the psychological continuity criterion.

I argue that they fail to address the range of issues involved, as they rely on obscure scenarios, at the expense of more relevant factors, such as accounting for mental content, the role of the body, or the nature of the self. I further argue that, in some important instances, the reasoning is faulty and inadequate, and does not yield the conclusions claimed. Specifically, it does not show that personal identity is encapsulated by the psychological continuity criterion, to the exclusion of other forms of continuity. Finally, I claim that the thought experiments' internalist view of mental content is flawed. Due to their detailed nature, the arguments for this last claim cannot be covered in the present paper, and are taken up elsewhere.

2 Thought Experiments Considered

Several issues pertinent to personal identity are raised in the following thought experiments, taken respectively from the work of John Locke, Sydney Shoemaker and Derek Parfit. Each is used in support of the psychological continuity criterion.

The Prince and the Cobbler

Like many recent theorists, Locke uses thought experiments when considering personal identity. A key tenant of Locke's philosophy is that substance, both material and immaterial, is unknowable. By contrast, consciousness is knowable, and is the means by which personal identity is preserved.

A problem for Locke is that he considers it possible that consciousness and substance could become separated:

But yet, to return to the question before us, it must be allowed, that, if the same consciousness (which, as has been shown, is quite a different thing from the same numerical figure or motion in body) can be transferred from one thinking substance to another, it will be possible that two thinking substances may make but one person. For the same consciousness being preserved, whether in the same or different substances, the personal identity is preserved (Locke 1959), 2.27.13.

But where consciousness and soul are conjoined, and the body becomes disconnected, it is consciousness which maintains the person's identity. To show this, we are asked to imagine that the soul and memories of a prince become manifest in the body of a cobbler. The question of which person is the prince permits the more general question about personal identity:

'For should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, everyone sees he would be the same person with the prince, accountable only for the prince's actions: but who would say it was the same man ?' (Locke 1959), p 457.

According to Locke, we would reject the person with the prince's body as being the prince, in favour of the person with the prince's mind. Even though confronted with a different body, with a different history, the mind and memories would have priority as the locus of identity. These points constitute part of Locke's vindication of the psychological continuity criterion.

Brown and Robinson

More recent thought experiments are similarly used to make points about personal identity. Due to increasing skills in neurosurgery, many thought experiments, such as this by Sydney Shoemaker, concern brain operations in which memories are transferred from one person to another:

Imagine the occurrence of a brain-operation in which the brain of one person, Robinson, was removed, and the brain of another person, Brown, was replaced into the empty skull. If the new person exhibited the characteristics and indicated the supposed memory-knowledge of the former person Brown, we might be inclined to think that the person Brown now inhabited the body of the former person Robinson, to be now referred to as Brownson (Shoemaker 1984), p 43.

It is not in virtue of the physical matter of the brain that Brown lives on in Brownson, but rather in virtue of inheriting the required psychological states, and experiencing them in the first-person mode. Shoemaker sees similarity between his 'brain-transfer' and Locke's prince and the cobbler (Shoemaker 1984), p 78.

While the physical matter of the brain is relevant as it is the vehicle of transfer, it is only a contingent fact that the relevant psychological states are instantiated in that bit of matter, rather than in any other. What is crucial is the inheritance of the right psychological states, as it is these which determine the continuation of identity.

Parfitian Teletransportation

Of Parfit's several thought experiments, the first concerns the putative teletransportation of a human being to Mars. Due to assumed advances in technology, it is now possible for human persons to become located on Mars without having to travel there. A brain-scanner is able to copy and record the complete state of a person's body and brain cells, at the same time at which the body is destroyed. This information is transmitted to Mars, arriving three minutes later. It is then used to create an exact duplicate of the person, using new matter. The person can return to earth in the same way, and in fact, 'travel' back and forth between Mars and earth any number of times. On each occasion he 'wakes up' as the self which entered the process three minutes earlier.

