

Wide Content and Psychological Continuity Views of Personal Identity

Alexander Kilcran

UCL

MPhil Stud Philosophy

I, Alexander Kilcran confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Externalism about mental content is the thesis that the content of at least some of a subject's mental states is individuated by things in the subject's environment. Psychological continuity accounts of personal identity claim that what it takes for a person to persist over time is to be share a sufficient degree of psychological similarity with themselves. I will argue that the truth of externalism about content implies that psychological continuity accounts of personal identity violate a plausible principle about identity, namely the Only x and y principle, which states that an object's identity and persistence conditions should be not be fixed by anything other than that very object. I will consider whether this result can be avoided by either denying that the Only x and y principle is true or applies in the case of personal identity or by adopting a version of psychological continuity grounded only in narrow content. I argue that taking either of these options means giving up some intuitively plausible views and that this counts against psychological continuity views of personal identity.

Contents

1	Introduction	4
2	Externalism and psychological continuity	5
2.1	Externalism	5
2.2	Psychological Continuity	9
3	Wide Content and the Only x and y Principle	14
4	Some Possible Neo-Lockean Responses	26
4.1	Narrow Psychological Continuity	38
5	Conclusion	50
6	Bibliography	51

1 Introduction

The aim of this work is to draw out some links between commonly held views in philosophy of mind and the personal identity debate. In particular the idea is that if externalism about mental content is true it has interesting implications for accounts of personal identity over time consisting in psychological continuity, often called Neo-Lockean theories as they descend principally from Locke's thesis that personal identity over time consisted in the continuity of memory. I will argue that if externalism about content is true, this implies that psychological continuity accounts of personal identity violate a plausible principle about identity, namely the Only x and y principle.

In the chapter 2 I set out the ideas I'll be discussing, first externalism about mental content and then psychological continuity views of personal identity. Chapter 3 makes the case that externalism and psychological continuity together imply that a person's identity and persistence conditions are dependent on their environment, and that this violates the Only x and y principle. In consider some possible objections to the first part of this claim, that externalism and psychological continuity imply that a persons identity and persistence conditions are dependent on their environment, but argue that they are ineffective. In chapter 4 I move on to discussing some options for resisting the conclusion of the third chapter. I briefly consider whether the Only x and y principle itself is sound, and whether there is an argument to be made for thinking that psychological continuity violating the Only x and y principle should count against the principle rather than the Neo-Lockean thesis. I then take a look at whether a version of the psychological continuity thesis grounded only in the narrow content of the persons mental states might allow psychological continuity to avoid violating the Only x and y principle. I'll argue that an entirely narrow psychological continuity faces serious difficulty in preserving the relationships we normally take to hold between the contents of a person's mental states and their behaviour.

2 Externalism and psychological continuity

2.1 Externalism

A great deal of literature in the philosophy of mind has been dedicated to debating both the truth of externalism about mental content and what implications it may have for other issues in philosophy of mind and other areas of philosophical enquiry. Roughly speaking, externalism is the thesis that the content of at least some mental states is individuated by things external to the subject in those states. It is best seen as the denial of internalism about mental content, the thesis that the content of mental states is individuated by things that are internal to the subject in those states.

The central claim of externalism is about how the content of mental states is individuated. The content of mental states is, roughly speaking, what those states represent. Owing to this, externalism is generally accepted as only making a claim about a restricted set of states, those that are intentional. Intentional states are those mental states that are “about” something, are of a representational nature or have veridicality conditions.

A mental state that does not represent the world external to the subject as being a certain way is not generally taken to be the kind of mental state that might have its content individuated by anything in the world external to the subject in that state. An example of such a non-representational mental state might be being in pain. It might be that pain does not represent the world as being a certain way, beyond representing it as “pain causing now” at least. Pain does not represent its cause, a pain would be the same mental state regardless of what causes it (McGinn 1989, 44-45). Depending on how broadly we construe the notion of representation we might reject the idea that pain or other sensations are not representational mental states. But for present purposes a complete inventory of representational states is not required. Some mental states definitely represent the world as being a certain way and it is these states the externalism is thought to make a claim about.

Non-representational states, should there be any at all, do seem to be within the scope of internalism’s defining claim though. Intuitively at least, as internalism does not claim any kind of dependence stronger than causal dependence between the world outside the subject and the contents of the subjects mental states, there is no need to restrict the scope of internalism to states that represent the world outside the subject. Non-representational states seem to be straightforwardly internalist, and straightforwardly not externalist.

Conversely, amongst the representational and intentional states that are within the scope of externalism’s defining claim there is a subset of mental states that seem to be straightforwardly externalist and straightforwardly not internalist. Intentional states that are de re - states that attribute some feature to a particular object directly, or in other words presuppose the existence of some particular object - look straightforwardly externalist. De re states are distinguished from de

dicto states - states that do not presuppose the existence of some particular object.

The distinction between de re and de dicto intentional states can also be characterised as a difference in the scope of the existential quantifier with respect to the scope of the propositional attitude verb that expresses the intentional state that a subject is in. Whether the more metaphysically robust characterisation above or the syntactic one just introduced is the correct way to characterise the de re/ de dicto distinction is controversial (see (McKay and Nelson 2014) for some discussion). I think I can safely set such concerns to one side and employ the syntactic characterisation here purely as a nice clear heuristic for identifying the kinds of mental states internalism and externalism are making interesting claims about.

If a subject is in a de re intentional state then this state is correctly captured by a sentence in which the existential quantifier has wide scope, it binds a variable that is within the scope of a propositional attitude verb, and the verb itself is also within the scope of the quantifier. For example, the sentence “There is some Guinness that Bob believes is good for you” attributes a de re state to Bob as we quantify over some Guinness about which Bob has a belief. In contrast, de dicto states are captured by sentences in which the existential quantifier has narrow scope, ranging only over a variable within the scope of the propositional attitude verb. “Bob believes that there is Guinness which is good for you” attributes a de dicto belief to Bob.

The upshot is that a subject being in a de re intentional state entails the existence of an object to which some feature is attributed, there can be no de re intentional state M^x that attributes some feature F to some object x without x existing and any mental state that does not directly attribute F to x is not identical to M^x . Therefore the content of state M^x , if the object x is external to a subject in state M^x , is clearly individuated by something external to the subject, namely x . Therefore de re intentional states are standardly accepted as being outside the scope of internalism. A subject being in a de dicto state does not entail the existence of any object and as such is not straightforwardly externalist in this way. The existential quantifier and the variable it binds both occur within the scope of “believes”, so it looks like we are not quantifying over an object about which Bob has a belief with the sentence “Bob believes there is Guinness which is good for you”. Rather, we are saying that Bob believes there is something, and that something is good for you.

In a similar vein, knowledge (should it be an intentional state at all) is clearly externalist, and not the kind of mental state that internalism is making a claim about. Being necessarily factive, the contents of knowledge states must be individuated by the things external to the subject that is in those states.

Non-representational states, de re intentional states and knowledge states then are all kinds of mental state that are fairly straightforwardly either internalist or externalist. They are therefore not the kind of states that either doctrine is thought to be making philosophically interesting claims about.

So the kinds of states internalism and externalism are thought to make interesting claims about are restricted to de dicto intentional states and perceptual states.

Among these de dicto intentional states are beliefs, desires, hopes and whatever other states consist in taking an attitude to a proposition. The content of such states is expressed by the proposition that appears in “that-clauses” of sentences that attribute those mental states to a subject. Returning to the example above the sentence, “Bob believes that Guinness is good for you” expresses the content of an ordered triple of *Guinness*, an “...is good for...” relation and *you*. The terms that appear in that-clauses of sentences that attribute intentional states occur opaquely. That is to say that the terms within the scope of a propositional attitude verb (such as “believes” or “hopes”) cannot be substituted for extensionally equivalent terms without potentially changing the truth value of the sentence in which they occur. This must be the case if such sentences are to accurately characterise the intentional states that they attribute to subjects (Burge 1979, 76-77). To illustrate the point, suppose it is true that “Bob believes that Eric Blair is not Orwell”. This sentence contains a that-clause that specifies the content - an ordered triple of *Eric Blair*, an “...is not identical to...” relation and *Orwell* - of the belief state Bob is in. Although “Eric Blair” and “Orwell” are in fact extensionally equivalent, Bob believes they are not. As such, the truth value of the sentence “Bob believes that Eric Blair is not Orwell” will change if the terms “Eric Blair” or “Orwell” were substituted for an extensionally equivalent term, such as “George Orwell”. This is because Bob does believe that George Orwell is Orwell. The terms “Eric Blair” and “Orwell” must therefore occur opaquely in order for the sentence “Bob believes that Eric Blair is not Orwell” to properly characterise the content of the intentional state that Bob is in (it is this opacity that gives rise to the de re/ de dicto distinction). That’s more or less all there is to say about the notion of mental content we are dealing with when discussing internalism and externalism. One final point to note is that content is typically taken to individuate mental states. That is, if mental state M^1 has content C^1 and mental state M^2 has content C^2 , then M^1 is not identical to M^2 . This point and much of the proceeding discussion needs to be made more informative by reflecting on what is meant by “individuation”. So individuation is the next notion employed in the crude formulations of internalism and externalism above that needs some explication.

We can start with the rough statement that if X individuates Y then X makes Y the thing it is rather than some other thing. That is to say that if X individuates Y then X fixes the identity and persistence conditions of Y . The individuation of Y by X occurs via a relation between the two. The most straightforward example of such a relation is identity; if X is identical to Y then X will fix the identity and persistence conditions of Y in virtue of its own identity and persistence conditions. Individuation can occur via other relations too, for example, sets have their identity and persistence conditions in virtue of their members. The set $\{a, b\}$ is the set it is rather than another set in virtue of having a and b as members, and if a or b fail to continue existing, so does $\{a, b\}$. But $\{a, b\}$ is not identical to its members. Rather the members of a set fix the identity and persistence

conditions of a set via the “... *is a member of...*” relation. Also *X* need not totally individuate *Y*. *X* might determine some necessary identity and persistence conditions of *Y*, but these conditions may not be sufficient to individuate *Y*. Some other thing, possibly *Y* itself, may contribute the remaining conditions for a total set of necessary and sufficient identity and persistence conditions for *Y* (McGinn 1989, 3-6).

The notion of individuation at work here is what Burge calls constitutive dependence (though I’ll stick to calling it individuation). Burge’s first point about individuation is that it does not occur via a merely causal relation. Individuation is a stronger notion than causal dependence, as we have already seen, for *X* to individuate *Y* is more than for *X* to cause *Y*. *X* must contribute to what makes *Y* the thing it is. He also notes that individuation implies nomological and metaphysical necessity, but is a stronger relation than these too. The idea here is again that to individuate *Y*, *X* must be more than metaphysically necessary for the existence of *Y*. Burge gives the example of it being necessary of a mental state that it either exists in a world where $2+2=4$ or it is made of sheep’s cheese. Clearly it is necessary that $2+2=4$ and therefore it is necessary of a mental state that it exists at a world where $2+2=4$, but this necessity does not contribute anything to an explanation of what that mental state is. It is in this sense that individuation is a stronger relation than metaphysical necessity, only a subset of all the necessities contribute to the identity and persistence conditions of a thing (Burge 2010, 64).

While discussing the metaphysics of his own version of externalism Burge gives some examples of instances of individuation that bear repeating here as they elucidate quite nicely the general notion of individuation at work in both internalism and other versions of externalism. The first example concerns reflecting on what it is to be a heart, what is it that individuates hearts. To be a heart an organ must pump blood around a circulatory system when suitably hooked up to a network of blood vessels full of blood and so on. Blood vessels, blood and the rest of the stuff that goes in to a circulatory system individuate hearts, they contribute to the identity and persistence conditions of hearts via a number of relations abbreviated to “suitably hooked up” in the previous sentence. A thing cannot be a heart unless it stands in these relations to the rest of a circulatory system. Burge’s second example is that of tectonic plates. Tectonic plates are only tectonic plates in virtue of standing in certain relations to other huge pieces of rock that slide around, bump into and pull away from each other. A large piece of planetary crust that does not stand in these relations is not a tectonic plate. Again, these relations individuate tectonic plates (Burge 2010, 66-67).

Of course, being an externalist Burge’s examples of individuation are intended to explicate how the individuation of a thing can, and often does, occur via relations that thing stands in to other distinct things. We could give examples that do not involve individuation via relations. A diamond for instance, looks like it is the thing that it is in virtue of the particular structural arrangement of the carbon atoms that compose it and nothing else (though Burge expresses doubts about whether even cases like this don’t involve some individuation via relations to other things (Burge 2010, 66).

Now to return our discussion of individuation to the rough formulations of internalism and externalism above. Internalism can be understood as claiming that the contents of intentional states are individuated only in a non-relational way, analogous to the individuation of a diamond by its chemical structure. Put another way, internalism claims that the contents of intentional states (and therefore the states themselves) are identical to whatever it is that fixes their identity and persistence conditions, which is usually taken to be some part of the subject that is in these mental states. Exactly which part of the subject depends on metaphysical commitments regarding the mind body problem. If the internalist is committed to irreducibly mental stuff this will be what individuates mental states. If they are committed to physicalism then (presumably) brains will be what individuates mental states. If their ontology of mind admits of both physical and irreducibly mental components then mental states will be individuated by both these components. Whatever the internalist takes the relevant part of the subject to be, they claim that the only relation between that part and the content of the subjects mental states via which individuation occurs is the identity relation.

Externalism on the other hand claims that at least some mental states have their persistence and identity conditions fixed by things that they are not identical to, analogously to Burges examples and to sets (McGinn 1989, 3). As we have already seen in discussing Burge's examples, this means that the relation between the contents of mental states and the things that individuate them will be various. Exactly what these relations are depends on the externalist's theory of content and, again, their commitments regarding the ontology of mental stuff. In what follows I'll be assuming that externalism about the content of at least some intentional states is true, and as such the conclusions of this essay are conditional on the truth of externalism. I won't be offering any arguments against internalism as such, though in later chapters I will employ some arguments for thinking that there is little reason to posit narrow content. This obviously weakens the overall conclusion that I wish to draw in this work and it would be better to be able to give some really good reasons for thinking that internalism was false, but I'm not entirely convinced of the case against internalism myself.

2.2 Psychological Continuity

Parfit has given a fairly influential Neo-Lockean account of personal identity in terms of psychological continuity.

Following a discussion of the inadequacy of Locke's original memory criterion, Parfit account begins with defining two general relations:

1. Psychological Connectedness; "is the holding of particular direct psychological connections."
2. Psychological Continuity: "is the holding of overlapping chains of strong connectedness."

The psychological connections here are such things as having direct memories of past events, continuing to hold a belief or desire, acting on an intention formed earlier, that kind of thing. Parfit maintains that connectedness can hold to any degree. There may be only one connection between S and S*, or there may be many. How many connections are enough for strong connectedness is impossible to say, but Parfit suggests the rule of thumb should be at least half the number per day of actual persons should be the minimum for S and S* to count as strongly connected.

