Why Parfit's Psychological Criterion Does Not Work as a Personal Identity Theory (And How it *Could*)

Chris Lay

University of Texas at El Paso

Abstract

On Parfit's Psychological Criterion of personal identity, I persist as some future subject if we can trace a chain of overlapping mental state connections from me to that future subject. When two subjects are connected in this way, we can say that they are psychologically continuous. Parfit offers up three different versions of the Psychological Criterion in *Reasons and Persons*, and what he calls the Narrow, Wide, and Widest views are distinguished from one another by what is acceptable as the cause of continuity on each. However, there appear to be problems with all three versions. On the one hand, the Wide and Widest views are so generous with how 'cause' is defined that they no longer deal with the persistence sense of personal identity at all. On the other hand, the Narrow View has a more appropriate definition of 'cause', but treats all types of mental state as equal contributors to continuity. This also becomes a problem when we consider cases where overall continuity obtains but certain moral features have been altered—as in some instances of traumatic brain injury where the original subject intuitively may not seem to persist through the injury. To this end, in this paper I examine Parfit's Psychological Criterion and argue that none of its three versions succeed as persistence accounts. Nonetheless, I do think that the Narrow View in particular can be appropriately modified to harmonize with these moral problem cases by the addition of a new relation that is necessary but not sufficient for persistence: moral continuity.

Keywords: Personal identity, Persistence, Philosophy of mind, Metaphysics.

1. Introduction

Although Derek Parfit first tested the waters with the paper "Personal Identity" (1971), it was of course the carefully curated refinement of those views in Part Three of *Reasons and Persons* (1984) that proved to be a watershed moment for the philosophy of personal identity. His talk of Teletransportation to Mars, brain transplants, and divided consciousness was not only enthralling, but shaped the nature of the personal identity debate for the next several decades. While Parfit

Argumenta 5, 1 (2019): 125-138
ISSN 2465-2334

DOI 10.14275/2465-2334/20199.lay

lays out two identity criteria, the Physical Criterion and the Psychological Criterion, it is certainly the latter with which he is most associated today (even if Parfit denies that identity is actually important, in the end).[1] The Psychological Criterion is about what is often called the *persistence* sense of personal identity—that is, the conditions under which a subject persists over time as one and the same thing.[2] Persistence thus concerns *numerical identity*, not just *exact similarity* in qualities.

Despite the fact that the Psychological Criterion is well-known as a kind of paradigm model for psychological approaches to persistence, I argue in this paper that none of the three versions of the Psychological Criterion that Parfit advances actually work as persistence theories. Two of these versions do not involve the persistence relation at all, and a third flounders when set against certain problem cases that suggest that continuity among particular mental states—moral features, in this case—may be a necessary condition for persistence. To make this case, I first provide a short analysis of the Psychological Criterion. Then, I produce arguments for why all three versions of the view fail. Lastly, I close with a short suggestion of how to modify one version of the Psychological Criterion, the Narrow View, to meet the demands of the problem cases from the third section.

2. The Psychological Criterion and What It Claims About Persistence

Parfit's Psychological *Criterion*—a word he takes to mean "what [persistence] necessarily involves or consists in" (Parfit 1984: 206)—gets its weight from two relations among mental states. The first is *psychological connectedness*: the holding of 'direct' connections between mental states. As standard examples of direct connections, Parfit offers the connection between an experience had and later remembered, the connection between an intention and the later action through which the intention is satisfied, and the connections between things like beliefs or desires that are retained over time. Problematically, direct connections of this sort are typically short-term. That is, there are few direct connections between, say, a ninety-year-old man and the little boy he was at age six. Yet, the six-year-old and the ninety-year-old *could* still be indirectly connected through sets of overlapping direct connections. If we follow the overlapping successions of years across the man's life, much like links in a chain, we find that every intermediate link is probably directly psychologically connected with *at least* the previous and subsequent links. So, despite the fact that the ninety-year-old and the six-year-old have few to no direct connections, we can trace a chain of successive direct connections between them. To Parfit, this means that there is, therefore, a kind of *psychological continuity* between the two subjects.

[1] A quick note: although I am critical in this paper of all three versions of Parfit's Psychological Criterion, I do not herein attack his view that identity is not what matters. Rather, because Parfit's insistence on the unimportance of identity is the more heterodox position, and because giving my reasons for thinking Parfit is wrong on this count would comprise a whole other paper, I take it for granted that personal identity *does* matter.

[2] To keep things simple and clear, I will mostly sideline the broader reaching phrase 'personal identity' in favor of 'persistence' throughout.