Parfit imagines that he travels in this way to Mars:

My replica thinks that he is me, and he seems to remember living my life up to the moment when I pressed the green button. In every other, way, both physically and psychologically, my Replica is just like me. If he returned to Earth, everyone would think that he was me (Parfit 1984), p 200.

Parfit distinguishes between numerical and qualitative identity, granting that the replica is the latter rather than the former. However, by application of his 'Widest criteria', he claims the replica is the original person:

Reconsider the start of my imagined story, where my brain and body are destroyed. The Scanner and the Replicator produce a person who has a new but exactly similar brain and body, and who is psychologically continuous with me as I was when I pressed the green button. The cause of this continuity is, though unusual, reliable. On the Physical Criterion and the Narrow Psychological Criterion, my Replica would not be me. On the two Wide Criteria, he would be me (Parfit 1984), p 209.

By allowing any form of causal continuity, rather than the 'normal' cause of memory retention, psychological continuity with the original person is maintained. Thus the person who continues life on Mars is the same person as the former person Parfit who lived on earth.

3 Thought Experiments Assessed

The above scenarios are examples of ways in which normal bodily and psychological continuity have become disconnected or disordered. In each instance, psychological continuity is favoured to retain identity over other forms of continuity. Although the scenarios themselves are highly problematic, their conclusions may seem initially convincing. It is intuitively plausible that a person's psychology relates significantly to a person's identity. On the face of it, it does seem that our relations with other persons frequently concern their mental rather than their physical characteristics. However, further consideration of these thought experiments reveals that many relevant issues are neglected, and that once they are considered, conclusions seem less secure.

The scenarios place emphasis on what observers might take to be the locus of identity, rather than on developing accounts about how transferred identity might operate in new surroundings. In addition, even if we accept that psychological continuity is a necessary part of personal identity, it is not clear that psychological continuity would be favoured over bodily identity were such a dispute to ever arise. We need to reconsider the above cases.

In the first case, just how certain is it we would take the prince to be the cobbler? If the prince was an accomplished pianist, we might ask him to prove his identity by playing us an item from his repertoire. But, with his long, slender fingers replaced with the gnarled, stubby fingers of the cobbler, how might this be accomplished? Further, would we really identify him as the prince if his voice was the gruff, heavily accented voice of the cobbler, rather than the soft, refined voice of the prince? In addition, if the prince was gay, and the cobbler heterosexual, once the prince was in the cobbler's body, would he be gay or straight? Things would be even more confusing if the cobbler was a pregnant woman. These variations suggest that personal identity may comprise more than the psychological continuity criterion entails.

In the second case, what if Brown was racially different to Robinson - say Brown was Chinese, short and small-framed, and Robinson was Jamaican, tall and muscular? In spite of the brain transfer, would the difference in physical stature and general appearance affect our judgment regarding Brownson's identity? Moreover, what would the wife or parents of the former person Brown think? Further complexity would reign if Robinson was a skilled horseman, but Brown had a severe fear of horses. Certain aspects of Brownson's identity might conflict.

In the third case, if Parfit's identity continued in a duplicate body, in what sense could his mental contents also continue? Due to the three minute time delay, there is discontinuity at the time of transfer. How do we know that some mental contents are not lost during that delay? Even if physical components are copied accurately, how could we know that mental contents are equally accurate? Further, what if Parfit's original failed to die, while contact with the duplicate was forever lost? If the duplicate appeared on Mars in an alien landscape, with unrecognisable laws of physics, amongst a group of aliens who neither recognised it, nor understood what it said, in what sense would it be the 'same' person as the original Parfit still alive on earth?

It seems that some thought experiments constructed to draw one conclusion can be modified to draw others. Examples above imply that factors other than those entailed in the psychological continuity criterion may be involved in personal identity. There are significant questions unasked, and crucial issues unresolved. To assist further exploration of these thought experiments, work is considered from Williams, Wiggins, Brennan and Elliot.