Psychological Connectedness, even strong connectedness, can't be the criterion of personal identity though. Connectedness is not a transitive relation, as is well known. Psychological continuity is transitive, so it is in terms of this notion that Parfit defines personal identity:

Psychological Criterion: There is psychological continuity iff there are overlapping chains of strong connectedness. S at t1 = S* at t2 iff (i) S* is psychologically continuous with S, (ii) the psychological continuity has the appropriate cause, and (iii) There is no S' that is also psychologically continuous with S at any time (Parfit 1986, 206-207).

With respect to part (ii), which says that psychological continuity must be appropriately caused, Parfit distinguishes between three appropriate kinds of cause; narrow, wide and widest. Narrow causes are the normal everyday causes of continuity, actual experiences causing memories and the like. Wide causes are unusual but reliable causes. Parfit's Teletransporter machine would be this wide kind of cause. (Parfit 1986, 199-201). Wider causes are any kind of cause. Parfit wishes to defend the view that only the narrow and wide versions of his psychological criterion give an account of personal identity, though the widest version is "as good as personal identity" (Parfit 1986, 208-209). He needs to allow for wide causes in order to help himself to the notion of quasi-mental states.

Parfit needs the notion of quasi-mental states to avoid his psychological continuity account being circular. If the only appropriate causes of continuity were narrow, the Psychological Criterion would presuppose personal identity. This is because the only appropriate way for S*'s mental states to be caused in order for them to be continuous with S's would be for them to be S*'s mental states, as this is the only way in which persons are actually psychologically continuous. Allowing for wider causes of continuity allows for the possibility of quasi-mental states.

Parfit stipulates that a S has an accurate quasi-memory of an experience iff:

1. S* has an apparent memory of an experience
2. Someone did have that experience
3. S*'s apparent memory is causally dependent, in an appropriate way, on that experience (Parfit 1986, 220).

The qualification in 3 that S's apparent memory be caused in an appropriate way just rules out cases where S misremembers an event. Parfit's example of inappropriate causation in this case

is where S might seem to remember events immediately prior to being knocked unconscious, but has in fact reconstructed this “memory” from information they are given after the fact (Parfit 1986, 206). This qualification does not, of course, rule out wide causes. It allows that S* has an apparent memory of an experience, but that S* did not in fact have that experience. S did have that experience, but thanks to advances in neural mapping and simulation technologies it has been possible to transfer that memory to S*. This process of memory transfer is reliable and not a case of reconstructed memory ruled out by the qualification in 3. Therefore, Parfit maintains, it is causally dependent in an appropriate but wide way on the actual experience had (Parfit 1986, 221-223).

Seeing as Parfit intends psychological continuity to involve other mental states as well as memory we can define the broader concept of a quasi-mental state.

S is in a quasi-mental state iff:

1. S* is in mental state m at t2
2. Someone was in mental state m at t1
3. S* being in state m at t2 is causally dependent in an appropriate way on state m at t1

Part (iii) is a non-branching condition. It is included to secure the sufficiency of the psychological account for personal identity once the possibility of fission is invoked to lend some plausibility to the notion of quasi-mental states.

Sydney Shoemaker also gives a Neo-Lockean account of personal identity. He argues that personal identity consists in psychological continuity, which in turn consists in sharing mental states that bear appropriate causal or counterfactual dependency relations to other mental states. His account leverages a functionalist account of mind in detailing what these appropriate causal or counterfactual dependency relations are.

The short version of Shoemaker’s account of the persistence of a person is this:

S at t1 = S* at t2 iff (i) S* has a mental state, M*, at t2 that are causally dependent in the appropriate way upon a mental state, M, had by S at t1 via a series of mental states (M, M1, M2...M*) each causally dependent in the appropriate way on its predecessor; and (ii) no series (M, M1, M2...M*) contains a member mental state had by more than one S at the same time.

Presumably we could add a third condition to the right hand side of the biconditional above specifying that a sufficient number of such psychologically continuous states must be had for S to be identical with S*. Unlike Parfit, Shoemaker (at least as far as I am aware) does not specifically mention such a condition or suggest a rule of thumb for what might count as sufficient. But it is implied in remarks made in his paper “Brown-Brownson Revisited” to the effect that the playing out of the functional roles of various mental states will be a holistic affair, involving a large number

of other states. Psychological continuity will therefore be a holistic affair, to be identical S and S* must share a large number of psychologically continuous mental states (Shoemaker 2004, 582).

Condition (i) is doing most of the philosophical work in this account. It is in (i) that Shoemaker puts functionalism to work in detailing what the appropriate causal dependency that must hold between two or more mental states for those states to count as psychologically continuous.

Functionalism, briefly sketched, is the idea that the nature of a mental state type is determined by the causal relations that state type bears to its inputs (proximal stimulus and other mental state types) and outputs (behaviour and other mental state types). A mental state is a functional state, it takes a given set of inputs to a certain set of outputs. Shoemaker's own example is the belief that it is raining. The belief that it is raining just is the state type that is brought about by the proximal stimulus of rain, that disposes its subject to take their umbrella when leaving the house given that they have certain other mental states such as the desire to remain dry and so on. It is the nature of a given mental state that it be caused in certain ways and causes certain other states and behaviours (Shoemaker 1984, 92).

Shoemaker notes that the mental states we are referring to in this characterisation of functionalism are restricted to states of the same person. Functionalism implies a certain degree of synchronic unity between causally relevant states. The belief that it is raining will not cause its subject to pick up their umbrella if they do not also have at the same time a desire to remain dry. This synchronic unity suggests that in part psychological continuity involves at least the possibility of "integration" - that all a person's mental states can form a consistent set (Shoemaker 1984, 96). Also, Shoemaker argues, this synchronic unity condition implies a degree of diachronic unity too. Mental states, in conjunction with other states or stimuli, will cause other states and behaviours at a future time. Even in the simple case of retaining a certain mental state we have this diachronic unity. For example, I might believe that "the race will start at 2pm tomorrow". The next morning I will have the belief that "the race will start in six hours". Shoemaker calls this latter belief a "successor state" of the earlier belief that "the race will start at 2pm tomorrow". The content of the successor state bears a causal or counterfactual dependence to the content of the initial state - this is what Shoemaker means by saying that the states have diachronic unity. Psychological continuity is "just the playing out over time of the functional natures of mental states characteristic of persons" (Shoemaker 1984, 95).

As it's expressed in this quote psychological continuity looks like it presupposes personal identity, and is therefore no good for giving a non-circular account of it. Shoemaker argues that this apparent circularity can be excised fairly easily. His idea draws on Lewis' method for excising circularity in functionalism itself. In short, the idea is that we replace mental predicates ("the belief that. . ." and so on) with variables (bound by the existential quantifier) to get an open sentence that specifies the causal relations that hold between the mental predicates (the mental predicates

functional roles) without mentioning them specifically. The mental predicates can then be defined in terms of one another without circularity, as none of the terms being defined appear in the definiens of any of the others. To extend this method to personal identity, we can understand the notion of personal identity as the relational predicate “is copersonal with” and replace this predicate with a relational variable (“...stands in R to...” and bound by the existential quantifier). This relational variable takes as values any relation that satisfies the functional role of securing the synchronic and diachronic unity necessary for mental states to have the causal relations essential to them (the role of securing psychological continuity, in other words). Personal identity is just a special case of this relation holding. As before, the term “personal identity” does not appear in its definiens, so circularity is avoided (Shoemaker 1984, 99-100).

This might seem like an odd or superficial position - that personal identity (in terms of persistence) consists in just a special case of a general psychological continuity relation holding, the case where it holds between the same person at different times. To motivate this idea the Shoemaker appeals to the possibility of “fission”. This is the idea that a person might split into two persons. Though entirely imaginary, the most plausible case in which we might think that fission occurs are split brain cases. In such cases each of the two hemispheres of one brain are transplanted into two brainless bodies. It is supposed that each of the two new persons would have a significant degree of psychological continuity with the original brain donor. But, Shoemaker urges, it would seem that neither of the new persons is identical to the original brain donor seeing as neither is identical to the other, and it would be arbitrary to say that one of them was identical to the original but not the other (Shoemaker 1984, 84-85).

Part (ii) of the biconditional above is the “non-branching” condition again. As before its role is to secure the sufficiency for personal identity of Shoemaker’s psychological continuity relation once the possibility of “fission” cases has been invoked to lend some plausibility to the notion of personal identity being just a special case of the psychological continuity relation holding.

In the discussion that follows I will turn to Shoemaker’s account pretty much exclusively when it is necessary to examine the the Neo-Lockean view in detail. To the best of my knowledge Shoemaker provides the most richly developed and consistently defended psychological criterion of personal identity.

3 Wide Content and the Only x and y Principle¹

It is natural to think that, if the persistence of a person over time consists in having a sufficient degree of psychological continuity over that time, it is required that the person's mental states have the same content over that time. For S to continue to hold the belief that water quenches thirst S must at a later time and all the times in between believe that water quenches thirst (though it need not be an occurrent belief at all those times). It is commonly held that intentional mental states are (type) individuated by their attitude type plus their contents. If state m and state m1 have different contents then they are not the same mental state. An externalist about content holds that environmental factors, both physical and social, will contribute to the individuation of the content of at least some mental states.

Consider a Twin Earth style scenario in which Bob, an Earthling, believes at t1 that water quenches thirst. According to the externalist his belief has the content that it does, and is the very belief that it is, in virtue of facts about Bob's environment. Specifically in this case, the particular nature of the natural kind referred to by Bob with the term "water", namely H₂O, determines in part the content of Bob's belief that water is thirst quenching (Putnam 1975). One night, unbeknownst to Bob, he is transported to Twin Earth, where the referent of the term "water" is not H₂O but another substance XYZ. XYZ differs in underlying nature from H₂O but it has all the same observable properties. Twin Earth is an exact duplicate of Earth except for this difference. Bob continues to live on Twin Earth ignorant of his change in location. After many years it seems reasonable to suppose, if content externalism is true, that Bob's belief that water is thirst quenching will have different content to Bob's belief at t1 on Earth. The content of his belief will have changed, his belief about H₂O at t1 is a belief about XYZ at t2. Consequently, it looks correct to say that Bob has not continued to hold the belief that water quenches thirst he had on Earth at t1, his belief on Twin Earth is a different belief. With respect to this belief he is no longer psychologically continuous with his former self.

This story leaves open the possibility that after the move to Twin Earth Bob acquired a new belief he would express with the sentence "water quenches thirst", rather than the content of his existing belief changing. He still has his belief with H₂O content, and continues to have it. But after being transported to Twin Earth he acquires a new belief with XYZ content and also continues to have that. He therefore remains psychologically continuous with his former self as far as his belief that water quenches thirst goes.

How plausible it is that Bob retains his belief about H₂O throughout the remainder of his life on Twin Earth is difficult to assess. There is nothing, as far as I can discern, in any of the major philosophical analyses of belief that rules out Bob continuing to have his H₂O belief indefinitely.

¹I owe the general line of argument in this section to Rory Madden.

On the functionalist analysis Shoemaker endorses, and makes integral to his account of personal identity, this certainly seems to be the case. Having said that, there is nothing in the functionalist analysis that suggests Bob must retain his belief about H₂O. Once Bob is on Twin Earth this belief will no longer be fulfilling the functional role that defines it on Earth. The inputs it would normally take to the outputs that characterise it are simply not present. But this isn't firm grounds for denying that Bob continues to have the belief. There's nothing about the functionalist analysis that suggests a mental state must continue to play its characteristic functional role in order for us to continue to have it. All of us have beliefs that are not always occurrent, and we don't think of them as disappearing and being re-acquired later on should their characteristic inputs occur again. At the same time there doesn't seem to be anything obviously wrong with saying that people cease to have beliefs when those beliefs cease to play an active role in their psychology for an extended period of time. Another area of philosophical inquiry that could settle the question of whether Bob hangs on to his H₂O belief on Twin Earth are theories of mental content. Again, I'm not aware of anything in any of the major theories of content that would settle the question conclusively either way. But I won't examine any of the literature in this area in defense of this claim. I'm just going to suppose that after a lengthy period of being psychologically inert, beliefs can and do disappear.

As well as the fanciful Twin Earth scenario sketched above there seem to be other more ordinary cases where it is plausible to think that there is a change in the content of a person's mental states resulting from an environmental change. *De re* states are fairly straightforwardly externalist. A person who has a *de re* perceptual state resulting in a true belief "that the liquid in the cup is water" at t₁ plausibly has a belief with a different content when they return to look at the cup at t₂ and the water in the cup has been switched out for some other clear, colourless liquid without their knowledge. Or consider a scenario much like the one sketched by Burge to show that the contents of belief can depend on a subject's social and linguistic environment too (Burge 1979). Jane may believe at t₁ that Caesar conquered Gaul. More specifically, Jane's belief is that Gaul was a large area of mainland Europe and also what we now call the British Isles, which is false, and that Caesar conquered these territories as one. Some time later expert historians come across new evidence and revise the proper use of the term "Gaul" so that it refers not only to a large area of mainland Europe but what we now call The British Isles too. Jane is ignorant of this change. Sometime later at t₂, after this change in her linguistic communities use of the term "Gaul", Jane's belief about exactly what territories Caesar conquered when he conquered Gaul is true. The contents of Jane's belief at t₁ and t₂ are different.

In each of these cases it appears that the subject's beliefs at t₂ are not psychologically continuous with their beliefs at t₁. This is not simply owing to the fact that the contents of their mental states has changed. The break in psychological continuity must be explained by the particular nature of the change in contents, as Shoemaker's functionalist account of psychological continuity allows for, and sometimes requires, that a subject's mental states can give rise to successor

states with different content (Shoemaker 1984, 95-96). As previously discussed, the temporally indexical belief that “the race will start at 2pm tomorrow” must give rise to a series of successor states with different content for the subject of those states to remain psychologically continuous. Shoemaker’s view that psychological continuity consists in the playing out over time of the functional roles of mental states requires, for at least some states, this sort of change in contents. The question then is what makes the kind of change in content required by Shoemaker’s account compatible with psychological continuity, but the kind of externalist change in content resulting from environmental change incompatible with such psychological continuity. In other words why think that an account of psychological continuity like Shoemaker’s cannot accommodate these changes in mental content resulting from environmental change?

The answer in this case seems to be that the causal relations between successor states that preserve the continuity of the kinds of mental states that Shoemaker’s account can accommodate do not hold when a state’s content changes due to an environmental change. S’s belief that “the race will start at 2pm tomorrow” is causally sufficient for a successor state with the different content “the race will start at 2pm today” when coupled with some other mental states of S such as the belief that “today is the day of the race”. The content of S’s later belief that the race will begin at 2pm today lies at the end of a chain of states each of which is causally dependent upon the previous state in the series for its content. The content of a preceding state is not always sufficient to fix the content of their successor, often further states or inputs are required. The causal chain from an initial state to a psychologically continuous successor state is more of a causal web (Shoemaker 2004, 581-582). But the initial states content is necessarily involved in preserving the diachronic unity relation between it and its successor states (Shoemaker 1984, 95).