However, not just any amount of direct psychological connections will suffice for continuity. After all, we would not want to grant that a *single* direct connection between intermediate links constitutes continuity. Unlike identity—which one either has or does not—connectedness admits of degree; a greater number of connections results in a stronger 'connectedness' relation, while fewer connections makes connectedness weaker. Parfit thinks that it is implausible to clearly define a kind of 'optimal' minimal connectedness for continuity to obtain, under pain of arbitrariness. All the same, he *does* give us a paradigmatic case of what counts as 'enough' connections for continuity: when "the number of connections over any day is *at least half* the number of direct connections that hold over every day, in the lives of nearly every actual person" (Parfit 1984: 206). Parfit calls this sense of 'enough' connectedness *strong connectedness*. Full-fledged psychological continuity, then, consists in overlapping chains of *strong connectedness* between states. This is what, on the Psychological Criterion, will allow a subject to persist despite often substantial psychological variation. Hence, continuity does not make overly demanding claims that subjects *never* change psychologically—just that they do so gradually, so that the chain of direct connections is not severed at any point along the way.

With both relations—connectedness and continuity—defined, we can state Parfit's proper formulation of the Psychological Criterion.

> *The Psychological Criterion*: (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) there does not exist a different person who is also psychologically continuous with Y. (5) [Persistence] just consists in the holding of facts like (2) to (4) (Parfit 1984: 207).

Clauses (3)-(5) give us a lot to unpack, so I will break down each clause individually, beginning with (4). Clause (4) is a rider meant to protect the Psychological Criterion against odd cases where psychological continuity could conceivably 'branch off', resulting in two or more later subjects that are simultaneously psychologically continuous with the same earlier subject. Of course, *one* original subject cannot be identical with *two* later subjects, as identity is a one-one relation, so branching would clearly create problems for any persistence theory. Clause (5) tells us that the Psychological Criterion is a *reductionist* view of persistence. What this means is that a complete description of how we persist can be given with only a description of the physical and psychological facts, like which physical brain states or psychological states persist. Extrapolating from here, *psychological reductionism* then rejects the notion that persistence involves some 'further fact' or separately existing entity beyond just facts about brains and mental states.

I have left clause (3) for last because of its importance to my arguments going forward. Clause (3) stipulates that continuity must have 'the right kind of cause'. Parfit cashes out 'right kind of cause' by providing three different versions of the Psychological Criterion: the Narrow View, the Wide View, and the Widest View (Parfit 1984: 207). These versions are distinguished entirely by what they treat as the 'right kind of cause'. On the *Narrow View*, the only permissible cause of continuity is the 'normal' cause. By this, Parfit seems to mean the regular functioning of a normal brain. Something like the continuity of the

Teletransporter case he suggests—whereby someone's body and brain are utterly destroyed and a qualitatively identical Replica is subsequently created on Mars—would obviously not constitute the 'normal cause' of the Narrow View, since "the continued existence of a person's brain is at least part of the normal cause" (Parfit 1984: 208). Curiously, both the Wide and Widest views *do* allow someone to persist through Teletransportation, despite the fact that every part of the Original has been vaporized. This is because the *Wide View* permits any reliable cause of psychological continuity to be the 'right' cause, whereby we can understand 'reliable' to be something like 'regularly reproducible'. The *Widest View* is even less picky: on this view, any cause whatsoever counts. So, because the Teletransporter is *ex hypothesi* reliable, and because the Martian Replica actually is psychologically continuous with her Earthbound Original, the Original is numerically identical with the Replica on the Wide View. And, since a reliable cause is of course also just a cause *simpliciter*, the Widest View must grant that the Original is numerically identical with the Replica, as well.

So, we now have a picture of what the Psychological Criterion of persistence means. It is a theory that reduces persistence to certain causal connections between mental and physical states, so long as continuity does not branch to multiple subjects. Still, owing to the scope of what Parfit allows the 'right' cause to be, a great number of cases that fit the Psychological Criterion will not be bona fide cases of persistence. Though he admits that, strictly speaking, the Original subject is not numerically identical with the Replica on the Wide and Widest, he calls going through Teletransportation 'about as good' as ordinary persistence. Indeed, Parfit (1984: 208) emphatically states that there is no good reason to favor the Narrow View over the other two. If it sounds like Parfit is playing fast and loose with the notion of persistence, that is probably because he goes on to deny that persistence is actually what is really important to us. All told, then, the Psychological Criterion ends up being not so much a persistence theory as a deflationary theory about the meaningfulness of persistence.

3. Why No Version of Parfit's Psychological Criterion Succeeds

More or less by Parfit's own admission, two of the three versions of the Psychological Criterion of persistence do not even engage with persistence but instead with a weaker relation that is supposed to do similar metaphysical work. This might be perfectly fine for Parfit and perfectly consistent with his goals, too. However, for anyone looking for a genuine account of persistence, the Psychological Criterion does not cut it.[3] Obviously, if the Wide and Widest views only deal with a relation that is 'about as good as' persistence, then those views do not exactly qualify as persistence theories. But I think it is worse than this. I do not think that the relation involved in the Wide and Widest views is anything like persistence but is instead closer to the relation between multiple instantiations of a pattern. Yet, though the Narrow View fares better, I still do not think it can meet the intuitive demands of some special—but not entirely uncommon—cases.