Charles, Guy Fawkes, and Robert

Bernard Williams presents a thought-experiment similar to Locke's. Instead of the cobbler inheriting the psychology of the prince, it is Charles who inherits the memories of Guy Fawkes, but the problem outlined by Williams could equally apply to Locke's example. Because the remembered events cannot be verified in the normal way, Williams suggests a way of overcoming this difficulty:

Let us imagine a person, say Charles, who awoke one day, to discover that he remembered having done certain things, which on an earlier occasion he could not remember doing, and being unable to remember doing other certain things which earlier he could remember doing. Even though these new memories appear to be first-personal, the strangeness of the situation may prompt us to seek verification of his claim that it was him who performed the deeds in question. For testimony to be valid, a witness would require to verify Charles' bodily presence at the event(s) concerned.

Failing the availability of such witness, the situation could be addressed from a different angle. Rather than firstly individuating Charles as an agent, and then appropriating a specific action to him, a particular action could first be individuated, and then uniquely appropriated to an agent, for example, 'the person who murdered the Duchess, whoever it was.' Under this approach, should Charles' actions prove to be those generally understood to have been undertaken by Guy Fawkes, we may be inclined to believe than somehow Charles has become Guy Fawkes (Williams 1973), pp 4-8.

According to Williams, even if we could overcome obvious objections to this account, such as our incredulity at the idea of re-incarnation, or the apparent difference of personal and bodily characteristics between Charles and Guy Fawkes, there is one particular reason why this account should not be accepted. What if, subsequent to the above, Charles' brother Robert also makes the same claim, that is, he also claims to have witnessed and carried out the relevant actions. On this account they might both be, not only Guy Fawkes, but also each other (Williams 1973), p 8.

Williams claims this outcome is 'absurd.' Yet, if Charles and Robert were equally good candidates for the identity in question, there would no principle to determine which of them was Guy Fawkes. Williams claims that analysis falls down due the failure to include the body:

We are trying to prise apart 'bodily' and 'mental' criteria; but we find that the normal operation of one 'mental' criterion involves the 'bodily' one (Williams 1973), p 5.

If Williams is right, there are serious problems with Locke's approach, and those inherited from it. Without specific spatio-temporal grounding, it does seem there would be nothing to stop the indefinite proliferation of a single set of psychological states. More consideration needs to be given to the relation between the body and personal identity.

Brown, Robinson, and Brown

Possible person reduplication is not only problematic for psychological states, but also for the brain and brain parts in which those states are instantiated. An early example of this problem was raised by David Wiggins. He considers the possibility of brain-identity duplication, based on Shoemaker's account of Brown and Robinson. Wiggins extends the scenario to Brown's brain being split into two equal halves prior to transplant. Each half is put into a different body, resulting in two persons with the former Brown's memories. Based on the psychological continuity criterion, they each have equal claims to now being Brown. This outcome entails the unacceptable consequences that they will initially appear to be the same person as each other, but later will appear to be two different persons:

if we say each is the same person as Brown, we shall have to say Brown 1 is the same person as Brown 2. That is an inescapable part of what was meant by saying that each was the same person as Brown. But Brown 1 will have all sorts of experiences which Brown 2 will not. They will be in different places and have separate experience from now on. And they will communicate interpersonally (Wiggins 1967), p 53.

This outcome is rife with paradox and confusion. Is Brown one person in two bodies, two different persons with two different lives, or does he cease to exist altogether? The consequences of brain-body discontinuity overturns our common-sense notions of personal identity. The problem would be amplified were additional brain dissections and transfers to be considered.

No Just Cause

Parfit's case of imaginary teletransportation is intended to justify the Widest criterion - psychological continuity with any cause. Based on this criterion, causal connections of any kind could exist between the psychological states of an individual at one time, and the identically similar states of another individual at a later time. This argument works because the relation which counts is the survival of those states, rather than the survival of their owner.