So we might think that when the content of a subject’s mental state changes due to a change in that subject’s environment the content of their initial state is not involved in fixing the content of a later state. Returning to the example of Bob’s belief that water quenches thirst and his later belief that twin-water quenches thirst, it looks like the content of Bob’s initial belief is not involved in fixing the content of his later belief. The content of his later twin-water belief is not causally dependent on the content of his earlier water belief, and the later belief state would have the content it does regardless of the contents of the earlier belief state. In other words, the diachronic unity relation necessary for psychological continuity does not hold between the two beliefs. Externalism about mental content entails the possibility of these breaks in psychological continuity resulting from environmental changes. And if this is true it seems there is something wrong with the psychological continuity account of personal identity. It would be absurd that a person could, in principle, be taken out of existence simply by moving them to a different environment. However, this conclusion is not warranted. With respect to Bob’s belief’s about the watery stuff in his environment at t1 and t2, it is very plausible that the content of his belief at t1 “that water quenches thirst” will cause him to believe “that XYZ quenches thirst” at t2. His having the belief about water at t1 will partly

explain why he has the belief about XYZ at t_2 , and this being so, the two beliefs would still be psychologically continuous.

Besides this, there are a number of other ways in which a Neo-Lockean might resist this claim that externalism results in breaks in the psychological continuity of a subject when their environment is drastically altered. One approach would be to appeal to their stipulation that S^* being psychologically continuous with S is a matter of S^* possessing a sufficient number of mental states that are continuous with states had by S at some earlier time. There is no reason to suppose that all of a person's mental states need be continuous with states they have at an earlier time for them to be the same person. In Bob's case, the fact that continuity fails with respect to one of his beliefs after his move to Twin Earth isn't all that important as long as he has a sufficient number of other mental states that are continuous. Also there is likely no precise number of continuous states or proportion of a total number of states that is sufficient for the persistence of a person. As already mentioned above, Parfit only suggests the metric that as long as the number of S^* 's mental states that are continuous with S 's from one day to the next is greater or equal to half the number of continuous mental states actual people have from one day to the next as a rough guide. Psychological continuity admits of some vagueness (Parfit 1986, 206). That Bob fails to be psychologically continuous with respect to a particular belief after his move to Twin Earth does not threaten his persistence, as it does not threaten his overall psychological continuity.

We might increase the number of environmental changes that occur when Bob is moved to Twin Earth. We could modify the Twin Earth scenario such that all the natural kinds on that planet have a different underlying chemical structure than their Earthly counterparts, though they have the same observable properties. We could even stipulate that all sorts of terms have slightly different extensions in the Twin-languages spoken there. If Burge is correct in regarding the scope of his argument for social externalism to be as wide as he claims (and he probably is), there are a huge range of terms that might be used differently in the linguistic communities on Twin Earth (Burge 1979, 79-82). If Bob were to be moved from Earth to a Twin Earth like that then a huge number, probably almost all, of his mental states would undergo a change in content and he would no longer be psychologically continuous with his former self.

Such a scenario is imagined by James Baillie in his (1997). He also suggests a line of thought the Neo-Lockean could employ to argue that even in moving to a Twin Earth as different as the one described above a subject could remain psychologically continuous. I say "suggests a line of thought" rather than "presents an argument" as Baillie's stated aim is to map the "logical geography" of issues that arise when a psychological criteria of personal identity meets content externalism (Baillie 1997, 323). Baillie sticks to his brief, but what he says can be expanded upon to provide an argument on the Neo-Lockean's behalf here. His suggestion is that a Neo-Lockean theory can accommodate the massive change in the contents of Bob's mental states after his move to Twin Earth provided the change in content proceeds slowly enough (Campbell (2004) also looks at how

the rate of psychological change might matter for psychological continuity views). Bob's move from his Earthly environment to the very different one on Twin Earth would not result in the contents of his mental states changing straight away. At least this is the view taken by Burge. Here he is, as quoted by Baillie; "The thoughts would not switch as one is switched from one actual situation to another twin actual situation. The thoughts would switch only if one remained long enough in the other (Twin Earth) situation to establish environmental relations necessary for new thoughts" (Burge 1988, 652). If this is indeed the case then the contents of Bob's mental states only change after he has spent long enough on Twin Earth to enter into the right environmental relations that fix those contents on the externalist's view. For a while at least, Bob will remain psychologically continuous with his former self on Earth as a large number, almost certainly a sufficient number, of his mental states will have yet to change content as he goes about his business in his new environment. This allows the Neo-Lockean to run a what is essentially the argument standardly employed against Thomas Reid's classic objection to Locke's original psychological criterion of personal identity.

Reid's objection was that Locke's memory criterion led to contradiction, as it followed by this criterion that a person could both be identical and not identical to the person that performed some past action. He argued for this position with the example of a young officer who does some valiant deed or other. The young officer remembers a flogging he received as a school boy. Many years later he is an old general who remembers his valiant deed but not the flogging he received as a school boy. By Locke's criterion the school boy who got flogged is identical to the young officer and the young officer is identical to the old general. By the transitivity of identity the general is identical to the school boy who was flogged but, seeing as the general has no memory of the receiving the flogging, by Locke's criterion he is not (Perry 1975, 114-115).

In response Neo-Lockeans standardly revise Locke's original account such that instead of the identity of S and S* consisting in S* actually having memories of S's actions (or more accurately, quasi-memories under a first-person mode of presentation) it is enough that S and S* form part of a chain of person-stages each of which has memories of the stage that preceded it (Shoemaker 1984, 81). Of course, this revision is reflected in the account's outlined above from Shoemaker and Parfit in setting out the Neo-Lockean's position. The stipulation that the contents of a person's mental states will change only slowly when they are relocated to a new environment might allow this revision to do more work for the Neo-Lockean. A subject moved to a very different environment will not have entered into the appropriate environmental relations necessary for the contents of many of their externalists states to change in contents. As such, immediately after their move to the new environment the subject will still have a very large number of the same mental states they did in their old environment, remaining psychologically continuous with their former self. As time goes by and the subject enters into more and more environmental relations in their new surroundings the degree to which they are psychologically continuous with their former self diminishes. But if this

process occurs slowly enough then we have a situation analogous to Reid's school boy becoming a young officer and later an older general. The subject just after a move to a very different new environment has a sufficient degree of psychological continuity with a subject before the move in the same way that the school boy and young officer have a sufficient degree of psychological continuity. The subject just after the move also has a sufficient degree of psychological continuity with a subject at a fairly distant future time. And though, like the school boy and the old general the subject just before the move and the subject at a fairly distant future time may not share a sufficient degree of psychological continuity to be identical they do form part of a series of person-stages that are each sufficiently continuous with the previous member of the series. And therefore, the Neo-Lockean can argue, even a move to as radically a different physical and linguistic environment as sketched above need not result in a person failing to be psychologically continuous with herself. If the contents of our mental states only change slowly, externalism does not present a problem for a personal identity consisting in psychological continuity.

In response to the Neo-Lockean case regarding this issue it's possible to question the premise that the changes in content must occur slowly. As Baillie also notes, there is no argument in Burge's (1988) for thinking that the contents of mental states must change slowly after an environmental change. It is not immediately obvious that mental content must change slowly as a matter of metaphysical necessity. But it is even less obvious that the rate of change of content being metaphysically contingent could form part of an argument against the Neo-Lockean here. The Neo-Lockean's argument above need not involve the claim that the contents of any single mental state can only change slowly over a lengthy period of time. The process by which one of Bob's mental states content changes from what it was on Earth to what it becomes on Twin Earth might be fairly rapid, even rapid enough that with respect to that particular mental state he becomes psychologically discontinuous with his former self on Earth. The Neo-Lockean only needs the claim that it will take a fair amount of time for Bob to enter into each of the relations to his physical and social environment that are necessary for fixing, and changing, the content of his mental states. Unless Bob has an unusual and extremely impoverished mental life, having only a small number of mental states about only a very small part of his environment this claim looks pretty hard to deny.

It might be argued that, though it is true a person experiencing a comprehensive environmental change will only go through a gradual change in the contents of the mental states they take with them from their old environment, it is also true that they will be acquiring a large number of entirely new mental states in their new environment. Once on Twin Earth Bob will quickly begin gathering a large number of de re beliefs, intentions, desires and perceptual states. These states clearly cannot be psychologically continuous with any states Bob had on Earth in virtue of their de re nature. But whether a person who has undergone a massive environmental change would acquire enough de re mental states quickly enough to break the psychological continuity they had with their former self doesn't look like a question that can be settled a priori in any principled manner.

What proportion of a typical person's psychology consists in de re states, how many of those states they would acquire anew and at what rate they would acquire them were they to undergo a massive environmental change as imagined above are not questions I care to guess the answers to. Having said that it seems unlikely that the number of newly acquired de re states would be enough to violate a sufficiency condition for psychological continuity like Parfit's.

Another line of thought that might be pursued to argue that a person's psychological continuity could not survive a comprehensive change in their environment would involve an appeal to holism regarding beliefs and other intentional states. Holism about belief can be characterised as the view that the content of each of a person's individual belief depends to a great extent on the content of, if not all, then a very large number of that person's other beliefs (Quine 1951). If it is true that the content of any given belief depends at least to some extent on the contents of other beliefs a person holds then the changes in the content of Bob's beliefs (and presumably other types of intentional state too) that result from changes in his environment when he moves to Twin Earth will have a more wide ranging impact on his psychological continuity than has so far been supposed. Instead of each of the environmental relations he enters into effecting the contents of a single mental state, or just a narrow range of mental states, new environmental relations will effect the contents of a large number of Bob's mental states. The slow change in contents that allows the Neo-Lockean to resist the charge that Bob's psychological continuity would be broken by changing his environment looks less plausible if each his mental states that undergo a change in content trigger a change in the content of a great many other states. Of course, pursuing this strategy against the Neo-Lockean means defending holism about beliefs and other intentional states, and holism is not an entirely uncontroversial doctrine. Mounting a defense of holism is not a project I will take on here either. But it is an option for anyone to intent on arguing that psychological continuity cannot survive massive environmental changes.

So far I have been focused on the question of whether a combination of externalism about mental content and a massive change in a person's environment entails that that person could not be psychologically continuous with their former self before the environmental change occurred. I don't think the arguments set out so far warrant a conclusion that strong. The Neo-Lockean can appeal to the kind of sufficiency condition suggested by Parfit and the plausibility of the notion that mental contents will only update slowly to make a good case that it certainly isn't necessary that a person could not be psychologically continuous with themselves after a massive environmental change. As for arguments against this defense there are a few options which I have not treated in any depth but on the face of it don't look especially promising as means of for showing that psychological continuity cannot survive a massive environmental change.

But none of the defense mounted on behalf of the Neo-Lockean view above implies that, if externalism about content is true, it is false that environmental changes make a difference to the psychological continuity of a person. It may not be true that a move to a very different environment

implies a break in a person's psychological continuity amounting to an end to their existence, and the beginning of a new person's existence in the new environment. However none of the arguments made on behalf of the Neo-Lockean to resist this apparent implication of their account can be put to use resisting the weaker claim that a person's psychological continuity depends at least partly, in virtue of content externalism, on that person's environment. Changes in a person's environment may never be enough to break their psychological continuity with themselves at various past times, for whatever reasons, but environmental changes do alter the identity and persistence conditions of that person.

This weaker claim may still pose a problem for the Neo-Lockean if it is paired with a fairly familiar principle regarding the persistence of things over time. This principle is the one expressed by David Wiggins as follows; "in retracing the past history of b in order to ascertain its identity link with a, I ought not need to concern myself with things that are other than a or other than b" (Wiggins 2001, 96). Wiggins dubs this principle the "Only a and b" rule. Harold Noonan calls it the "Only x and y" principle and states it a little more clearly; "whether a later individual x is identical with an earlier individual y can depend only on facts about x and y and the relationships between them: no facts about any other individual can be relevant to whether x is y" (Noonan 2003, 13).

The Only x and y principle is typically introduced to argue against "best candidate" theories of identity over time. These theories are motivated chiefly by Hobbes' "Ship of Theseus" example (Noonan 2003, 131). Discussing this example helps bring out what exactly is being claimed by the Only x and y principle. Suppose Theseus has a ship that he thinks is in need of extensive repair. He completely renovates the ship, replacing every part of the ship down to the last nail. Suppose also that an opportunist Athenian ship-builder gets her hands on all the old parts of Theseus' ship and assembles them into a vessel like the one Theseus now has, only differing in that each of its parts is a little shabbier than its counterpart in Theseus' renovated ship. Both Theseus and the opportunist ship builder claim that they are the sole possessor of the original Ship of Theseus. In such a case it sounds plausible to say that Theseus' renovated ship would definitely be the original Ship of Theseus if the opportunist ship builders' ship did not exist, and vice versa.

The plausibility of this idea motivates "best candidate" theories of personal identity such as Robert Nozick's closest continuer theory (Nozick 1981). Such theories can be roughly characterised as claiming that in a case such as the Ship of Theseus example, the ship that is identical to the original Ship of Theseus is the ship that is the best candidate according to some criterion of qualitative continuity. The details of this criterion can be filled in differently by different theorists. Exactly how they do this isn't important here. What is important is that this implies that whichever candidate, y, for identity with x from a field greater than one is deemed the best in light of some criteria is identical with x in virtue of facts about some other candidate, y* say. The relevant facts being whatever makes y* a worse candidate than y according to the relevant criteria, and/ or the non-existence of a better candidate. What this means is that with respect to y*, it is only in virtue

of the existence of at least one other candidate, y , that it can be true that y^* is identical to y^* rather than x . If the better candidate y , did not exist then y^* would be identical to x .

Wiggins criticises this result for violating the necessity of identity. It should be necessary that y^* is identical with x but on a best candidate theory this is not the case, y^* is only identical to x in virtue of the contingent non-existence of y (Wiggins 2001, 98-99; Noonan 2003, 134). Four-dimensionalism or a constitution view of identity (a view Shoemaker has some sympathy for) might offer a way off the hook for best candidate theories with respect to Wiggins criticism. Claiming that the relation between y^* and x is not identity but “being a temporal part of the same entity” or constitution means a best candidate theory need not deny the necessity of identity (Noonan 2003, 134-135).