[3] I take this opportunity to remind the reader that my purpose in this paper is not to argue against Parfit's claim that identity does not matter, so I go on assuming that persistence *is* an important relation and worth discussing.

To this end, I argue in this section that no version of the Psychological Criterion is successful as a theory of persistence. First, I critique the Wide and Widest views by showing that, on those views, there is not actually anything that persists. Then, I critique the Narrow View by demonstrating that the fact that the Narrow View treats all mental states as equally constitutive of persistence means that it cannot accommodate cases where certain types of mental state—in particular, moral features—seem to play a necessary role in determining persistence. The first problem case presented is admittedly fantastical, so I close the section with a more grounded case involving something that actually happens with some regularity: traumatic brain injury.

3.1 The Wide/Widest Views as Pattern Instantiation, Not Persistence

Starting with the Wide/Widest views, a slightly modified version of the Teletransporter case will reveal why neither view concerns persistence identity. This modified example is also Parfit's (1984: 199-200). Rather than being immediately destroyed when the Martian Replica is created, suppose that the Original survives the Teletransportation process. There is surely no question whether the Original persists—or if she is numerically identical with the Replica—in this case. She plainly survives the procedure and even gets the opportunity to talk to her space-bound copy over an inter-planet intercom. In virtue of this, she is quite clearly *not* one and the same entity as her Replica, as one thing cannot be in two different places at once! We can call this case Teletransporter 2, or T2, for short (the plain old Teletransporter case will be T1). In T2, the Replica's status is straightforwardly a *clone*. That the Replica is just a clone is less obvious in T1 since the Original perishes before the Replica comes into existence. After all, the case is framed as if Teletransportation moves one and the same subject from Earth to Mars; this is just what teleporters do (or are supposed to do, I guess). Things are different in T2, however. Here, the Original stays right where she is, and another entity is created whole cloth at the 'destination' location on Mars.

Parfit (1984: 201) calls T2 the "Branch-Line Case", which seemingly violates clause (4) of the Psychological Criterion and thus does not count as persistence anyway. But I would hesitate to call T2 a case of branching. The creation of a clone is pretty transparently different from standard branching cases. In standard branching—whether it is a splitting amoeba-man (Chisholm 1969), separated hemispheres from the same brain placed in different bodies (*à la* Wiggins 1967), or whatever else—the question is 'who, if either of these branching subjects, is the Original?' This is not at all the case in T2. Here, there is no issue telling the Original apart from the Replica. Nothing even happens to the Original in T2; a new entity is created that is *exactly similar* to the Original, but there is no change that the original subject undergoes. If persistence is a question of what sort of change a subject can endure, then T2 does not engage with this question at all. This is because T2 is *instead* a case of simple cloning, with a clear distinction between the Original and the Replica.

Now, what applies to T2 can also be applied to T1, since the only difference between the two is that the Original survives Teletransportation in T2 but not T1. This means that T1, like T2, does not involve the persistence of the Original—the Original is merely cloned. In the most direct sense, she clearly *does not* persist. Per Parfit's stipulation, she is destroyed as part of the Teletransportation process and does not just spring back to life as her Replica. No one

thing can have two beginnings of existence. According to reductionism, we can describe a subject just by giving the physical and psychological facts. There is no separately existing entity or 'further fact' in which the subject consists. If there are no separately existing mental entities, like souls or pure Cartesian Egos, a subject's mental states cannot exist absent a body (or at least something physical in which these states are instantiated). So, when the Original is completely physically obliterated in T1, her mental states go with her. This means that the mental states reproduced in the Replica are numerically distinct but qualitatively exactly similar states. We might say that the reproduced states just follow the same abstract pattern as the states in the Original, and this pattern of mental states are now instantiated in the numerically separate Replica.

Therefore, I conclude that T1 is not a case involving the persistence relation at all, unless we want to accept the conclusion that the Original's 'mental state pattern' is a separately existing entity that persists through its later instantiation in the Replica—so, there would really be *three* entities in T1 instead of two: the Original, the Replica, and the separately existing 'pattern'. But the reductionist cannot accept this. Instead, the relationship between the Original and the Replica is closer to that held between two numerically unique instantiations of the same pattern. Without granting Parfit's further claim that identity does not matter, a fresh instantiation of the same pattern will not be 'about as good' as identity.

Teletransportation thus presents a problem for the Wide and Widest views of the Psychological Criterion. T1 fits both the Wide and Widest views and ought to count as persistence on both. The Wide View implies persistence because the Teletransportation process is regularly reproducible—it is reliable. And, if the Wide View implies persistence, so does the Widest View *a fortiori*. Yet it cannot be the case that I persist through complete destruction of *all* my parts just in case a timely clone is created. That just is not persistence. Contra Parfit, then, we *do* have good reasons to prefer the Narrow View to the Wide and Widest views, at least if we are interested in theories of persistence, because only the former is actually a persistence theory.