Various objections to this argument have been mounted, two of which are now considered. They both question Parfit's causal requirements, concluding that his argument does not yield the conclusions he claims. Andrew Brennan claims that the causal requirements lead to the conclusion that the survival of psychological states is as equally indeterminate as is personal identity (Brennan 1987), p 225, while Robert Elliot claims the causal requirements do not entail the survival relation, are virtually meaningless, and could therefore be dispensed with.

Parfit's claim is that the Widest criterion shows that what matters is the survival of particular mental states, regardless of either their particular physical embodiment, or of the cause of their survival. Brennan's claim is that the Widest criterion shows that no such gulf exists between survival and personal identity, each is equally indeterminate (Brennan 1987), pp 225-230. This is because the Widest criterion generates a dilemma by setting up a tension between survival and personal identity.

Two points of Parfit's theory are at issue: the need for a causal connection (whatever kind) for the survival of mental states, and the irrelevancy of whether the person who causes the states survives or not. These two requirements generate a contradiction. According to the Widest criterion, a relation exists between the survival of mental states, and the person who causes them. This criterion also states that the survival of particular individual persons is unimportant. Thus, if the survival of the particular persons who cause mental states is unimportant, then their survival is independent of those states. But, if their survival is independent of those states, they cannot be the cause of them. Alternatively, if particular persons are relevant to the cause of mental states, and their survival is unimportant, then those states are similarly unimportant.

Put another way, if having a cause matters, then so must the person who causes - that person *is* the cause. If the person had not survived, the states in question would not have been caused at all. But, if the identity of the person is merely trivial, then so also must be the causal role attributable to that person:

If causal differences of the sort mentioned are trivial, then either personal survival is independent of causal role in the way just suggested, or, if it does depend on causal role, it depends on merely trivial circumstances and is thus - given Parfit's view - no better off than personal identity (Brennan1987), p 226.

If, however, the causal role is abandoned, Brennan claims there would be an unacceptable proliferation of mental states. Without causal connections to anchor them to a legitimate origin, any selection of mental states deemed to be sufficiently like the originals could count as cases of survival. Without refinements attributable to particular individuals, there would be 'too many cases to count as cases of survival' (Brennan 1987), p 226. Brennan concludes that unless the role of persons as causes is clarified, arguments about survival are subject to 'crippling ambiguity' (Brennan 1987), p 230.

Elliot also questions Parfit's notion of cause (Elliot 1991), pp 55-75. He claims it is inadequate to include the elements significant to the survival of mental states. Yet, Parfit's causal continuity requirement (CCR) is crucial to his version of psychological continuity. It is this which permits the retention of personal continuity when psychological continuity is disconnected from other forms of continuity, such as spatio-temporal, bodily, or brain (Elliot 1991), pp 56-57. Either way Parfit's theory is threatened. Without CCR, it cannot work, but if CCR proves to be fruitless, its retention is meaningless:

My concern is rather to show that psychological continuity theories which include CCR are unstable; either CCR must be dropped or the psychological continuity approach must be abandoned (Elliot 1991), p 58.

Elliot uses two sets of thought experiments, designated Aa, Ab, Ba, and Bb to argue his case. The first in each set are cases where the presence of CCR warrants the conclusion that identity is retained. The second in each set are cases where the absence of CCR implies that identity is not retained. Elliott argues that the difference in cases is not sufficient to warrant different conclusions to each - if identity is maintained in the first examples, it should also be retained in the second.

In case Aa, a super-being creates Y following the death of X. The super-being creates Y to be psychologically similar to X because it wants X to live on in Y. Both X and Y know this, and expect it to happen. No bodily or brain continuity is involved, but there is a causal connection in virtue of the super-being's specific intentions and actions.

Based on CCR, we should accept that X lives on as Y. Case Ab is similar to Aa, in that Y appears following the death of X, complete with similar psychological states to X, and these events were expected by them both.