But, Noonan argues, a properly formulated Only x and y principle is still violated by best candidate theories. His idea here is that even if the necessity of identity is not violated the best candidate theory still leads to absurd looking results. In short, the best candidate view implies that a set of two or more events may form part of the origin and causal history of a particular entity in one possible world, w , but not be part of the origin and causal history of that particular entity, or any other entity, in another possible world, w^* , even though that set of events and all the other events that were part of the particular entities origin in w occur in w^* . To bring out the apparent absurdity of this result a little more clearly, consider again the Ship of Theseus case. Say there is a possible world, w_1 , at which Theseus renovates his ship and an opportunist ship builder gets hold of the ship parts Theseus discards and reassembles them into a ship as described above. Also say that at w_1 the renovated ship is the best candidate for being original Ship of Theseus. So the opportunist’s reassembled ship is not identical to the Ship of Theseus (or a temporal part of the same entity or constitutive of the Ship of Theseus). It is a different ship entirely. Now say there is another possible world, w_2 , at which, for whatever reason, Theseus has no interest in renovating the original Ship of Theseus. However he does dismantle it part by part in exactly the manner he did at w_1 . And the opportunist ship builder gets hold of the parts and reassembles them into a ship in exactly the manner she did in w_1 too. On a best candidate theory the opportunist’s reassembled ship is surely the original Ship of Theseus at w_2 , as it has no rival candidate at all. It therefore follows that on a best candidate theory a set of events, the dismantling and reassembly of all the old parts of the original Ship of Theseus, belong to the origin and causal history of a particular ship at w_1 , the opportunist’s reassembled ship, which is not the original Ship of Theseus, as Theseus’ renovated ship is the best candidate at that world. At w_2 those same events are not part of the origin and causal history of the opportunist’s reassembled ship, as that ship does not exist at w_2 , there is only the Ship of Theseus, which as been dismantled and put back together again. The same events took place in the same way in both possible worlds but those events constitute the origin and belong to the causal history of a particular ship in w_1 and not in w_2 , where that particular ship failed to come into existence (Noonan 2003, 136-137).

On the face of it such a result should not be acceptable consequence of a criteria of identity.

Noonan's diagnosis of what's going wrong here is that the best candidate theory allows for mere Cambridge changes in an object to make a difference to the identity and persistence of that object. The notion of a mere Cambridge changes is best explained using examples. A typical example of a Cambridge change is a the change undergone by a person when their younger sibling grows taller than them. At a certain point in time it is true of that person that they are taller than their younger sibling. At some point they stop growing and at some later time their younger sibling grows taller than them. It is no longer true of this person that they are taller than their younger sibling and it is now true of them that they are shorter than their younger sibling. The change in the truths values of these predications indicates a change in this person. But there is a very real sense in which that person has not changed at all. In this example, that person simply stopped growing while their sibling continued to grow. The change that this person has undergone only concerns one of their relational properties. Though it would not be correct to characterise a Cambridge change as a change that only concerns something's relation properties. This is because some relational properties clearly involve changes in the object that undergoes that change. For example having a one-to-one ratio between between your fingers and toes is a relational property, but a change in the truth value of that property resulting from the loss of a hand is definitely not a mere Cambridge change in the individual who has that property.

It would be more accurate to characterise Cambridge changes as those changes in an object that do not involve that object's intrinsic properties. Exactly what that means depends on how we define intrinsic properties, but I don't think much turns on precisely which analysis of intrinsicness we opt for here. On any of the major contenders for the correct analysis of intrinsic properties the properties of objects that are intrinsic involve parts of that very object, and nothing that is not part of that object (Weatherson 2014). This very swift analysis is enough to capture the notion of a mere Cambridge change, at least in the way that Noonan is construing it. Though Noonan himself shies away from any attempt at analysis, preferring instead to note that we just seem to have a decent grip on what a Cambridge change is insofar as there is almost always agreement regarding any given case (Noonan 2003, 138). In short we can say that a real change in an object involves a change in one of that object's parts or within that object's spatial boundaries. A mere Cambridge change in an object involves a change in something that is not part of that object or is outside that object's spatial boundaries that results in a change in at least one of that object's relational properties.

So with this notion of real and mere Cambridge change in hand Noonan defines in a fairly self explanatory way a number of related notions, such as a Cambridge property and a Cambridge difference. Most importantly he defines the notion of two possible situation differing in a mere Cambridge way with respect to what occurs at a particular location. This notion is simply that two possible worlds differ in a mere Cambridge way with respect to what occurs at a certain location if the only differences between those two possible worlds are mere Cambridge changes or mere

Cambridge properties (Noonan 2003, 138-139). The point of defining this notion of Cambridge difference with respect to a particular location seems to me to be another way of expressing the view above that Cambridge change cannot involve a change in an object's intrinsic properties, as intrinsic properties are necessarily properties within a particular spatial region, the region within the spatial boundaries of a particular object.

That being said we can restate the Only x and y principle as the principle that whether a later individual x is identical to an earlier individual y can only depend on facts about the intrinsic properties of x and y and cannot depend on facts about any individual that is not part of x or part of y.

If true this version of the Only x and y principle adequately explains why it is unacceptable that the same events occurring in the same way in two different possible worlds should constitute the origin of a particular individual in one world but fail to do so in the other. Returning to the example of the Ship of Theseus, in the world where Theseus's renovated ship exists, and is therefore the Ship of Theseus in virtue of being the best candidate. The opportunist ship builder's reassembled ship is the ship it is in virtue of facts about an individual that is not a part of the reassembled ship, namely facts about Theseus' renovated ship. The difference with respect to the particular location carved out by the boundaries of the reassembled ship between the two possible worlds is a mere Cambridge difference, as the only difference at that location across the two possible worlds is a mere Cambridge change in the reassembled ship. The difference does not involve any of the reassembled ships intrinsic properties, the only difference is the relational property it has of being a worse candidate than the renovated ship for being the Ship of Theseus.

This formulation of the Only x and y principle also looks like it can explain why it is a problem for a psychological continuity account of personal identity like that advocated by Shoemaker or Parfit that mental content is individuated by things in a person's environment. On the externalist view a great many mental states have their identity and persistence conditions fixed by things that are outside the person that is the subject of those states. According to a psychological continuity account of personal identity person's have their identity and persistence conditions fixed by the mental states they are in. It follows that a psychological continuity criterion of personal identity must violate the Only x and y principle. Returning to the environment switching example above in which Bob is transported to Twin Earth all of his intrinsic properties remain the same. With respect to the spatial location within the boundaries of his skin there is no change after his move to Twin Earth. The only differences that occur with respect to Bob between his being on Earth and Twin Earth are Cambridge differences. Bob's relational properties change as he changes environments but these are mere Cambridge changes. But on a psychological criterion these Cambridge changes make a difference to the identity and persistence conditions of Bob as they result in changes in the mental states that he is in.

This diagnosis of the prima facie difficulty for psychological continuity accounts of personal iden-

tity given externalism about mental content is not the only one we might offer. Instead it might be thought that it would be more accurate to say that there is tension between the two positions is due to the fact that psychological continuity views place a great deal of emphasis on phenomenology. Externalism makes the claim that the phenomenology at least some of a subject's intentional states is not sufficient to fix the identity conditions of those states. Psychological continuity accounts of personal identity on the other hand look like they come close to saying that phenomenology is sufficient. Recalling Parfit's notion quasi-memories, it is enough for for psychological continuity that a subject has an apparent memory of a certain experience, and provided that this apparent memory is caused in the right way, it can form part of a series of mental states continuous with the actual experience. It's clear that what is left of the actual experience in the apparent memory is the phenomenology, how the experience seemed to the person having it. Sameness of how an experience seems to the person having it sounds a lot like what internalism posits as being sufficient for the identity of mental states. So perhaps the tension between externalism and psychological criteria is better explained by psychological criteria involving an implicit commitment to internalism. There is no need to invoke the Only x and y principle at all in explaining why externalism about content might pose a problem for Neo-Lockean accounts of personal identity. However, while it is true that the Neo-Lockean requires sameness of phenomenology between mental states for psychological continuity, this falls a little short of the claim that phenomenology is sufficient for the identity of those mental states, or even for the continuity relation to hold between them. The "caused in the right way condition" usually does the additional work needed for that, and moreover seems to imply that to some extent psychological continuity views require externalism, at least to an extent. For example, for a person to have knowledge of a first person nature via memory of some past event, it is necessary that that person's quasi-memories of the event are initially caused by them experiencing that very event. It is not enough that the mental state they are in now shares the phenomenology of a past experience some other person had for them to have knowledge of a first person nature (this is one reason why a move to a kind of psychological continuity grounded only in narrow content is unlikely to help the Neo-Lockean, which I'll discuss later). For this reason, it seems to me that any tension between externalism and psychological continuity accounts involves something other than a disagreement over how important phenomenology is in determining the identity/ continuity of mental states. Adding the Only x and y principle into the mix is required to make externalism about content, and the resulting environmental dependence of psychological continuity a problem for the Neo-Lockean.

4 Some Possible Neo-Lockean Responses

This criticism of psychological continuity criteria of personal identity is similar in spirit to Bernard Williams' famous duplication argument as well as Wiggins' criticism of best candidate theories in that all of them are underpinned by the Only x and y principle (Noonan 2003, 129; Williams 1956-7, 238-239). Wiggins also maintains that any criterion of personal identity that includes a non-branching condition, as both Shoemaker and Parfit's accounts do, will violate the principle in so far as admitting the possibility of fission amounts to admitting the possibility of duplication (Wiggins 2001, 208). Though widely endorsed the Only x and y principle is not entirely sacrosanct. One obvious option for a Neo-Lockean looking to resist the weak criticism above is to deny the truth of the Only x and y principle. Shoemaker does pursue this option, calling into question whether an account of a person's identity and persistence conditions really does need to avoid violating it.

Shoemaker's argument for this view takes the form of a thought experiment in which we are to imagine a species that is as psychologically sophisticated as our own. Unlike us this species reproduce by a process of fission, splitting in half, with each half then regenerating the biological matter necessary to make up the rest of its body. Such creatures were introduced by Parfit in his (1971) and as such Shoemaker dubs them Parfit People. When the Parfit People undergo their reproduction by fission the resulting persons have full psychological continuity, they share all the same quasi mental states. In virtue of their reproductive mechanism they also have partial bodily continuity. Shoemaker asks us to imagine that Parfit people occasionally contract an illness that leads to one half of their body falling off. The Parfit People who suffer this illness are able to grow a new body-half in much the same manner that they do after they undergo reproductive fission. Shoemaker claims that if we are committed to the Only x and y principle we are committed to the view that a Parfit Person who contracts this illness is not the same person after they lose half their body as the person that fell ill, despite having full psychological and partial bodily continuity with the person who fell ill. And this he thinks is a highly counterintuitive result. The reason advocates of the principle are committed to this view is that it is the only alternative to holding a non-branching account, and non-branching accounts are ruled out by the Only x and y principle (Shoemaker 2004, 590).

To make sense of Shoemaker's reasoning here it helps to go back to the "Brown-Brownson" transplant cases about which Shoemaker and Wiggins have had a long running debate (Shoemaker (2004a); (2004b); Wiggins (2004a); (2004b)). The relevant variation on the Brown-Brownson case in this instance is the one described by Wiggins in which we are to imagine Brown's brain being bisected such that its left and right hemispheres are separated, and each of these hemispheres is transplanted into a different de-brained body. Call the resulting individual "splinters" Brownson (1) and Brownson (2), each of whom are psychologically continuous with Brown. If we deny that either Brownson (1) or Brownson (2) are identical to Brown, and we are committed to the truth

of the Only x and y principle like Wiggins, then the reason we have for denying identity of Brown and Brownson (1) or (2) cannot be that there is a better candidate or that branching has occurred. The Only x and y principle rules out this option. Instead it must be claimed that the intrinsic properties of Brown and the splinters are different in some way such that neither of the splinters is identical to Brown. In a slightly different case, only one of Brown's brain hemispheres is successfully transplanted, and the other is destroyed somehow. The resulting individual, Brownson Sole, also cannot be identical to Brown, for the same reason that neither Brownson (1) or (2) were, as the Only x and y principle entails that it cannot be for lack of another candidate that Brownson Sole is identical to Brown. On the Only x and y principle it must be that there is some intrinsic difference in the between Brown and Brownson Sole that means they are not the same person. The only intrinsic difference in both of these cases appear to be that Brown, Brownson (1), Brownson (2) and Brownson Sole have different bodies.

The Parfit Person and resulting Parfit People from normal Parfit Person reproduction by fission are supposed to be analogous to Brown and Brownson (1) and (2). The Parfit Person who undergoes the body-loss illness is supposed to be analogous to the case of Brown and Brownson Sole. The ill Parfit people undergo an intrinsic change in that the matter making up their body post illness is not the same matter making up their body prior to the illness. Shoemaker maintains that an advocate of the Only x and y principle must hold that a sick Parfit Person goes out of existence when they lose half their body to this illness and a new, psychologically continuous, Parfit Person begins their existence on the same spot - just as they must in the Brownson-Brownson Sole case. They must hold this view because the only way that they could affirm the identity of the Parfit Person before they got ill with the Parfit Person after they regenerate lost body matter would be to appeal to a lack of another candidate or a lack of branching. But as already stated, these are not options for holders of the Only x and y principle. But surely, Shoemaker urges, this is highly counterintuitive. A Parfit Person undergoing a natural (for them) process like losing and regenerating body matter shouldn't mean them going out of existence. As such we should reject the Only x and y principle so that we can deny this result by including a non-branching condition among the identity and persistence conditions for Parfit People (Shoemaker 2004a, 586-590).

In response to this argument Wiggins writes; "I wrote "in the case of Brownson, the surgeon .. could have made two of him." The point ... relates to splitting. A continuity relation which can be realized, with whatever difficulty, by a case of artificial fission is no basis for a sameness relation. Brown is the same what as Brownson? The same person. Decide next what a person is. The explicate identity of persons by reference to the non-fissive 'immanent causality' proper to persons. ... Either that or think through (and then live) the consequences of the thought that Brownson is a (however specific) universal" (Wiggins 2004b, 615).

This quote does not directly address the sick Parfit Person case of course, only the Brown-Brownson case. But it does serve to illustrate what must be the relevant difference between the two

for Shoemaker. The key word in the quotation above is “artificial”. Fission of Brown into Brownson (1) and Brownson (2) is induced by some non-natural process, whereas fission occurs naturally in the Parfit People. Shoemaker’s argument was that the principle forced us to accept a counterintuitive result in the case of ill Parfit People, and we should therefore reject, or at least have reason to doubt, the results entailed by the principle in other cases, such as Brown-Brownson. Whatever intuitive force Shoemaker’s ill Parfit Person case has against the Only x and y principle it must come from the naturalness of the process, or possibility, of fission going on. This is the only relevant difference between the cases at hand. It is worth asking why the naturalness of the process of fission might be thought to make a difference to how intuitively plausible we find the entailment of the principle in the Parfit Person case. Presumably it is because naturalness just carries with it a kind of plausibility of its own, and the Only x and y principle goes against this plausibility. Perhaps this is right, but we might still wonder exactly what contribution the naturalness of the process in the Parfit People case is making relative to the surgeon’s interventions in the Brown-Brownson cases. It’s difficult to adjudicate the relative naturalness of an imaginary surgical procedure and an imaginary species undergoing an imaginary illness.