3.2 The Narrow View and Moral Problem Cases

While definitely a view about persistence and not something 'about as good' as it, the Narrow View still faces its own problems. The Narrow View, like most psychological approaches to persistence, treats all mental states as equally constitutive of persistence. Indeed, this is at first blush one of its advantages. Because all mental states contribute in a kind of value-neutral way on the Narrow View, a subject can persist through tremendous changes to any given type of mental state as long as sufficient direct connections among her *overall* psychology are retained. So, the Narrow View accepts that a subject with a suitable number of other direct connections could persist through complete amnesia (in terms of episodic memories), full desire apathy, or sweeping personality shift.

At the same time, I think that there are cases involving change to specifically moral features that might make us doubt whether someone could persist, even if overall psychological continuity obtains. Here's an unproblematic case of persistence for the Narrow View:

> *Mental State Booth*: in a future society, direct manipulation of one's mental states is a simple process. Subjects can enter 'mental state booths', apparatuses that

> scan the user's brain and map his mental states onto a display, whereupon the user can then selectively remove undesirable states and add desirable ones. Perhaps the subject would prefer to replace a desire for sugary foods with one for something healthier, to forget an especially traumatic memory, or to become courageous when he is, in fact, a coward.

In *Mental State Booth*, only a handful of direct connections are severed in a given visit. So, even though these changes occur by way of an abnormal cause—the manipulation of the booth's technology rather than ordinary brain function—this does not contravene the Narrow View's requirement that continuity obtain through the 'normal' cause. There are more than enough remaining connections made via ordinary brain function that overall continuity obtains easily. Suppose, though, that we alter the case:

> *Malfunctioning Mental State Booth:* everything from the description of *Mental State Booth* is the same, but a particular booth has started malfunctioning and performs the requested operation for *all* instances of a given type of mental state. So, Jones enters the booth to remove an unpleasant memory, but she leaves a full amnesiac (again, in terms of episodic memories). Next, Smith enters the booth wanting to be kind rather than inconsiderate but leaves with all of his character traits inverted. (As a matter of convenience, we can consider these traits in the manner of virtues and vices, but I do not think the argument turns on this).[4] For both Jones and Smith, all non-targeted types of mental state are unaltered. Let us further suppose that at least half of all of both subjects' total mental states are unchanged.

Malfunctioning Mental State Booth stipulates that at least half of both subjects' total psychology is unchanged so that there will still be *strong* connectedness between the pre- and post-booth subjects. Recall that without strong connectedness, there can be no continuity on the Psychological Criterion. Given this stipulation, I hope that it is clear that the Narrow View has it that both Jones and Smith persist as one and the same subject when they leave the booth as when they entered it. That is, in both cases, the subject who enters is psychologically continuous with the subject who leaves, and this continuity obtains through normal brain functioning. Despite the manipulation of a substantially greater number of states than the vanilla *Mental State Booth* case, continuity still obtains because the vast majority of unaltered connections between mental states *do* still hold.

Now, I propose that this case is just as unproblematic for the Narrow View as *Mental State Booth*. And I further think that the fact that the Narrow View handles *Malfunctioning Mental State Booth* so easily is not a point in its favor. That Jones could persist through her amnesia is intuitively plausible in a way that

[4] It may be helpful to see a few examples of moral character traits to make both *Malfunctioning Mental State Booth* and the discussion in later sections clearer to the reader. Possible character traits—paired with their moral opposites, for brevity's sake—might include: kind-ness/meanness, empathy/indifference, tolerance/intolerance, respectfulness/insensitivity, compassion/cruelty, generosity/selfishness, sincerity/deceitfulness, fairness/inequity and temperance/profligacy. This list is by no means exhaustive, and I think that the rabbit hole of possible character traits can be as deep as you like. Consider peculiarly specific traits along the lines of *loves animals mostly but is exceptionally cruel to squirrels* or *kind only to convenience store clerks*. It seems next to impossible to classify such traits under broader headings just because they are built around exceptions.

Smith's persisting through complete shift in moral character is not. Why? In Smith's case, the post-booth subject has a radically different relationship with others and the world from the pre-booth subject. In fact, the pre- and post-booth subjects will exhibit wholly *opposite* responses and behaviors. The post-booth subject will not act in Smith's characteristic way. I think that this indicates that, perhaps unlike other mental states, selectively altering only moral features does not *ceteris paribus* amount to persistence. The Narrow View is not sensitive to this apparent problem, though, because it treats all mental states as equally constitutive of persistence.