However, there is no causal connection between the death of X and the appearance of Y. The fact that no such connection exists does not affect the fact that the psychological states in Y happen to be similar to those of X. Elliot claims that if we accept that X survives as Y in case Aa, we should also accept it in case Ab. The reason is that, although a causal relation is present is Aa, that causal relation is not one of survival. It is no more the case that X caused the states of Y in Aa than it is in Ab. Thus if we are prepared to accept X as being Y in Aa, we have no legitimate grounds for denying it in Ab.

The second set of thought experiments is similar to Parfit's teletransportation case. In Ba, X becomes located at a distant place as Y, by having her body biochemically recorded, reconstituted as Y, and then destroyed. CCR is met due to the causal connection between the various stages of the process. In this case, we would, in accordance with CCR, accept that Y was the former X. Case Bb is similar, except that the blueprint containing X's records is lost. During the malfunction, a person Y appears. She has been constituted from a stockpile of elements, but is coincidently just like X. Further, Y believes that she is X. Things, for both X and Y seem to have occurred just as they had expected. Thus, even if Y discovers the malfunction, she still believes she is X. But, although X mirrors Y, no causal connection between them exists.

Elliot claims that in spite of this, if we accept that Y is X in Ba, we should also accept this in Bb. This is because, although CCR was present in Ba and not in Bb, the causal connection was not one in which the states of the original person actually survived. In essence, the person Y in Ba is no different to the person Y in Bb. This means that if the psychological continuity of Ba is sufficient to satisfy CCR, then so also is the psychological similarity of Bb. The conclusions drawn from both sets of thought experiments is that Parfit's version of causal continuity does not capture the elements involved in the survival of mental states, and thus has no force. Consequently, psychological continuity is rendered virtually meaningless, as there is no discernible difference between instances when it pertains and instances when it does not.

4 Conclusion of Assessment

Under scrutiny, the thought experiments of Locke, Shoemaker, and Parfit prove too sketchy and incomplete to bear the weight of the conclusions drawn from them. In Locke's example, it was claimed that in virtue of having the prince's memories, we would take the former cobbler to be the prince. Similarly, for Shoemaker, due to the brain transfer, we would take it that Brownson was the former Brown. Finally, the precisely controlled process of copying and duplication encourages us to take Parfit's Martian duplicate as being him. But, were we to take into account additional issues, such as appearance, physical characteristics, and other contingencies mentioned we might easily draw different conclusions to those suggested. Whether these would be more or less valid is unclear.

Work of other theorists also throws doubt on the outcome of the above thought experiments. Williams and Wiggins demonstrate the potential of thought experiments for person reduplication by extending cases of inherited memories, and increasing the amount of brain dissection carried out on individual brains. Further extension of both ideas could lead to uncontrolled person proliferation. Brennan and Elliot call into question the causal continuity requirements of the Widest criterion. They reveal insecure reasoning, in which the relation between causal continuity and survival is problematic. Their work demonstrates that other thought experiments are possible which yield different, or opposing conclusions, showing that reasoning based on thought experiments is sometimes questionable, and therefore should not be taken for granted.

Locating the Difficulties

Although thought experiments have long been used as a tool of philosophical inquiry, if we are to benefit from them, we need to be aware of their limitations. They are, after all, merely a tool, to be used in conjunction with other available tools. If we take them too much at face value, we are in danger of accepting the fantastical as the seriously possible. This appears to have happened in the case of those used in favour of the psychological continuity criterion. Those discussed here treat the issues involved in psychological continuity in a very simplistic way, such that a particular answer as to 'wherein lies the identity of the original person?' seems obvious and uncontroversial. We are wooed by the elements of the stories, to presume all too easily what the answer might be.

To accurately assess the value of these stories, we need to consider them in light of the actual world, rather than just the imaginary. We need to be aware of the limitations of our tools of inquiry.