Relatedly, it is also worth enquiring why we might judge the process in Parfit People case as natural at all, regardless of how naturalness is supposed to make us find what the Only x and y principle implies about the identity of Parfit People either side of their illness counterintuitive. The thought here is that there has to be some reason we think it is natural, and therefore plausible or intuitive, that Parfit People should be able to go through the body-loss illness and remain the same person. Presumably this reason cannot be that personal identity consists in psychological continuity plus non-branching and therefore it is natural that the Parfit People remain the same person. The truth or falsehood of this claim is conditional on the truth or falsehood of the Only x and y principle, so it won’t do for Shoemaker to appeal to it in arguing that the principle should be rejected. Stated another way, the worry for Shoemaker’s criticism that the Only x and y principle leads to a counterintuitive conclusion is that unless we already buy his account of personal identity it is unclear why we should accept that the result of the principle in the Parfit People case is any more counterintuitive than in the Brown-Brownson case.

For these reasons I think we should be dubious that Shoemaker does enough to successfully rebut the Only x and y principle with the sick Parfit People case. There are other objections to the principle of course, and other defenses have been mounted against them (Brennen 1986; Garrett 1988; Kingsley 2004). This is, however, the point at which I will step out of that debate.

Rather than reject the Only x and y principle altogether a defender of the Neo-Lockean account might instead suggest that the principle is violated by externalism about mental content itself. When we are tracing the history of mental state a, enquiring whether it is identical to mental state b, don’t we have to concern ourselves with things other than a and b if externalism is true? Isn’t this exactly what we shouldn’t need to do according to the Only x and y principle, and therefore

shouldn't we reject externalism?

It is of course true that externalism claims that the identity and persistence conditions of some mental states are fixed by things other than those mental states, namely things in the physical and social environment of the subject in those states. It is precisely because this is the defining thesis of externalism and that this thesis violates the Only x and y principle that its truth would make a psychological criterion of personal identity incompatible with the principle in the first place. But this does suggest we owe the Neo-Lockean an explanation of why we can accept a thesis about the identity and persistence conditions of mental states that violates the Only x and y principle but reject a thesis about the identity of persons guilty of exactly the same offence. In other words, there must be an explanation for why an account of the identity of persons should respect the Only x and y principle. An explanation why a Neo-Lockean can't just accept that externalism about content does mean that a psychological criteria of personal identity violates the principle, but also maintain that this is fine, because all it means is that persons just aren't the sort of thing that the principle applies to. There are, the Neo-Lockean might say, a number of kinds of things for which the Only x and y principle sounds highly plausible. Rocks, cats, potatoes and other concrete objects all seem like perfectly good candidates for things that we should individuate according to their intrinsic properties and nothing else. There may also be, recalling Burge's examples for illustrating what he means by "constitutive dependence", some things that look like they must be individuated according to their relational properties. A heart is a heart in virtue of being suitably related to other parts of an animal such as its blood, blood vessels and the rest of its circulatory system. A heart is not itself a relation, but an explanation of a heart's identity and persistence conditions requires reference to relations it bears to objects that are not part of it. As briefly touched upon earlier, it is doubtful that Burge would accept the Only x and y principle at all, maintaining as he does that the intuition to individuate things according only to their intrinsic properties is "nearly always mistaken" (Burge 2010, 66). Footprints are another classic example of things that are individuated by relations they bear to other things (Stalnaker 1999, 171) Externalism claims that mental states are in the latter category of things along with hearts and footprints. For the argument that a psychological criterion of personal identity is incompatible with the Only x and y principle if externalism is true to have any bite there must be reason for thinking that persons are in the former category, with rocks and cats. Without some reason for thinking that persons are the kind of thing that must be individuated only according to their intrinsic properties, the truth of the claim that the individuation of mental states violates the Only x and y principle will be not be sufficient to show that personal identity does not consist in psychological continuity. The argument, a Neo-Lockean might say, is missing at least one premise.

At this stage it is helpful to note that the argument I have been making is similar to another argument (though it is not really advanced as such) found in Colin McGinn's book "Mental Content" (McGinn 1989). McGinn's aim with these remarks is to offer a rough diagnosis of the intuition(s)

that have lead some philosophers to find externalism about content extremely difficult to accept. The diagnosis he offers also suggests how if externalism is true then psychological criteria of personal identity are false. McGinn suggests that what some find so objectionable about externalism is that it offends an implicit assumption that minds are to be modelled on, and according to some doctrines identified with, substances. That is, externalism entails that minds do not have the two features essential to substances, which according to McGinn are;

Ontological autonomy; substances have intrinsic properties and do not depend on the existence of other things for their own existence and fix the identity and persistence conditions of the substance. Exclusive spatial boundaries; substances have boundaries determined by their intrinsic properties and other substances are excluded from the space within these boundaries.

Though somewhat rough around the edges it is clear that the notion of substance that McGinn has in mind is more or less the typical thing philosophers mean by the word "substance" (Robinson 2014). If we were committed to thinking of minds as having these features, writes McGinn, it would be very difficult indeed to see how we could accept externalism about content. Externalism is the thesis that at least some mental states depend for their identity and persistence conditions on things other than themselves, in direct contradiction of ontological autonomy and the notion of spatial boundaries. McGinn sharpens this diagnosis with the observation that no one finds it especially problematic that logical operators, universals or sense-data might be among the constituents of thoughts, it is only really concrete objects, and concrete objects in the environment at that, that some feel are not the sort of thing to be the constituents of mental states. Their intuition appears to be that mental states are located where the subjects of mental states are located, so whatever fixes the a mental state's identity and persistence conditions must be located where the subjects of mental states are too. This must be so if we conceive of minds as substance-like things with intrinsic properties that fix their identity and spatial boundaries like tables or dogs. But this is precisely what externalism denies. Further clarifying his point here, its not that McGinn is claiming that anyone literally takes minds to be a substance-like thing that is the subject of mental property attributions. Persons are the appropriate subjects of such attributions. McGinns point is if externalism about content is true then attributions of mental properties to persons are not "substantial" attributions. They do not predicate intrinsic properties of persons that underpin their ontological autonomy and spatially exclusive boundaries. Rather, they predicate relational properties that can be instantiated only in virtue of things in the person's environment.

Regardless of whether McGinn's diagnosis really does get to the core of what some find objectionable about externalism, his way of framing what externalism entails for the nature of mental states leads naturally to a conclusion very similar to the one I have been arguing for. We need only add the supposition that persons are substances, as McGinn himself does, and we reach the conclusion that minds just aren't the right kind of thing with which to identify persons (McGinn 1989, 17-26). McGinn's way of putting his point here also serves to illustrate what the relevant

distinction between the categories of things mentioned earlier, with cats and rocks in one and hearts and footprints in another, might be. The former category of things are substances, the latter are not. We might think therefore that the premise missing from my argument so far is something along the lines of “persons are substances, and substances are exactly the sort of thing for which the Only x and y principle is appropriate”. Plausibly then, the question at hand is “are persons substances?”. The Neo-Lockean could look to deny that that they are, and that the Only x and y principle is appropriate with respect to them.

Exactly what question “are persons substances?” is however is not entirely straightforward. Exactly what makes something a substance is a topic of considerable philosophical pedigree in itself. But I don’t think much turns on what the correct and complete account of substance is as far as the arguments here are concerned. I’ll just stick with the generic and intuitively appealing notion that a substance is an ontologically independent thing as this is more less as sophisticated a notion as we need, at least if the work of McGinn, Olson and Shoemaker is anything to go by (McGinn 1989, 16; Olson 2007, 5; Shoemaker 1997, 287). Furthermore, following Olson, we might distinguish between a number of questions we could be asking when we are tackling the problems of personal identity.

First there is the question of what we are. That is, what are human persons, what is their metaphysical nature. An answer to this question will tell us what our essential properties are and whether we are composed entirely of matter or entirely of some immaterial stuff or some mixture of the two and, if we are not entirely immaterial, it will tell us the extent of our spatial boundaries. It will also tell us whether or not we are substances or some other kind of thing, like states, events or processes. This question does not, in Olson’s view, amount to the question of what persons are. There may be persons that do not share the underlying metaphysical nature of human persons. Alien species or technologically advanced thinking machines might count as people too, but we needn’t assume that they are necessarily what we are (Olson 2007, 9-10).

This first question naturally suggests another, which is what is the metaphysical nature of persons. The answer to this question will be a list of necessary and sufficient conditions for being a person, conditions that might be met by a variety of otherwise quite different kinds of being. Olson calls this the personhood question. Closely related to this question is the question of what it takes for a person to persist from one time to another, or what it would take to stop a person persisting from one time to another. Together the correct answer to these questions would give us the identity and persistence conditions of persons (Olson 2007, 16-17). It is clear that the answer we give to the personhood and persistence question will put constraints on the possible answer we can give to the question of what we are (Olson 2007, 19).

The question of whether or not persons are substances, in order to be sufficiently general, looks like it should be a partial answer to the personhood question rather than a partial answer to the

question of what we are. But given that the complete and correct answer to the personhood question is a set of necessary (and sufficient) conditions for a thing to count as a person, to argue that persons are substances or not means arguing that they are necessarily substances or not. I'm not aware of any such argument, nor do I plan to attempt one. Philosophers who endorse bodily criteria of personal identity or animalism are naturally going to endorse the view that persons are substances. But they don't do so in virtue of some more general commitment to that position. Interestingly Shoemaker himself argues for a position that looks a lot like the view that persons are substances. It too falls out of his particular account of the identity and persistence conditions of persons rather than being built in from the start. Similarly, arguments for bundle theories don't take it as a premise that persons are not substances, it is the conclusion towards which they are arguing. Campbell's (2006) is an example of such a view, where persons are identified with a series of mental events rather than any substance.

Shoemaker begins his discussion of the matter by drawing a broad distinction between two schools of thought on personal identity that have their roots in the work of Locke and Hume on the one hand and Reid and Butler on the other. Locke and Hume were the originators of what Shoemaker calls the reductionist approach to personal identity, an approach that denies that persons are substances. Reid and Butler laid the foundations for anti-reductionism about persons, maintaining that persons are substances. Somewhat surprisingly he claims that a Neo-Lockean view such as his own is the most promising if we wish to be anti-reductionists about persons.

We can, Shoemaker writes, distinguish between a number of senses of the word "substance" that look like what we might have in mind when asserting or denying that persons are substances. In one sense substance means "subject of properties". In another it means "parcel of stuff". Holding subject-predicate propositions about persons to be true does not entail a commitment to holding that persons are substances in the subject of properties sense, as like Hume we might think that "x having the thought that P" only implies that a certain bundle of mental states counts a certain state among its members. Instead, a person is a substance in the subject of properties sense iff:

Subject-predicate propositions about persons like the one above are sometimes true
Subject-predicate propositions about persons like this are not analysable in way that does not quantify over the subject of the predicates in those propositions

Denying that subject-predicate propositions are not analysable in way that does not quantify over the subjects of the predicates in those propositions is plausibly what Parfit does in saying "we are not separately existing entities, we could fully describe our thoughts without claiming that they have thinkers" (Shoemaker 1997, 285). Like Hume, it appears Parfit's account of personal identity consisting in psychological continuity and connectedness is reductionist with respect to persons as subjects of properties. On the other side of the personal identity debate there are philosophers who have strong intuitions that in the sort of teletransporter story Parfit tells and Shoemaker's

own Brain State Transfer procedure where a person's psychology is transferred into another body without transferring any matter, such that there is branching of psychological continuity, it is the being that has bodily continuity that is identical to the original person. Psychological continuity is not sufficient for personal identity. We might think that these intuitions come from thinking of persons as substances, and that if psychological continuity were sufficient for personal identity persons could not be substances. The being that has the same body in the branching cases has substantial continuity with the original person.

But Shoemaker argues things are not so simple. Holding that personal identity consists in psychological continuity and also holding some variety of physicalism does not obviously commit us to denying that subject-predicate propositions about persons are not analysable in a way that does not involve quantifying over the subject of such propositions. So, with respect to the "subject of properties" sense of substance there is room in logical space for advocates of psychological criteria of personal identity to be anti-reductionists too. And, Shoemaker adds, with respect to the "parcel of matter" sense of substance presumably everyone agrees that every normal case of a person persisting through time does not involve exactly the same matter composing that person throughout. Proponents of psychological criteria are, on the face of it, no more committed to reductionism about persons in this respect than their opponents (Shoemaker 1997, 286-287).

Shoemaker distinguishes a third sense of "substance", and it is this sense that we are interested in. As we have already observed in discussing McGinn above, substances are often thought to be ontologically independent in a way that marks them out from other kinds of things. Shoemaker adds a little more detail to this claim, saying that it is not far removed from the idea of substances as the subjects of properties. This basically just means that something like condition 2 (suitably generalised to apply to more than just persons) above still holds of substances considered as being those things that are ontologically independent. This Shoemaker sums up as saying that substances are those things upon which other kinds of entities such as states are "logically parasitic" or "adjectival" and which are not themselves adjectival on anything. Additionally he claims that ontological independence also involves causal and counterfactual dependence of a substance's current properties on that substance's earlier properties. The idea here is that a substance's properties are at any given time the result of immanent causation - causation internal to that substance and not involving any interaction with anything else - by that substance's earlier properties. Of course a substance's properties can be changed by interaction with other things, but left to its own devices a substance is the way it is in virtue of the way itself was at an earlier time. Substances, Shoemaker sums up, are "relatively autonomous self-perpetuators". The "relatively" qualifier enters to accommodate the fact that organisms self-perpetuate via a great deal of interaction with their environment as well as a great deal of immanent causation (Shoemaker 1997, 287-288).

The introduction of the "relatively" qualifier is accompanied by a footnote in Shoemaker's (1997) that is of interest regarding my overall argument here. Here is a quote from that footnote;

“It should be noted, however, that the independence of persons, qua subjects of mental states, is compromised in one way in which that of other organisms is not. Assuming an ‘externalist’ view about mental content, the content of a person’s mental states is determined in part by her causal relations to things in her environment and, if Tyler Burge is right, by what linguistic practices exist in communities to which she belongs. To the extent that personal identity consists in a psychological continuity that involves the content of mental states, it requires a certain amount of constancy in the external factors that enter into the determination of such content. This is a further reason, beyond the dependence of persons, qua biological organisms, on an appropriate environment, for qualifying the term ‘autonomous self-perpetuator’ with the term ‘relatively.’” (Shoemaker 1997, 302).

It is clear that Shoemaker is aware of the sorts of considerations that I have detailed above that suggest that the idea personal identity consists in psychological continuity implies that a person’s identity and persistence conditions are at least partly conditional on their physical and social environment. Indeed remarks he makes elsewhere (see for example (Shoemaker 2004a; 2004b)) suggest that he would count himself as an externalist about content. It appears fair to say that not only is he aware that externalism suggests environmental dependence of personal identity on the psychological continuity view but that he thinks this is indeed the case. Though as we have already seen, he is not so sympathetic to the Only x and y principle, so I don’t think he would be inclined to see this as a problem for his psychological account.