Cases like *Malfunctioning Mental State Booth* imply that the Narrow View is wrong to adopt a value-neutral stance toward mental states: some states—namely, moral features—may actually be more important than others in determining persistence. That a certain kind of 'moral connectedness' might be a necessary condition for persistence is not an altogether strange idea, though it may sound that way on the face of it. We often talk informally as if moral character plays a special part in persistence, such as when the convict emerges from a prison sentence and earnestly declares himself 'a new man', or when the religious convert is said to be 'another person' entirely after her conversion experience.

Of course, we could be talking figuratively in these cases. Yet, there *is* empirical evidence that people do actually regard moral features with just this sort of constitutive significance to persistence. For one, discontinuity of moral features is seen as most disruptive to the persistence of other individuals when compared to discontinuity of features that have traditionally received more attention in the persistence literature, including memories, agency, the ability to construct coherent self-narratives, and non-moral personality traits (Strohminger and Nichols 2014, Prinz and Nichols 2016). The same results are observed in studies that ask respondents about the persistence of oneself instead of others (Prinz and Nichols 2016, Molouki and Bartels 2017). Moreover, friends and family members of individuals suffering from neurodegenerative diseases also report that moral change is most threatening to the persistence of their loved ones, even in cases of near total memory loss (Strohminger and Nichols 2015).

Likewise, data suggest that people tend to view changes to specifically moral beliefs—especially beliefs widely-held by large communities—as more impactful on persistence than changes to other types of belief, leading researchers to propose a kind of moral essentialism in how people perceive themselves, others, and their persistence (Heiphetz, Strohminger, and Young 2016). This is further evident in research that shows that people believe moral features in large part constitute a deep "true self" that represents the features an individual *really* has, even if they are not always explicitly expressed (Strohminger, Newman and Knobe 2017). Additional studies link up with claims about moral essentialism, indicating that people intuitively see moral features as among the most causally central of the many features ordinarily taken to be part of identity (Chen, Urminsky and Bartels 2016) and that perceived moral character contributes substantially to person perception—the impressions we form about the kind of person someone else is (Goodwin, Piazza and Rozin 2014, Goodwin 2015).

Now, it is important to note that none of the authors cited above argue that moral features are the *only* meaningful features in determinations of persistence. Indeed, the same data support the notion that non-moral psychological features typically privileged in the persistence discussion—like memories, desires, and

intention—play a valuable role in persistence. As will become clearer in Section 4, I also think that an account of persistence must take stock of these non-moral psychological features. My point in calling attention to the empirical evidence is to give a basis for the plausibility of a necessary moral persistence condition that goes beyond the intuitive force of my proffered counterexamples to the Narrow View, like *Malfunctioning Mental State Booth*. If nobody actually believed moral features made a special contribution to persistence, then my position would be much harder to accept, on the face of it. But beliefs about the significance of moral features to persistence are in fact fairly widespread.

Of course, that people think that moral features are importantly constitutive of persistence is not proof of a necessary moral persistence condition. What people think might be wrong. I am not here prepared to make the stronger claim that Jesse Prinz and Shaun Nichols (2016: 450) do when they say that our beliefs about persistence actually determine the conditions of persistence. However, the possibility that people might be wrong in their persistence beliefs does not definitively mean that they *are* wrong, either. Given that intuitions about a necessary moral persistence condition appear to be widespread, I do not think it *prima facie* implausible to postulate such a condition as an explanation for what seems to be missing in certain problem cases in which the individual intuitively does not persist, despite the fact that overall psychological continuity obtains. Parfit presents cases like Teletransportation and the Combined Spectrum to make reductionism more palatable; reductionism is, he argues, the answer that makes the most sense in these cases. Analogously, my strategy is to suggest cases that reveal explanatory weakness in the Narrow View—by holding all mental states to be value-neutral, the Narrow View gets the wrong result in cases involving severe changes localized to moral features.

3.3 A More Grounded Example: Moral Change and Traumatic Brain Injury

An immediate rejoinder from the advocate of the Narrow View might be that cases like *Malfunctioning Mental State Booth* are just too fanciful to offer much trouble. Selectively altering mental states in the way of *Malfunctioning Mental State Booth* is sci-fi conjecture; in all *actually* possible cases, the Narrow View gets it right because our moral features and the rest of our psychological profile just do not ordinarily come apart like that. We should not be so easily convinced by the advocate, as there in fact *are* real-world cases where there appear to be great changes in character while other mental features more or less stay the same. In particular, this can happen in cases involving traumatic brain injury. One classic example that is often trotted out is the case of Phineas Gage, the railroad worker whose head was impaled by a tamping iron in 1848. The injury allegedly resulted in extreme change (for the worse), primarily in his character traits but also among other mental features. Currently, there is skepticism about whether Gage's purported changes in character, or the degree to which he was famously claimed to be 'no longer Gage' actually held true (Macmillan and Lena 2010). Whether or not Gage's accident really caused profound psychological changes, one can imagine a case where a similar event *does* cause such changes. Indeed,

we do not have to look far to find a timelier example with Gage-like circumstances.[5]