Wilkes notes that thought experiments lack background conditions. These would be essential in scientific experiments, as they affect the legitimacy of results. Of course, thought experiments are not commensurate with scientific experiments, so they should not be expected to carry the same weight. But we need to be aware just how far we should take them. Wilkes points out an important difference between thought experiments in philosophy, and in the use of fantasy in literature. In the case of fantasy, an environment is supplied in which fantastical events can occur, permitting our suspension of belief. The world of Carroll's Alice is one in which it is legitimate to abrogate the laws of nature, we know her world is not intended to be one commensurate with our own:

A world in which one can walk through mirrors is, as explicitly indicated, a world of a dream; in such a world mushrooms can make one grow or shrink, a shop can turn into a boat, Queens can believe six impossible things before breakfast. For such fantasy, we have another world sketched for us, against the background of which the events are intelligible (Wilkes 1988), p 10.

However, in the thought experiments under discussion, such background conditions are not sketched. It is not clear whether we are operating under the same laws of nature, or some completely different or even contrary to our own. If the world of the foregoing scenarios was one in which persons did inherit the mental states of others, or in which it was possible to transfer brains, or that persons could be reduplicated, would there be other changes commensurate with these which would also affect our judgments about personal identity? The situations presented may not only be metaphysically dubious, but also logically inconsistent.

As Wilkes points out, a world in which gold has a different atomic number, or where water is no longer H2O, is as impossible as one in which a fish could be a whale (Wilkes 1988), p 18. In other words, if we reinterpret one concept, we need to recognise the impact this could have on the concepts to which it is related. Thus, if in our world the body and brain are significant to our mental states, we need to consider whether it is intelligible to assume we can radically change them without our mental states being affected.

These points are relevant if we are to draw strong conclusions based on theoretical brain transfers or person reduplication. It matters whether or not the difference between these imagined cases, and the cases of real life are significant differences. For example, even if we lack the skill to successfully separate brains from bodies (at present anyway), does the theoretical separation make any sense? Brains are not discrete objects, but are part of the nervous system, a complex network which exists throughout the whole body. Would the laws of nature be different in a world where transferring such a large and fragile system was possible?

Similarly, even on a materialist interpretation, does the idea of person replication make sense? Persons are not static objects, but, from the micro level of cellular activity, to the life-preserving functions of the body's major systems, are subject to constant dynamic change. At the sub-atomic level, particles cannot even be definitely located. There seems to be genuine doubt that 'copying' a person makes any sense.

These considerations raise two crucial questions. First, do the thought experiments in question address the issues involved in personal identity? Second, do they yield the conclusions claimed? In other words, are the issues raised in these scenarios the right ones, and do the conclusions claimed follow from the premises? Let us grant for the moment that the various forms of mental content inheritance, memory transplant and reduplication discussed are possible. In what way does this actually show that personal identity is a matter of psychological continuity, to the exclusion of other forms of continuity? Consider the following scenario:

The earth has been devastated by nuclear winter, resulting from a massive nuclear war. As a result, persons have a much shorter life-span, only twenty years as adults. Technology is such that old bodies can be recycled, and re-constituted into new bodies. Persons expecting to die can thus order a new body just like their own ahead of time. When death is impending, brain states can be copied, such that the person continues life much the same as before. In order to keep life as normal as possible, persons voluntarily agree to undergo no more than three such transplants in their life-time , so that death finally occurs at about age eighty years.

According to the psychological continuity criterion, because they have a continuing psychology with the right kind of cause, the persons above remain the same persons over time in virtue of the continuation of their psychological states, rather than because of the continuation of their bodies. The elements in the story are sufficiently similar to those of the earlier stories to recognise the points at issue. Let us now apply the two questions mentioned above.

First, have the issues involved in personal identity been addressed? We could answer in the affirmative, on the basis that we have referred to the continuation of persons' psychology and persons' bodies. Old bodies die, while psychology lives on in new bodies. Based on this, personal identity is preserved because psychology is preserved. But this answer does not show that the full range of issues involved have been raised.