To argue that persons are substances understood as ontologically independent things, which in turn is understood as being relatively autonomous self-perpetuator and non-adjectival, Shoemaker assumes (as usual) both functionalism and physicalism.

His argument for the non-adjectival status of persons starts with the familiar functionalist claim that to be in a particular mental state entails being in, or being disposed to be in, a number of other mental states contingent on sensory inputs and other mental states a subject is already in. Assuming physicalism mental states must be realized in some physical state that plays the functional/ causal role characteristic of that mental state. We can distinguish between the core and total realizations of a thing (Shoemaker 1981). The distinction is most briefly explained with an example. A particular pound (GBP) coin is realized in a particular piece of alloy shaped in a particular way. But it is not a pound coin unless it has been manufactured at the Royal Mint. A duplicate coin manufactured somewhere other than the Royal Mint would not be a pound coin, it would be counterfeit. The piece of alloy shaped just so is the pound coin’s core realization and the piece of alloy shaped just so plus its having been manufactured at the Royal Mint and other similar facts is its total realization. Just as for the property of “being a pound coin”, we can distinguish between the core and total realizations of the property of “being in mental state m”. Given functionalism, being in a mental state necessarily involves being in a number of other mental states, so if we wish to identify a token mental state with something physical we must identify it

with a token total physical realization rather than just its core realization. Due to the necessary dependencies that hold between different mental states on functionalism, the total realizations of many states must overlap each other. That is, the physical states - which, being states, are adjectival - that form the total realizations of mental states must overlap. In order for them to overlap these states must be states of not just something but the very same thing. Therefore persons must be substances in the sense of being non-adjectival (Shoemaker 1997, 293-294).

The parallel between Shoemaker's notion of substances as autonomous self-perpetuators and his account of the persistence conditions of persons almost draws itself. Psychological continuity suffices for personal identity and psychological continuity is in his view "just the playing out over time of the functional natures of mental states characteristic of persons" (Shoemaker 1984, 95). Being in certain mental states on this view implies that a person will be in certain mental states at a later time, and often this process will occur entirely autonomously (Shoemaker 1997, 295). To supplement this picture Shoemaker turns to the core/ total realization distinction again. Any number of mental states bearing the kind of causal dependencies to each other they do on the functionalist account - and therefore belonging to the same person according to Shoemaker - will necessarily have a component of their total realizations common to all of them. This part of their total realization cannot also be part of their core realization. This non-core component of their total realization will be whatever does the job of being in various mental states for that person. In human persons this common non-core component of mental states total realizations is the matter that makes up the cerebrum. But in other kinds of person, should they be possible, it will be something else. So Shoemaker is arguing that the kind of physicalist functionalist psychological continuity that he thinks suffices for personal identity necessarily implies that there is a continuous component of the realizations of a continuous series of mental states. This continuous component being common to all of the realizations of mental states a person might have looks very much like an autonomous self perpetuator, like the brain and central nervous system of human persons is (Shoemaker 1997, 295-298).

So not only does Shoemaker think that his psychological continuity account of personal identity is compatible with persons being substances, it necessarily implies that they are non-adjectival autonomous self-perpetuators, or in other words substances. However it should be said that Shoemaker recognises that his view that I have been expounding over the last few paragraphs is in tension, but apparently not strictly inconsistent, with his intuitions in brain-state transfer cases and the like. In a situation where a person's psychological state is recorded and transferred to a different body and brain (and the original body is destroyed) Shoemaker wants to say that the person does not survive the process, but finds a Parfitian stance on what matters in survival intuitively appealing too i.e. psychological continuity is what matters in survival rather than identity (Shoemaker 1997, 299). On the face of it this looks like a retreat from the claim that persons are substances, making Shoemakers overall position on the matter somewhat unclear.

Though the water is slightly muddy regarding Shoemaker's position on the question of whether persons are substances we can be certain he is making the following two claims. First, given physicalist functionalism, for there to be a psychologically continuous series of mental states it is necessary that those states are partly realized in some substance. Second, the existence of that substance is not itself sufficient for psychological continuity and therefore not itself sufficient for the existence of a person. For the existence of a certain substance to be sufficient for the existence of a certain person that substance must be in certain states. So the substance that forms the common part of the physical realizations of a person's mental states is not identical to that person. On Shoemaker's account certain substances constitute persons in virtue of being necessary for realizing mental states that are synchronically and diachronically unified as dictated by the functional profiles by which such states are individuated according to functionalism. But those substances are not identical to persons as they have different persistence conditions. That substance might cease to be in the states it must be in to realize the common part of the total realization of a number of synchronically unified mental states but the substance can continue to exist. A cerebrum, for example, might be damaged in such a way that it can no longer be the common component of the total realization of a psychologically continuous series of mental states (ending the existence of a person) but nonetheless survive the damage done to it (Shoemaker 2011).

If Shoemaker's argument above is correct it shows that even if a Neo-Lockean accepted the Only x and y principle they need not accept that it applies to persons. Plausibly the principle applies only to substances, construed as ontologically independent things. A constitution view such as Shoemaker's leaves the Neo-Lockean room to deny that persons are substances, without retreating to a bundle view. And if persons are not substances, then persons being individuated by things other than themselves, as a psychological continuity view and externalism about content implies, does not violate the Only x and y principle.

Of course, that is not to say that Shoemaker's constitution view is correct. It faces a number of well known objections. As a general metaphysical view constitutionism is widely defended but faces some difficult problems too. It entails that more than one object can occupy the same part of space and time, which some find objectionable. It also means that two or more objects can differ in their sortal properties despite sharing all their other properties and it has been argued that we need an explanation of how this can be so. We might also wonder if why seeing as, in the case of a human person, there is both a cerebrum and a person occupying the same bit of spacetime there aren't a vast multitude of other things also in that bit of spacetime (see (Wasserman 2015) for an overview of constitutionism and objections to it). Additionally the constitution view, and other psychological criteria face, the well known objection particular to the personal identity debate known as the Too-Many-Thinkers problem. Briefly stated this is the objection that wherever there is a person on the constitution view there is also a physically indistinguishable but not numerically identical human

animal. Presumably normal adult human animals are capable of thought, and what is more we have no reason not to suppose they would be in exactly the same mental states as the persons with whom they share their location. So there are two thinking beings wherever there is a person on the constitution view. This raises a couple of problems for the constitution view. For one there is an epistemological difficulty in that there appears to be no way for us to know which of the physically and psychologically indistinguishable beings we are. More problematic is the implication that if human animals are psychologically indistinguishable from persons then they are also persons. But human animals have the kind of persistence conditions typical of other organisms, psychological continuity is not necessary or sufficient for their persistence. They can persist as unthinking beings as long as a sufficient number of their biological processes are functioning adequately. Fetuses and comatose humans, for example, are not psychologically continuous with themselves at later or earlier stages of their existence but are still human animals. And two or more human animals that are (obviously) not numerically identical to each other might be psychologically continuous. I won't discuss how well Shoemaker's constitution view fares against the objections sketched above here ((Shoemaker (2008) addresses some of these worries). I mention them here just to say that Shoemaker's view (like all philosophical positions) is by no means a free lunch. And accepting that persons are not substances also runs counter to our everyday understanding of what we are. It might be telling that Shoemaker and other proponents (Campbell in his (2001) for instance) of psychological criteria of personal identity have put forward arguments like the one above. Such arguments attempt to make the view that persons are not substances seem less counterintuitive. These arguments have to employ a great deal of sophisticated metaphysical machinery to square the intuitive and attractive view that persons are substances of a certain sort with commitments to a psychological criteria of personal identity (there is good a case to be made for taking a step back from larger metaphysical issues and focusing on the merits of the different views of personal identity in their own right (Snowdon 2014, 5). Assessing the arguments for and against the view that persons are substances and whether the Only x and y principle is true of them, in light of such arguments becomes difficult, and therefore a little beyond the scope of this work. Here I just want to emphasise that the Neo-Lockean's options if they wish to deny that persons are not the sort of things that are individuated without reference to things other than themselves marks a departure from our everyday understanding of what we are. And this surely makes those options a lot less attractive. In saying this though, I must acknowledge that I'm asking a lot of psychological continuity views. It is entirely reasonable to suppose that our everyday notions of personal identity needn't be accommodated by our philosophical theories. How we think of ourselves in everyday contexts depends upon everyday concerns like tracking the moral status of other individuals and the relationships we maintain with those around us. When we sit down to do metaphysics however, there may be little reason to give the intuitions we've developed in everyday life much weight. But like I say above, I'm not going to offer any arguments regarding this here.

4.1 Narrow Psychological Continuity

Besides calling into question the truth or applicability of the Only x and y principle the Neo-Lockean could claim that there is a narrow component of mental content that is sufficient to secure psychological continuity. The idea here would be that if there is a narrow component of mental content that a series of appropriately caused mental states have over some period of time, and sharing this narrow content is sufficient for this series of states to be psychological continuous, then psychological continuity need not violate the Only x and y principal. There is no danger of externalism about content and a psychological continuity criterion of personal identity resulting in personal identity being environment dependent if sharing narrow content is sufficient for psychological continuity.

In unpacking this idea a little further I'll start with narrow content. Narrow content is mental content that is determined entirely by a subject's intrinsic properties. That is to say that for any subject of mental states S, S's mental states have narrow content iff the content of those states is fully determined by S's intrinsic properties, where intrinsic properties are understood as those properties that S has independently of her relations to other things. Any duplicate of S would be in mental states with the same narrow content as S, regardless of any differences there might be between S and her duplicates environment. The Neo-Lockean need not argue that all mental contents are narrow rather than wide, just that there is a narrow component of mental content and that this component is sufficient for psychological continuity. They can agree or disagree that at least some of the contents of our mental states are wide as they please. I won't distinguish between each of the many versions of narrow content that have been proposed in the literature because I don't think any of what I have to say will turn on which version of narrow content .

Parallels can be drawn between this strategy and the Neo-Lockean appeal to quasi-mental states in order to avoid Butler's circularity objection. Quasi-mental states are mental states with first person reference stripped away, leaving a state that represents how things seem to a subject without saying anything about which subject that is. In moving to a notion of psychological continuity grounded only in narrow content the idea would be to strip away reference to anything in the subject's environment too, leaving how things seem to the subject without saying anything about the environment the subject is in (or which subject is in that environment).

To re-iterate ground covered earlier, in order to be sufficient for psychological continuity this narrow component of content would need to, on Shoemaker's account at least, be capable of explaining the playing out over time of the functional natures of mental states characteristic of persons. The task for the Neo-Lockean then would be to defend a conception of narrow content that can sustain the causal and counterfactual dependence relations characteristic of the functional roles of mental states. Narrow content must explain why a subject being in certain mental states, such as believing it is raining now and desiring to remain dry, is caused to form the intention to pick up their umbrella before leaving the house and caused to act on that intention. And, of course, this narrow content

must be something shared by Bob at t1 on Earth and Bob at t2 on Twin Earth. Narrow content is supposed to capture how things seem to the subject in such a way that when Bob on Earth and Bob on Twin Earth are each thinking that “water quenches thirst” the narrow content of their belief is the same.

Put like this, the sort of position the Neo-Lockean might adopt might appear quite attractive and well motivated for the same reasons that have motivated philosophers to posit narrow content in general. One especially pertinent consideration, given Shoemaker’s functionalist account of psychological continuity, is the apparent difficulty that externally individuated contents have in providing causal explanations of a subject’s behaviour. If the contents of a subject’s mental states are fixed by things outside that subject it is hard to see how that content can be involved in a causal explanation the subjects behaviour, as causation must be local - no action at a distance - so everything that need be referenced in a causal explanation of an action must be local too. This suggests that sustaining the causal and counterfactual dependency relations between a person’s mental states and behaviour over time involved in psychological continuity requires a narrow component of content.

Of course there is a distinct and more general worry about how mental content, wide or narrow, could be causally efficacious, seeing as content is abstract. The thought here is put nicely by Martha Klein; “At the moment, I don’t see how one can reconcile the claim that what-is-thought is not spatio-temporal and the belief that we do what we do (partly) as the causal consequence of what we think” (Klein 1996, 160). Supposing there is a solution to the more general causal-efficacy worry and we don’t have to endorse epiphenomenalism, the problem of causal-explanatory redundancy is supposed to remain for externalism about content in a way that it does not for internalism.

The problem has been put forward in numerous ways by numerous thinkers some of whom call it the Doppelganger challenge, a name I will also use (Jackson & Pettit 1988, 389-390; Kallestrup 2012, 189). Stated in a bit more detail the Doppelganger challenge goes like this. The content of our intentional states is supposed to provide causal explanations of our behaviour. It is supposed to be because a person has the belief with the content that “it will rain soon” and an intention with the content “arrive at work without getting wet” that that person picks up their umbrella before beginning their commute. If that person’s intentional states had different content then that person would have acted differently. If they had believed “there is no way it will rain soon” they would not have reached for their umbrella. Externalism is the thesis that the contents of intentional states are individuated by things extrinsic to the subject in those states. An intentional state has the content it has in virtue of the subject’s relations to objects or properties extrinsic to the subject. The subject’s intrinsic properties could all remain the same while the content of their intentional states could be different. But causation is local. The causes of a subject’s behaviour, understood as the movements of that subject’s body, must be caused by things intrinsic or internal to that

subject's body. So if what causes a subject to behave in a certain way is having some intrinsic property (being in a certain neurological state, say) they could have been in a mental state with different wide content and behaved exactly the same way. In a different linguistic community where "rain" refers to snow as well as liquid water falling from the sky, a duplicate commuter might have the belief that "it will rain soon", and this belief will have different wide content from that of the normal English speaking commuter. Or perhaps a different commuter might be merely hallucinating the presence of an umbrella. All the commuters will both make the same movements in the same way in virtue of being in the same internal states. A causal explanation of the two commuters movements in the act of reaching for their umbrellas will only make reference to the goings on in their brains, central nervous systems, muscles and so on. It will not make reference to anything in their environments. If it is still true that the content our intentional states provide causal explanations of our behaviour then this content must be narrow content, must depend only on intrinsic features of the subject. Narrow content is sufficient for causal explanations of the commuter's behaviour understood as the movements of their bodies in some particular way which they all share.

Narrow content still does not explain why one the normal English speaking commuter reached for that umbrella and the commuter in the different linguistic community reached for that other umbrella. But it looks like all we need to add are the relevant facts about the commuter's environment. The commuters each go to pick up that umbrella because it happens to be the one that is actually there in front of them. Narrow content explains the movements they would make regardless of their environment, and a few facts about the environment explain whatever particulars need explaining beyond that. So the wide content of their mental states need not figure in a causal explanation of their behaviour. It does not add anything of explanatory value that is not already explained by their internal states plus some facts about their environment.