Alissa Afonina is a woman sensationalized by the tabloids as turning "from a star student to a lusty dominatrix" (Natalie O-Neill, *New York Post*, Feb. 6 2015). If we set aside the ridiculous exaggerations of yellow journalism, we find that court documents from the case really do support a dramatic shift in moral character after a brain injury Afonina suffered during a car accident. In his summary of *Afonina v. Jansson*, Presiding Justice Joel Groves remarks that, prior to the accident, Afonina's Grade 11 teacher, Mr. Byrne,

> found Alissa to be a student who was very bright and interested in her pursuits, and in matters generally. He described her as being in the top 2% in terms of engagement in class activities and assignments [and] testified that he had no sense that she was troubled emotionally in any way (*Afonina v. Jansson,* 2015 BCSC 10, section 114).

After the accident, however, Afonina's character changed remarkably, as her teacher now

> observed a very different Alissa. He said that she showed signs of no impulse control […]. She became socially isolated and began to have outbursts in class. She made sexual comments during these outbursts that were inappropriate for the class setting. Mr. Byrne was of the view that she did not appear to filter her thoughts and acted as if she was unaware of her social environment [...]. Her talk was unfiltered, random, and as he described, not logical for the school social environment, or "out of left field" (*Afonina v. Jansson,* 2015 BCSC 10, section 116).

Justice Groves was convinced by these claims and corroborating evidence presented by Afonina's legal team that her accident indeed caused a pronounced change in character in the form of a personality disorder with symptoms that ranged among "emotional liability, cognitive fatigue, reduced insight, reduced judgment, disinhibition, mood swings, apathy, and inflexible thinking" (*Afonina v. Jansson,* 2015 BCSC 10, section 129). Because of her disorder, Afonina had tremendous difficulty holding down a job and became increasingly depressed and unmotivated. Eventually, she turned to sex work to support herself financially, including—as the tabloids were so ruthlessly quick to point out—work as a dominatrix for hire.

As Afonina's unfortunate case illustrates, there are legitimate Gage-type cases that mirror the consequences of the more far-fetched *Malfunctioning Mental State Booth*. Before her accident, Afonina appears to have been a driven and creative young woman mindful of her education—at least in areas of interest to her—and concerned about how she related to social peers. Post-accident,

[5] I use Alissa Afonina as a but one instance among a multiplicity of cases involving massive changes in moral character following traumatic brain injury. Other representative examples include a man who developed spontaneous and uncontrollable sexual aggression, including pedophilia, due to a tumor (Burns and Swerdlow 2003), Mr. L., who fell from his roof and later began displaying unprovoked hostility and obsessive jealousy toward his wife (Salas 2012), and American photographer Edward Muybridge, who became dishonest, unstable, and aggressive to the point of murder after a stagecoach accident (Manjila et al. 2015).

though, Afonina became listless, intellectually closed-off, and, most tellingly, was highly inconsiderate of others. Not all of these changes are overtly changes in moral character; Afonina also suffered an incapacity for retaining new memories and saw changes to personality traits that are decidedly morally neutral.

At the same time, there is no clear indication that she endured anything like widespread loss of or change to other mental features. For instance, Afonina did not report amnesia-like loss of episodic memory, and there seemed to be no significant change in her foundational belief structure. It could even be argued that many of her desires—such as aspirations towards filmmaking and other creative media—were still present and accounted for, as Afonina *did* continue pursuing these interests post-accident in a specialized program for skilled students. It was only when she found herself unable to fulfill these desires due to cognitive impairment from the accident that she became frustrated and withdrew. This means that there appears to be overall psychological continuity in Afonina's case, and this continuity is the result of normal brain functioning. The parts of the brain responsible for Afonina's unchanged features presumably continued to function normally, even if there is abnormal function in changed areas. So, since the pre- and post-accident subject are psychologically continuous by way of the normal cause, the Narrow View would pronounce that Afonina persisted through the accident.

Still, we can see that, as in *Malfunctioning Mental State Booth*, changes to Afonina's character traits at least in part led to an extreme difference in how she participated in and related to her social ecosystem. In particular, her failure to meet standards of social propriety looks to indicate a lack of concern for others. It was not that Afonina no longer recognized what was socially expected of her. Rather, she just did not seem to care anymore whether her words and actions negatively affected those who shared her classroom. There is a whole web of character traits involved here: kindness, respect for others, and concern for the welfare of one's peers, probably among many more.