It just shows that we have raised the issues which we are going to address, that is, that we have confronted personal identity by addressing psychological continuity and bodily continuity, because the issues involved in personal identity are psychological continuity and bodily continuity, and therefore by addressing these things, we have addressed personal identity.

In other words, for the above thought experiment to work, we must already have decided what is involved in the question of personal identity. The thought experiment itself did not reveal anything we did not already 'know.' This also applies to the previous thought experiments. In each instance we are given a set of factors pertaining to personal identity, and asked to choose amongst them, for example the prince's memories, the prince's body, the cobbler's body, Brown's memories, Robinson's body, Parfit's duplicate, and so on.

Because we are presented with a set of factors, amongst which a choice of alternatives is possible, it is assumed that the total set of factors has been presented. But it is not at all clear that this is the case. If factors other than those presented in the above thought experiments are involved in personal identity, we would not learn this from the thought experiments themselves. By confining analysis to what is involved in them, we are giving tacit approval to the range of possibilities that they provide. We are thus lured into thinking we have given personal identity a comprehensive analysis when in fact we have not.

Second, we need to consider whether the above thought experiments yield the conclusions claimed. In the previous example, we are to assume that persons remain the same persons over time in virtue of retaining appropriate psychological continuity, in spite of periodical body replacement. But, we might ask, in what way does the scenario show that personal identity is a matter of psychological continuity rather than of bodily continuity, or any other form of continuity?

How has the story proved that it is the fact of psychological continuity alone which preserves identity, rather than say, social custom, community acceptance, stability of relationships and so forth. There is no specific argument to show that it is psychological continuity over and above these other things which encapsulates personal identity.

To accept that this is the case I must already be sympathetic to that view. In other words, unless I already believe that personal identity is a matter of psychological continuity, I am not going to be convinced by a story which merely reiterates that view, without arguing for it.

To conclude from these thought experiments that persons retain their identity in virtue of psychological continuity, in spite of various bodily changes, whether due to swapping of minds and bodies, brains and bodies, or complete reduplication of bodies and so on, one must already be convinced of psychological continuity. The thought experiments themselves do not argue for this, they merely demonstrate 'cases' of it. This point could be put in the following way:

Personal identity is a matter of psychological continuity. We can see this because in cases where bodily continuity and psychological continuity become disconnected, personal identity is maintained in virtue of psychological continuity. That is, we can see that personal identity is maintained because we can see that psychological continuity is maintained. Therefore personal identity is a matter of psychological continuity.

This argument does not justify psychological continuity, as it does not actually prove its case. It demonstrates neither that psychological states operate independently of physical states, nor that personal identity does not involve issues other than mental or physical continuity. If other issues were involved, they would not be revealed in the type of arguments discussed so far.

Finally, there is a serious flaw in the general strategy of the above thought experiments. It concerns the type of analysis which is given to mental states. In the various scenarios, they are presented as insular. That is, they are considered in terms of their intrinsic content. This view, sometimes known as internalism, stipulates that what makes a mental state the particular state it is, is describable in terms of the mental state itself, as opposed to something external to the state, such as some item in the person's environment.

For example, in the case of Locke, what makes the former cobbler the prince is that he has the former prince's mind contents. It is not altogether clear just what this means. Perhaps it entails having mental pictures of the palace where the prince lived, of his courtiers, jewels and fine clothes. If the cobbler lives in a crude village setting, it is not clear how he would recognise these mental pictures.

Similar questions could be asked of any thought experiments which discuss mind contents as if they are items contained wholly inside a mind or a brain. This position is controversial and highly questionable. Unless there is some external reference point which gives mental contents meaning, it unclear how meaning can be derived from within the mind itself.

5 Conclusion

This paper has considered and evaluated the efficacy of key thought experiments in resolving the personal identity debate. Investigation of specific examples reveals several inadequacies. First, their sketchy and incomplete nature renders them inconclusive, and open to counter-examples and counter-arguments. Because background conditions are inadequately supplied, logical inconsistencies and impractical or contradictory states of affairs are not revealed.