So perhaps narrow content and some relevant environmental facts are enough to explain the synchronic and diachronic unity relations between mental states and action that psychological continuity consists in too. Sure, the wide content of Bob's intentional states at t1 on Earth and t2 on Twin Earth might be different but all that is required for the maintenance of the causal and counterfactual dependency relations that characterise the functional natures of those states is their narrow content plus a few facts about his environment when required. This, I think, looks fairly plausible if we buy the idea that wide content is explanatorily redundant. Narrow psychological continuity could be defined something like this:

S at t1 = S* at t2 iff (i) S* has a mental state with narrow content, M*, at t2 that are causally dependent in the appropriate way upon a mental state with narrow content, M, had by S at t1 via a series of mental states with narrow content (M, M1, M2...M*) each causally dependent in the appropriate way on its predecessor; and (ii) no series (M, M1, M2...M*) contains a member mental state had by more than one S at the same time.

Narrow psychological continuity would not violate the Only x and y principle seeing as narrow content supervenes only on the intrinsic properties of the subject. If it is plausible it promises a neat way for a Neo-Lockean psychological continuity view of personal identity to side step the counter-intuitive result that a person's persistence and identity conditions are environmentally dependent. And it would not mean having to deny that ordinary content is wide or the Only x and y principle. Unfortunately I think that there are good reasons to think narrow psychological continuity is implausible. First, the motivation for narrow content is a little thin. There is of course a huge amount of literature on whether there is sufficient motivation for positing narrow content. I can't hope to address the issue properly here but I will mention a couple of arguments for the view that there is little reason to posit narrow content for the purpose of psychological explanation. Naturally these arguments also count as very general reasons for doubting the feasibility of narrow psychological continuity. Second, there is good reason to think that a version of psychological continuity grounded only in narrow content will not be able to do all the work that the Neo-Lockean needs it to do. In particular there is reason to think that wide content does play a distinctive role in psychological explanation and that this role cannot be fulfilled by narrow content. I will argue that in turn this implies that narrow psychological continuity is not really psychological continuity at all.

So starting with the case for positing narrow content in the first place, if it turned out there was no such thing as narrow content the Neo-Lockean would obviously have a very difficult problem for narrow psychological continuity on their hands. In particular a number of thinkers have put forward a case for thinking that we have little reason to posit narrow content at all on the basis of considerations regarding psychological explanations of actions. Everyday wide content does all the explanatory work we want it to do just fine. One such argument comes from Frances Egan (Egan 2009). She argues persuasively that general considerations about computational systems support the ascription of wide content in psychological explanation, and fail to warrant the positing of a narrow component of content.

We interpret a computational system as a visual system by specifying a mapping from internal states of the system to properties such as colour variance, change in depth in the scene and the like. Under such an interpretation the internal states of the system are understood as representations of colour variance. Covariance of internal states of the system with colour variance requires "fit" between the system and its environment, which for biological systems is a product of natural selection.

Consider two type-identical computational visual (V) systems i.e. they share the same computational characterisation of their internal causal operations. One system, V1, might be fit to track colour variance, F1, on Earth. The second system, V2, might track some other property, F2, in a different environment, E2. If the two systems swapped environments V1 would track F2 and V2 would track F1, seeing as they are internal duplicates. F2 might not be a useful property to track, it might be some function on the kinds of properties visual systems on Earth track, but may be

more noise than signal for the V systems. If it is the case that F2 is not a useful property to track then we might not want to call V2 a visual system on E2 at all. However it is still natural to say that V1 represents changes in colour on Earth and that V2 represents changes in F2 on E2. It is not natural to say that V1 represents F3, where F3 subsumes F1 and the useless property F2. Put another way, the computational characterisation of the internal operations of the two systems can predict how they will behave in counterfactual environments but it does not tell us what they will represent in those environments. The semantic interpretation of the system is given by the correlations between states of the system and properties of the environment it is in, how it behaves in counterfactual environments are irrelevant.

The reason it is natural to ascribe the wide content F1 to V1 and F2 to V2 is that theories of cognition seek to answer questions about how a system represents the environment it is in. Interpreting the system as representing local rather than more general environment independent properties allows for answers to these questions. The computational characterisation of the system details the basic causal operations of the device and allows for explanations of the environment independent behaviour of the system but this component of the system is entirely independent of its semantic interpretation. There is no need to posit some narrow component of content that tells us how the system will behave in other environments as we have the computational characterisation for that. Indeed, even if V1 and V2 do in some represent a property that subsumes F1 and F2 this content is still wide, though it is narrower in that it omits the environmental contributions from F1 and F2. If we were to further narrow down the content we ascribed to the V systems by omitting the environmental contribution they would track in more and more environments we would eventually end up with systems that could not properly be said to be representational at all. A semantic interpretation of the V systems that was this general would not be much good for answering the questions about how they came to represent the environments they are in (Egan 2009, 357-361).

Egan argues that there are crucial parallels between this general account of content attribution for computational systems and content attribution in psychological explanation. However, one important difference between the role of content in computational systems and psychological explanation is that in the latter content must do the job of specifying the causal role of our propositional attitudes as well as their semantics. In Egan's words, we rely on content to "get at" the functional/causal roles in virtue of which intentional states feature in explanations of behaviour. Content does this job pretty well for the most part, but this is not always the case.

As brought to our attention by Brian Loar, sometimes it is too fine grained (Loar 1988). A person on Earth may decide not to swim because they believe the water (H_2O) is too rough. Their twin on Twin Earth might decide not to swim because they believe the water (XYZ) is too rough. We might conclude from these twins performing the same action that their beliefs share some narrow content which explains this. But just like with the V systems considered above, there is little reason to think that this sort of shared content is genuinely narrow content. It will be narrower than the

ordinary wide content of the twin's beliefs in the same way that F3 was narrower than F1 and F2 in the story about the V systems. But were their content to be entirely independent of their environments there is little reason to think the twins beliefs would be genuinely representational at all. And it certainly wouldn't be much use in giving psychological explanations of the twins actions, brought about by local, environmental concerns as these explanations are.

And sometimes wide content is too coarse, and requires the use of more sophisticated content ascriptions in order for successful psychological explanation. Going with another example from Loar, we might imagine that prior to going to see her doctor Burge's arthritis patient had taken a trip to France and learned about an illness called "arthrite" there. Not realising that arthrite is French for arthritis, she forms the belief that she has arthrite in her ankles. In such a situation the patient might be thought to have two beliefs, individuated by their distinct functional/ causal roles. One belief about arthritis that causes her to take painkillers and another belief about arthrite that does not. On the basis of this we might conclude that we need narrow content to capture the way the world seems to the patient and explain why she takes painkillers for arthritis and not arthrite. But again there doesn't seem much reason to think the content we need to capture the way the patient believes the world to be is narrow (Egan 1999, 362-364).

At this point we could also look to Robert Stalnaker's treatment of the same example from Loar. Again the case Stalnaker is putting forward is that Loar's claim that wide content ascriptions often do an imperfect job of capturing a person's beliefs gives us little reason to posit narrow content as necessary for psychological explanations. Stalnaker agrees in so far as there is some difficulty in saying whether we are ascribing a belief about arthritis or a belief about arthrite when we say that "the patient believes that she has arthritis in her ankles". But we need to ask what makes it true that the way the world seems to the patient is the in fact the way the world seems to her, what made the world seem that way to her. Or in other words what makes content attributions true. Stalnaker sees the only viable account of how intentional states get the content they do as being the causal-information theoretic account he defends elsewhere (Stalnaker 1999, 210-222). Briefly stated the causal-information theoretic view is that a mental state carries information (their content) when there is a counterfactual dependency between that state and some state of the environment. If a subject S is in state M, and M is counterfactually dependent under normal conditions on the fact that P then state M carries the information (or has the content) that P. The "under normal conditions" qualifier makes room for misrepresentation. Assuming that representational states normally represent things as they in fact are we can say that M has the content that P iff under normal conditions it is counterfactually dependent on the fact that P. To be beliefs these information carrying states must also carry information that can be used (not always consciously) by the subject in those states in determining their behaviour. Counterfactual dependencies between facts in a subject's environment and the way the subject is disposed to act make the subject's behavioural dispositions representational states too (the content of other intentional states, ac-

ording to this picture, is parasitic on the contents of the beliefs with which they interact to shape behaviour) (Stalnaker 1999, 214).

So returning to Loar's case of the confused arthritis sufferer, we can say that if normal conditions obtain then the patient will be in the internal state she is in, and take the world to be the way she does if her internal state is counterfactually dependent on the world in fact being the way she takes it to be. The content (and the way the world seems to her) of her belief would be wide as it would depend on her environment being a certain way. Or in other words, the way the world seems to the patient that Loar suggests we need to posit narrow content to account for can be accounted for perfectly well by wide content. All that the case of the confused arthritis sufferer shows is that when people have beliefs that are false in peculiar ways it can be difficult to work out exactly what belief ascriptions are true of them (Stalnaker 1999, 203-250).

Egan and Stalnaker's arguments against the need to posit narrow content for the purpose of psychological explanation aren't the whole story of course. But they do represent a general hurdle for a narrow version of psychological continuity to overcome. More interesting is that even if we accept that there is a narrow component of mental content I think there is still reason to doubt that it could ground a version of psychological continuity that can do all the work the Neo-Lockean needs it to do. Returning to the Doppelgänger Challenge in particular, not everyone is convinced that there is no distinctive work for externally individuated content to do in explanations of a person's behaviour. Christopher Peacocke has put forward a case for thinking that wide content has a distinctive role to play in psychological explanation that narrow content plus a few environmental facts could not be expected to perform. I'll set out Peacocke's argument then attempt to show how it would spell trouble for narrow psychological continuity if his conclusion holds true.

Peacocke argues, psychological explanations of a person's action explain that actions relational properties. For example, the act of reaching out with one's hand towards one's umbrella may have a number of relational properties. It may have the relational properties of "being a reaching movement towards the umbrella stand in the hall", "being a reaching movement to the right of the person's body" or "being a reaching movement westwards" amongst many others. A distinctive feature of explanations of relational properties of actions such as reaching for an umbrella is that they explain the truth of various counterfactuals regarding the action undertaken. Suppose the person we are talking about here is our commuter on a rainy Monday morning. They are in a number of mental states (beliefs about the weather and the location of their umbrella plus an intention to remain dry) that explain the movement of their hand out towards their umbrella. If this is the case then the following counterfactual is true, all else being equal; if the commuters umbrella had been on the other side of the hall then she would not have reached to her right or westwards, but she would still have reached out towards her umbrella. If on the other hand the commuter had been in a number of mental states that explained her hand moving westwards a different counterfactual is true; if the commuters umbrella had been on the other side of the hall

then she would not have reached towards her umbrella but would still have reached westwards and to her right. These counterfactuals have antecedents that make reference to objects in the person's environment and their consequents involve the person's relations to various things in her environment.

Peacocke therefore argues that the worry that there is no distinctive work for wide content to do in psychological explanations arises out of a misunderstanding about what is to be explained by the content of mental states. Intrinsic states of the subject that fully individuate narrow content are sufficient to explain the movements of the subject's body. But we need wide content in order to explain relational facts about a person and their environment. These relational facts are counterfactuals like those above; if the umbrella had been on the other side of the hall then the commuter would not have reached westward and to her right. Counterfactuals like this are explained by the externally individuated content of mental states and in turn explain more than the mere bodily movement of a hand in one direction or another (Peacocke 1993, 204-207). The commuter's behaviour with respect to her umbrella is essentially relational, she moves toward it, away from it and so on. To explain this relational behaviour we need relationally individuated content. Without wide content, with only her internal states to go on, we can only explain a series of movements in one direction or another. Similar remarks are made by Timothy Williamson, including some additional examples showing how factive (very much externalist) states also have something distinctive to add to psychological explanations of behaviour (Williamson 2000, 61-62).

Jackson offers a simple response to the argument that it is only wide content that can explain the truth of counterfactuals regarding essentially relational action. The movement of the commuter's hand towards her umbrella is causally downstream of the formation of the internal state that is causally sufficient for the movement of her hand one way or another. She saw the umbrella or was made aware of its location sometime before she went to pick it up. The information she received about the umbrella's location updates the narrow content of our commuter's mental state. So narrow content alone can still explain the truth of the counterfactual; if the umbrella had been on the other side of the hall then the commuter would not have reached westward and to her right. If she had seen the umbrella on the eastern side of her hallway prior to reaching out to pick it up she would have been in a state with different narrow content, and therefore she would have reached to the east and her left and towards the umbrella (Jackson 1995, 266; 1996, 391).

I am not entirely sure that Jackson's response really answers the question at hand (it is very possible that I am misinterpreting Jackson's point here). Though it is correct that the commuter's movements are causally downstream of her seeing the umbrella in a particular location, but it's not immediately apparent to me how this establishes that wide content is not doing some work in explaining the commuter's relational behaviour. Being in a state with the wide content that her state has if the umbrella is in some particular location is also causally upstream of her relational behaviour with respect to that umbrella. The commuter could also be put in a state with the same narrow content

as a result of a hallucinogen and this would also be causally upstream of her actions concerning her umbrella. But this state looks unlikely to sustain the same counterfactuals that her normal wide perceptual state would. So there's little reason to suppose that being put in a state with certain narrow content causally upstream of some relational behaviour is enough to explain the we're interested in counterfactuals by itself. The wide content of the subject's mental state might still be playing a distinctive role in relational behaviour.

Anyway, Peacocke also offers some reasons for thinking that we'd be mistaken if we took narrow content plus some environmental information as enough to explain essentially relational behaviour. First we need to consider what additional information about the person's environment needs to go with their internal state to explain the truth the kind of counterfactuals above. Suppose we add a bunch of information about which direction the commuter would reach out in had her umbrella been in various other locations about her house. If the umbrella was at this end of the hall she would reach this direction and if the umbrella was hanging from this hook on the coat rack she would reach in that direction and so on. For the her internal state plus this information to give us the counterfactuals we are after we would need bridge principles linking each of these various possible umbrella locations with specific different internal states. And each of these specific different internal states must be described in a suitably abstract functional manner so that it is in some sense the same state for each of the umbrellas possible locations, seeing as the commuters intention to pick up her umbrella is the same in each possible situation.

Even if we add all of this, Peacocke argues, it is not enough to explain relational behaviour as well as wide content can. This is because our commuters intention to pick up her umbrella is itself explained by her being in a number of other intentional states, a belief that it is raining outside and desire to remain dry in this case. We'd also need to add a bunch more information as these states also explain their own counterfactuals regarding the commuters relation behaviour. As we add more and more information to go with our commuter's internal state the question has to be asked, whether we aren't just adding back in the wide content we were supposed to be leaving out of our explanation altogether (Peacocke 1993, 208-209).