In light of this, I suggest that to say that the post-accident person is one and the same Alissa Afonina is dissonant with her behavioral changes and the underlying disruptions to her character that they represent. Despite the fact that the pre-and post-accident subjects seem to be psychologically continuous, the post-accident subject's behavioral changes are both quite severe and of such a nature that they make the attribution of sameness seem unnatural—substantially more so than if the psychological changes were, say, solely to learning aptitudes. After all, it is not her learning disabilities that made the post-accident 'Afonina' so unnerving to Mr. Byrne and others, but her extraordinary and sudden character changes. Thus, while the Narrow View would unproblematically affirm that Afonina persisted through her accident, there does seem to be something wrong with this assessment. Even if Afonina is psychologically continuous overall, the discontinuities in her moral features challenge the Narrow View's claim of persistence.[6] Hence, from counterexamples like *Malfunctioning Mental State Booth*

[6] It is prudent to call attention to a potential objection here. Perhaps one could say that moral features are interlinked with a whole network of other mental states: beliefs, desires, etc. If this is the case, loss of moral features would require the simultaneous loss of too many other features with which those moral features are interwoven—and so, the whole idea of selective change in mental states, on which examples like *Malfunctioning Mental State Booth* and the Afonina case depend, is incoherent. Ultimately, I do not think

and Gage-type injury cases like Afonina's, I conclude that the Narrow View—like the Wide/Widest views—is also unsuccessful as a persistence theory.

4. How the Psychological Criterion Might Succeed

Though I have argued that the Psychological Criterion—whether on the Narrow, Wide, or Widest views—fails as a persistence theory, I *do* think that at least the Narrow View can be appropriately modified to account for the above problem cases. To close the paper, I will offer a broad suggestion as to how this might work. In short, I posit the amendment of the Psychological Criterion with a second sort of continuity, *moral continuity*, as a necessary condition on persistence. This results in a slightly expanded Narrow Psychological Criterion that I call the *Narrow Moral Psychological View*.

Cases like *Malfunctioning Mental State Booth* and more grounded Gage-type cases show that discontinuity among moral features appears incompatible with the claim that such-and-such subject persists. In turn, these cases imply that a kind of moral continuity might be *necessary* for persistence. Such cases do not, however, imply that moral continuity is *sufficient* for persistence. To see why, consider the following case:

> *The Moral Accident Victim*: a man, Quaid, is involved in a catastrophic accident that results in a peculiar impairment. Quaid loses access to all previously held memories, beliefs, desires, intentions, and personality traits. In fact, the only previously held mental features that remain are his moral character traits. Quaid still interacts with others in his social environment with the same warmth, kindness, and generosity that he did before, even though he does not remember anything about himself or others, lacks a belief structure that verifies that these acts are things that he ought to do, and has no real desire to do these things.

Quaid is not psychologically continuous with the post-accident man. There are exceedingly few direct connections between them. Despite the fact that his behavior has not changed, we can and should say that the post-accident man does not relate to others in the world in the way that the pre-accident man did; indeed, just about all of his relations with others in the world were severed by the accident. So, it is clear that we cannot accept that Quaid persisted through the accident if we believe any sort of psychological continuity is necessary for persistence. And *The Moral Accident Victim* implies that, absent psychological continuity, moral continuity is not sufficient for persistence. There just is not enough of 'Quaid' left over, and this is clear by the impoverished nature of the post-accident man's relations to the world.

Between Gage-type cases and *The Moral Accident Victim*, I can now make a case for moral continuity as *necessary but not sufficient* for persistence. From here, the simplest way to define moral continuity and integrate it into the Narrow View with minimal alterations is as a kind of subset of standard psychological continuity. So, this would mean that moral continuity consists in the following parallel premises to the Psychological Criterion:

this objection goes through, though I lack the space in this paper to either fully dissect the objection or present suitable responses. I hope that it will satisfy the reader here to just anticipate the objection and say that I believe I have answers for it.

> *Moral Continuity*: There is moral continuity among X today and Y at some past time if and only if (1) there are overlapping chains of strong connectedness (of specifically moral features) between these subjects, (2) this continuity has the right kind of cause, and (3) there does not exist a different person who is also morally continuous with Y. (4) Moral continuity just consists in the holding of facts like (1) to (3).

My main intention in giving moral continuity this formulation is to clearly convey that moral continuity does not preclude *any* moral change at all. If the Psychological Criterion is to largely retain its character, the notion of moral continuity must permit the same kind of gradual change that the Psychological Criterion does as a whole. Beyond this, (2), (3), and (4) have the same consequences as their counterparts in the Psychological Criterion. Given this formulation, we can append a clause to the standard Narrow View that includes the moral continuity claim as a necessary condition on persistence, giving us

> *The Narrow Moral Psychological View of Persistence*: (1)' There is psychological continuity if and only if there are overlapping chains of strong connectedness. (2)' There is moral continuity if and only if there are overlapping chains of strong connectedness among the relevant moral features. X today is one and the same subject as Y at some past time if and only if (3)' X is psychologically continuous with Y, (4)' X is morally continuous with Y, (5)' these continuities have the right kind of cause, and (6)' there does not exist a different subject who is also psychologically and morally continuous with Y. (7)' Persistence just consists in the holding of facts like (3)' to (6)'.