Second, the scenarios do not offer cogent arguments in support of the psychological continuity criterion, but operate as if the case had already been proven.

Finally, the thought experiments addressed adopt an insular view of mental states, such that no explanation of the meaning of those states is furnished, other than that of internal reference to the state itself. This conception of mental states, as is argued elsewhere, is ultimately unsound, and cannot be sustained.

In conclusion, due to their limitations, the key thought experiments discussed above fail to vindicate the psychological continuity criterion, and ultimately leave the issue of personal identity unresolved.

6 Bibliography

Brennan, A. A. (1987). "Survival and Importance." Analysis 47(1): 225-230.

Elliot, R. (1991). "Personal Identity and the Causal Continuity Requirement." The Philosophical Quarterly 41(162): 55-75.

Locke, J. (1959). Of Identity and Diversity. An Essay Concerning Human Understanding (1894). New York, N Y, Dover Publications Inc. 439-470.

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

Shoemaker, S. (1984). Personal Identity: A Materialist's Account. Personal Identity. Oxford, Basil Blackwell. 67-132.

Shoemaker, S. (1984). Persons and their Pasts. Identity, Cause, and Mind. Cambridge, Cambridge University Press. 19-48.

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

Wilkes, K. (1988). Real People. Oxford, Clarendon Press.

Williams, B. (1973). Personal Identity and Individuation. Problems of the Self. Cambridge, Cambridge University Press. 1-18.

I am speaking here of normal experiences, excluding situations such as claimed astral travelling, near death experiences or similar. For a good discussion of thought experiments, see Kathleen Wilkes, Real People 1988, Clarendon Press, Oxford. Some ideas in this paper are drawn from this work. Wilkes cites two basic kinds of thoughts experiments: 1) those which are scientifically possible, such as carrying out a particular manoeuvre to prove a scientific law, as in the case of Galileo contemplating the outcome of objects of different weights falling to the ground. In this case, the procedure was not carried out, but it could have been - it was imagined instead; 2) those which amount to playing around with ideas and words in the mind, an example being imagining the grammatical merit of a phrase such as 'colourless green ideas sleep furiously' (Wilkes 1988), pp 2-3.

(Indented accounts presented in italics indicate that they are paraphrased rather than quoted. This is done either for the sake of brevity and economy, or because the item in question is available in a variety of formats.)

For this famous objection, see Bernard Williams, 'Personal Identity and Individuation' in Problems of the Self 1973, Cambridge University Press, Cambridge, pp 1-18. Williams solution was later criticised on the grounds that bodies also can be duplicated. This has of course meant that further thinking is required to overcome the problem of reduplication (fission), but Williams' work is nevertheless opened up the issue to discussion.

Another seminal article in this debate. See David Wiggins, Identity and Spatio-Temporal Continuity 1967, Basil Blackwell, Oxford, pp 50-58. Wiggins takes his discussion of this example from Shoemaker's presentation of the case in Self-Knowledge and Self Identity 1963, Cornell University Press, pp 23-24 - see Wiggins 1967, P50 & p78, but the example I draw on is Shoemaker's abbreviated version of the same scenario in: Sydney Shoemaker, 'Persons and their Pasts' in Identity, Cause and Mind 1984, Cambridge University Press, Cambridge. Shoemaker refers to Wiggins' variation on p 40 of the above work. This thought experiment is my own, although I may have read one like is somewhere, but cannot be sure.

An early version of this view was presented by Tyler Burge. Burge uses the term 'Individualistic' to refer to mental states which are described and individuated purely in terms of reference to the person who owns them, and to nothing outside that person. An 'individualistic' view treats 'a person's intentional mental phenomena ultimately and purely in terms of what happens to the person, what occurs within him, and how he responds to his physical environment, without any essential reference to the social context in which he or the interpreter of his mental phenomena are situated'

AAP Conference 2001.