If Peacocke is correct about the need for wide content to explain essentially relational behaviour it looks bad for a version of psychological continuity grounded only in narrow content. Psychological continuity is supposed to involve, on Shoemaker's view at least, the causal or counterfactual dependence of current mental states on earlier states, the playing out over time of the functional natures of mental states characteristic of persons. A key part of the functional role of intentional mental states is of course causing behaviour. And what is more intentional states qua contentful states are supposed to cause behaviour, what you believe about mathematics is supposed to be causally involved in the act of answering questions in a mathematics exam for example. So it would be very odd if the playing out over time of the functional natures of intentional states stopped short of playing this action guiding role. Peacocke's arguments suggest that where behaviour is

relational (and that is an awful lot, if not all behaviour), wide content is required to properly explain it. Plausibly ascribing wide content is the only way to get the resources we need to explain a range of counterfactuals resulting from the relational properties of a person's actions.

If we suppose that psychological continuity is a matter of sharing only narrow content it appears we must conclude that psychological continuity cannot explain these counterfactuals that are in turn explained by the relational properties of actions. That is to say that two persons being psychologically continuous would not guarantee that the same counterfactuals would be true as a result of the relational properties of their actions. Put in terms of the commuter example once again, a person might reach out with their hand towards their right, westwards and towards their umbrella and in a counterfactual scenario that commuter might also reach out with their hand towards their right, westwards and towards their umbrella, all else being equal. But for the first commuter the following counterfactual is true; if their umbrella had been on the other side of the hall then they would have reached out to their left, eastwards and towards their umbrella. However for the counterfactual commuter a different and incompatible counterfactual might be true, for example; if their umbrella had been on the other side of the hall then they still would have reached out their hand to their right and westwards but not towards their umbrella because they were merely gesturing westwards. A version of psychological continuity grounded only in narrow content would provide no means for distinguishing and explaining the distinction between the actual and counterfactual commuters actions in these two scenarios. The actual and counterfactual commuters narrow content would be the same and therefore incapable of explaining the counterfactuals true of them and the difference between the commuters action in the actual and counterfactual scenario. But intuitively there is a very real difference between them to be explained. Such a difference that I think its pretty odd to even call the commuter in the actual and counterfactual scenario psychologically continuous at all. At least with respect to the particular intentional states relevant to the actions the commuter undertakes in the actual and counterfactual scenario.

Putting this point slightly differently, the playing out over time of the functional natures of the mental states of the commuter in the actual and counterfactual scenario looks like it fails to play the action guiding role we'd expect intentional states (qua contentful states) to do if wide content is left out of the picture. Without involving wide content the difference between the action - understood as more than just a movement in some direction or other - the commuter undertakes in the actual and counterfactual scenario is underdetermined by the content of her belief about the location of the umbrella, desire to remain dry and resulting intention to pick the umbrella up. It could just as easily be a gesture westwards based on the narrow content of the commuters intentional states, as the narrow content fails to sustain the counterfactuals that distinguish the two acts. For this reason it looks very odd to say that the commuter is psychologically continuous across the actual and counterfactual situations. Or at least if the actual and counterfactual commuter are psychologically continuous in virtue of sharing the same narrow content it is a form of psychological continuity that

allows for divergence with respect to the counterfactuals true of a person's relational behaviour. I don't think this is an especially plausible notion of psychological continuity. Were we in a situation in which two individuals were reaching out to their right towards their umbrella as in the actual and counterfactual scenarios above, but different counterfactuals were of them with respect to that action, I think we would normally be inclined to judge that they were in different mental states. I am therefore doubtful that a notion of psychological continuity grounded only in narrow content could really do the work the Neo-Lockean needs it to do.

The point I am trying to make here can perhaps be brought more sharply into focus by using a different example. I mentioned earlier that Timothy Williamson had shown that states of knowledge play a distinctive role in psychological explanations of a person's behaviour. Williamson asks us to consider an example in which a burglar invests a great deal of time and effort searching a house. He spends so long searching despite an ever increasing risk of being caught in the act because he knows that there is a diamond in the house. The content of the burglar's knowledge state is wide, but if we are persuaded by the Doppelganger Challenge we will claim that there is a narrow component of the burglar's mental state that when supplemented with some facts about his environment explains his staying so long to search the house. Following Williamson's use of the terms for ease of explication, let's call the narrow component of content, the way things seem to the burglar, his belief and this narrow "belief" plus some environmental facts his believing truly. If we were to substitute "believed truly that there was a diamond in the house" for "knew there was a diamond in the house" we would be giving a worse explanation of the burglars behaviour. As Williamson puts it, "the explanans and explanandum become less closely connected" (Williamson 2000, 62). This is because it is consistent with the explanation involving only the burglar's narrow content and the fact there was a diamond in the house that the burglar's true belief was derived from false information. He might have been told that there was a diamond under the bed but this was false. However there was a diamond elsewhere in the house. In this situation it is reasonable to suppose that he would not continue his search for as long as he did once he had checked under the bed and not found a diamond. His true belief that there was a diamond in the house wouldn't explain his behaviour nearly as well as his knowing that there was a diamond in the house. If he knew there was a diamond in the house, if he had seen a diamond being taken into the house and knew it was not possible that it had been taken from the house since, he would be less likely to call off his search upon having checked under the bed. Williamson says that if we were to substitute "believes that there is a diamond in the house" for "knew there was a diamond in the house" the explanation of the burglar's behaviour would be even worse than if we substitute "believes truly" (Williamson 2000, 62). I'm not sure this is true, as the narrow content of the burglar's corresponding "belief" and "true belief" states would be the same for each set of false premises he might have derived his belief from, we should expect both would give an equally weakened explanation. The fact there is a diamond in the house remains inaccessible to the

burglar in both cases and so his prolonged risky searching looks equally poorly explained whether or not his belief is true or not. But that is somewhat besides the point.

Returning to the point, Williamson's example of the explanatory loss involved in giving psychological explanations only involving narrow content presents much the same problem for narrow psychological continuity as before. The burglar's behaviour is much less tightly connected to the content of his earlier mental states if wide content is left out of the picture. It is a crucial feature of psychological continuity that the content of a person's current mental states and the actions they explain are causally and counterfactually dependent on the content of their earlier states. Or at least this is the case on Shoemaker's functionalist version of psychological continuity. In his own words, "it [is] central to the functional nature of mental states that they, in conjunction with other mental states of the same person, 'rationalize', i.e., make it rational for the person to do (in the case of behaviour) or have (in the case of mental states)" (Shoemaker 1984, 94). It is pretty clear that in Williamson's example the burglar's behaviour is better rationalised by his state of knowing than it is by a narrow component of that state, whether or not that narrow component is supplemented with some relevant environmental facts, all else being equal. As such a version of psychological continuity grounded only in the narrow content of a person's mental states would often fail to rationalise behaviour as well as a version grounded in ordinary wide content. The explanatory power lost by omitting wide content from psychological continuity would result in a person's actions being underdetermined by the content of their earlier intentional states. This could easily amount to the diachronic unity relation not holding between their earlier states and actions at all. At the least we can say that a narrow version of psychological continuity would yield a much more tenuous, fragile link between a person's intentional states and their actions than we normally think there is.

5 Conclusion

To sum up, on the externalist view a great many mental states have their identity and persistence conditions fixed by things that are outside the person that is the subject of those states. According to a psychological continuity account of personal identity person's have their identity and persistence conditions fixed by the mental states they are in. It follows from the truth of externalism about content that a psychological continuity criterion of personal identity must violate the Only x and y principle. There is room for denying the Only x and y principle but being forced to give it up surely counts against the Neo-Lockean view. I think the prospects for a version of psychological continuity grounded only in narrow content aren't great. Narrow psychological continuity is not likely to be able to sustain the tight relationships that must hold between the contents of a person's mental states and their behaviour that even Shoemaker's own account of psychological continuity demands. Of course there is a lot more that could be said on the issue. Applying Peacocke and Williamson's arguments in the way I have above doesn't in itself show that narrow psychological continuity is logically impossible. At most it shows that a person at t1 narrowly psychological continuous with themselves at t2 would have their actions at t2 less well explained by the content of their mental states at t1 than if psychological continuity also included wide content. As I have said above I think this makes narrow psychological continuity a much less attractive option for the Neo-Lockean. But it is still an option. Faced with defending an account of psychological continuity involving only narrow content or denying the Only x and y principle though, I'd suggest that a psychological continuity account of personal identity would be better served by doing the latter.

Overall the argument I've put forward here definitely does not do enough to refute psychological continuity views of personal identity. At most it suggests that, conditional on the truth of externalism about content, adopting the Neo-Lockean approach means sacrificing one of two intuitively attractive positions. Either the Only x and y principle must be denied or narrow psychological continuity must be defended.

6 Bibliography

- Baillie, James. (1997) "Personal identity and mental content" *Philosophical Psychology*, 10:3, 323-333
- Brennan, Andrew (1986) "Best candidates and theories of identity", *Inquiry: An Interdisciplinary Journal of Philosophy*, 29:1-4, 423-438, DOI: 10.1080/00201748608602096
- Burge, Tyler. (1978), 'Individualism and the Mental', *Midwest Studies in Philosophy*, 4: 73–121.
- Burge, Tyler. (1988). "Individualism and self-knowledge". *Journal of Philosophy*, 85, 649-683
- Burge, Tyler .(2010-03-04). *Origins of Objectivity*. : Oxford University Press. Retrieved 27 Aug. 2015, from <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199581405.001.0001/acprof-9780199581405>
- Campbell, Scott (2001) "Persons and Substances" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* , Vol. 104, No. 3, pp. 253-267
- Campbell, Scott (2004) "Rapid Psychological Change" *Analysis*, Vol. 64, No. 3 pp. 256-264
- Campbell, Scott (2006) "The Conception of a Person as a Series of Mental Events" *Philosophy and Phenomenological Research*, Vol. 73, No. 2 pp. 339-358
- Egan, F. (2009-09-02). "Wide Content". *Oxford Handbooks Online*. Retrieved 28 Aug. 2015, from <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199262618.001.0001/oxfordhb-9780199262618-e-21>.
- Garrett, B.J. (1988) "Best candidate theories and identity: Reply to Brennan" *Inquiry: An Interdisciplinary Journal of Philosophy*, 31:1, 79-85, DOI: 10.1080/00201748808602139
- Jackson, Frank and Pettit, Philip (1988): "Functionalism and Broad Content". *Mind*, 97, pp. 381-400.
- Jackson, Frank (1995): "Mental Properties, Essentialism and Causation". *Proceedings of the Aristotelian Society*, 95, pp. 253-68.
- Jackson, Frank. (1996) "Mental Causation" *Mind, New Series*, Vol. 105, No. 419, pp. 377-413
- Kallestrup, Jesper (2012), *Semantic Externalism*, Oxon: Routledge
- Kingsley, Alan C. (2004) "The only X and Y principle" *Inquiry: An Interdisciplinary Journal of Philosophy*, 47:4, 338-359, DOI: 10.1080/00201740410004179
- Klein, Martha (1996) "Externalism, Content and Causation", *Proceedings of the Aristotelian Society, New Series*, Vol. 96, pp. 159-176
- Loar, Brian. (1988), 'Social Content and Psychological Content', in R. Grimm and D. Merrill (eds.), *Contents of Thought* (Tucson, Ariz.: University of Arizona Press), 99–110.
- McGinn, Colin (1989), *Mental Content*, Oxford: Basil Blackwell

McKay, Thomas and Nelson, Michael, "Propositional Attitude Reports", The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2014/entries/prop-attitude-reports/>>.

Noonan, Howard (2003); *Personal Identity*, London: Routledge

Nozick, R. 1981: *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Olson, E.(2007-11-01). *What Are We?: A Study in Personal Ontology*. : Oxford University Press. Retrieved 27 Aug. 2015, from <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195176421.001.0001/acprof-9780195176421>.

Parfit, Derek. 1971: "Personal Identity." *The Philosophical Review*, 80,1, 3-27.

Parfit, D.(1986-01-23). *Reasons and Persons*. : Oxford University Press. Retrieved 27 Aug. 2015, from <http://www.oxfordscholarship.com/view/10.1093/019824908X.001.0001/acprof-9780198249085>.

Peacocke, Christopher 1993: "Externalism and Explanation". *Proceedings of the Aristotelian Society*, 92, pp. 203-30.

Perry, John (1975) "Personal Identity" London: University of California Press

Putnam, Hilary (1973), "Meaning and Reference", *The Journal of Philosophy*, Vol. 70, No. 19, Seventieth Annual Meeting of the American Philosophical Association Eastern Division , pp. 699-711

Quine W. V. (1951), "Main Trends in Recent Philosophy: Two Dogmas of Empiricism", *The Philosophical Review*, Vol. 60, No. 1, pp. 20-43

Robinson, Howard, "Substance", The Stanford Encyclopedia of Philosophy (Spring 2014 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2014/entries/substance/>>.

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

Shoemaker, S. (1981) "Some Varieties of Functionalism" *Philosophical Topics*, Vol. 12, No. 1, Functionalism and the Philosophy of Mind, pp. 93-119

Shoemaker, S. (1997), "Self and Substance", *Noûs*, 31: 283–304. doi: 10.1111/0029-4624.31.s11.13

Shoemaker, S. (2004a) "Brown-Brownson Revisited", *The Monist*, Vol. 87, No. 4, *Personal Identity* , pp. 573-593

Shoemaker, S. (2004b) "Reply to Wiggins", *The Monist*, Vol. 87, No. 4, *Personal Identity*, pp. 610-613

Shoemaker, S. (2008) "Persons, Animals, and Identity" *Synthese*, Vol. 162, No. 3, pp. 313-324

Shoemaker, S. (2011-05-02). *On What we are*. Oxford Handbooks Online. Retrieved 28 Aug. 2015, from <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199548019.001.0001/oxfordhb-9780199548019-e-16>.

Snowdon, P.(2014-10-09). *Persons, Animals, Ourselves.* : Oxford University Press. Retrieved 28 Aug. 2015, from <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198719618.001.0001/acprof-9780198719618>.

Stalnaker, R.(1999-04-08). *Context and Content: Essays on Intentionality in Speech and Thought.* : Oxford University Press. Retrieved 28 Aug. 2015, from <http://www.oxfordscholarship.com/view/10.1093/019823709780198237075>.

Wasserman, Ryan, "Material Constitution", *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/spr2015/entries/material-constitution/>>.

Weatherson, Brian and Marshall, Dan, "Intrinsic vs. Extrinsic Properties", *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2014/entries/extrinsic/>>.

Wiggins, David. (2001), *Sameness and Substance Renewed*, Cambridge: Cambridge University Press

Wiggins, David. (2004a) "Reply to Shoemaker" *The Monist*, Vol. 87, No. 4, *Personal Identity*, pp. 594-609

Wiggins, David. (2004b) "Reply to Shoemaker's Reply" *The Monist*, Vol. 87, No. 4, *Personal Identity* pp. 614-615

Williams, B.A.O. "Personal Identity and Individuation" *Proceedings of the Aristotelian Society, New Series*, Vol. 57 (1956 - 1957), pp. 229-252

Williamson, Timothy. *Knowledge and its Limits.* Oxford: Oxford University Press, 2002. Oxford Scholarship Online, 2003. doi: 10.1093/019925656X.001.0001.