We can see, then, that only a little refiguring is needed to get the Narrow View to the more precise Narrow Moral Psychological View.

All the same, I think that this small change is enough to comfortably integrate the necessity of moral continuity to persistence in a way that satisfies the examples from the previous section. For instance, Alissa Afonina does not persist as the post-accident victim, despite the overall psychological continuity she and the post-accident victim enjoy. Rather, the accident severed moral continuity between Afonina and the post-accident victim. On the Narrow Moral Psychological View, X cannot be one and the same subject as Y unless moral continuity obtains. So, moral continuity is necessary for persistence. However, persistence still requires overall psychological continuity, as in the ordinary Narrow View, and moral continuity is not *alone* sufficient for persistence.

With this modification, I believe that the Narrow View succeeds as a persistence account. Perhaps just as importantly, this modification largely preserves the spirit of Parfit's original Narrow Psychological Criterion. There may be other ways to salvage the Narrow View in light of the problem cases I posited in the previous section; this is just a brief attempt to resolve the difficulties those cases raised. Contrarily, I do not think there is anything to be done with the criticisms leveled against the Wide and Widest views because those views ultimately run too far afield of persistence. As I mentioned before, this is no real problem for Parfit: to him, the Wide/Widest views were mostly a springboard for jumping headlong into the controversial—but highly influential—conclusion that persistence identity does not matter. Yet, in arguing that only the Narrow View is even a persistence account and then modifying that view to accommodate moral problem cases, I would like to take steps to shift discussion of Parfit back to per-

sistence. Identity *does* matter to many of us, and, with the right adjustments, his Psychological Criterion is well-positioned to show us why.

References

Burns, J.M. and Russell, H.S. 2003, "Right Orbitofrontal Tumor with Pedophilia Symptom and Constructional Apraxia Sign", *Arch Neurol.,* 60, 437-40.

Chisholm, R. 1969, "The Loose and Popular and the Strict and Philosophical Senses of Personal Identity", in Care, N.S. and Grimm, R.H. (eds.), *Perception and Personal Identity*, Cleveland: Press of Case Western Reserve University, 82-106.

Chen, S.Y., Urminsky, O. and Bartels, D.M. 2016. "Beliefs About the Causal Structure of the Self-Concept Determine Which Changes Disrupt Personal Identity", *Psychological Science*, 27, 1398-1406.

Goodwin, G.P. 2015, "Moral Character in Person Perception", *Current Directions in Psychological Science,* 24, 38-44.

Goodwin, G.P., Piazza, J. and Rozin, P. 2014, "Moral Character Predominates in Person Perception", *Journal of Personality and Social Psychology*, 106, 148-68.

Heiphetz, L., Strohminger N. and Young, L.L. 2016, "The Role of Moral Beliefs, Memories, and Preferences in Representations of Identity", *Cognitive Science*, 41, 744-67.

Macmillan, M. and Lena, M.L. 2010, "Rehabilitating Phineas Gage", *Neuropsychological Rehabilitation,* 20, 641-58.

Manjila, S., Gagandeep, S., Ayham, M.A. and Ramos-Estebanez, C. 2015, "Understanding Edward Muybridge: Historical Review of Behavioral Alterations after a 19th-Century Head Injury and Their Multifactorial Influence on Human Life and Culture", *Journal of Neurosurgery,* 39, 1-8.

Molouki, S. and Bartels, D.M. 2017, "Personal Change and the Continuity of the Self", *Cognitive Psychology*, 93, 1-17.

Parfit, D. 1971, "Personal Identity", *The Philosophical Review,* 80, 3-27.

Parfit, D. 1984, *Reasons and Persons*, New York: Oxford University Press.

Prinz, J. and Nichols, S. 2016, "Diachronic Identity and the Moral Self", in Kiverstein, J. (ed.), *The Routledge Handbook of Philosophy of the Social Mind*, New York: Routledge, 449-64.

Salas, C.E. 2012, "Surviving Catastrophic Reaction after Brain Injury: The Use of Self-Regulation and Self-Other Regulation", *Neuropsychoanalysis*, 14, 77-92.

Strohminger, N. and Nichols, S. 2014, "The Essential Moral Self", *Cognition*, 131, 159-71.

Strohminger, N. and Nichols, S. 2015, "Neurodegeneration and Identity", *Psychological Science*, 26, 1469-79.

Strohminger, N., Newman, G. and Knobe, J. 2017, "The True Self: A Psychological Concept Distinct from the Self", *Perspectives in Psychological Science*, 12, 551-60.

Wiggins, D. 1967, *Identity and Spatio-Temporal Continuity*, Oxford: Blackwell.