

## Moral Responsibility without Personal Identity?\*

This is a post-peer-review, pre-copyedit version of an article published in *Erkenntnis*. The final authenticated version will be available online at: [http://dx.doi.org/ DOI: 10.1007/s10670-018-0092-7](http://dx.doi.org/10.1007/s10670-018-0092-7).

### 0. Introduction

Many people think that one can only be responsible for what one did *oneself*. This seems to suggest

ENTAILMENT: Moral responsibility presupposes personal identity.

ENTAILMENT is very plausible. John Locke, for example, famously *characterizes* PERSONAL IDENTITY as a *forensic concept*

“*appropriating actions and their merit*, and so belongs only to intelligent beings, capable of a law, and happiness, and misery. This personality extends itself beyond present existence to what is past, only by consciousness – whereby it becomes concerned and *accountable*; *owns* and *imputes to itself* past actions, just upon the same ground and for the same reason as it does in the present.” (Locke 1694: Book 2, Chapter 27, § 26)

Despite its initial plausibility, however, ENTAILMENT is threatened by Derek Parfit’s (1984: 261-264) famous considerations for the thesis that personal identity isn’t what matters for our practical concerns. While Parfit himself discusses moral responsibility only in passing, rejection of ENTAILMENT has recently been championed by arguments developed by David Shoemaker (2012) in a Parfitian spirit.<sup>1</sup> According to these arguments, all accounts of personal identity have counter-intuitive implications when it comes to moral responsibility, so ENTAILMENT must be abandoned.

---

\* For helpful feedback, comments, and discussion I would like to thank Vuko Andrić, Radim Bělohrad, Cameron Boulton, Ben Ferguson, Mike Ridge, Neil Roughley, Hanno Sauer, Christine Tiefensee, the participants of the Oberseminar Philosophische Anthropologie und Ethik, the audience of my talk at the OZSW 2017, as well as anonymous referees.

<sup>1</sup> Shoemaker calls ENTAILMENT “Slogan.”

This paper does two things: First, it argues that the arguments against ENTAILMENT only apply to particular *kinds* of views about personal identity, namely reductionist views. Non-reductionist views are unaffected by those problems and, hence, do not threaten ENTAILMENT's truth. Second, it argues that non-reductionist views do so only at the cost of undermining the legitimacy of ascriptions of moral responsibility. However, the paper introduces a novel view about personal identity that can also preserve ENTAILMENT's truth, but isn't affected by the worries that plague non-reductionism. This is non-representationalism about personal identity, according to which we shouldn't account for the concept PERSONAL IDENTITY in terms of what it represents, but in terms of a distinctive non-representational function. The paper also fleshes out one variant of this view, inspired by meta-ethical expressivism, on which judgements about personal identity are plans for holding to account.

The paper proceeds as follows: Section one clarifies ENTAILMENT and presents the problems that affect it. Section two argues that these problems don't speak against ENTAILMENT, but rather against reductionist views about personal identity. A non-reductionist view can escape the arguments. Section three then introduces and motivates non-representationalism about personal identity as a way of preserving ENTAILMENT, without succumbing to worries about non-reductionism.

## 1. Troubles for ENTAILMENT

We should start by clarifying ENTAILMENT.<sup>2</sup> First, I will assume that identity is *numerical* identity in this context and – following common practice in the debate about personal identity – I will understand “person” as the kind of thing beings like

---

<sup>2</sup> What follows draws on D. Shoemaker, 2012. The examples are my own, but built on Shoemaker's.

you and me are essentially.<sup>3</sup> So, ENTAILMENT says that moral responsibility presupposes numerical identity of persons. How should we understand “moral responsibility” in this context, though?

Assessments of morally responsibility have various dimensions. Suppose Black was killed with a candlestick in the conservatory. One question is whether *anyone* is responsible for this. Maybe Black’s pet monkey accidentally got loose and killed Black. Even if it is clear that *someone* killed Black, however, one still needs to determine *who* – whether it was Plum or Mustard, for example. This, though, *still* doesn’t fully settle who is responsible. If, for example, Plum killed Black because factors beyond his control turned him insane, he is the proper *candidate* for bearing responsibility, but relevant factors settle that he isn’t responsible.

This shows that we can distinguish at least two issues regarding moral responsibility.<sup>4</sup> The first is relevant when figuring out *who* committed a deed, i.e. to whom it is in some relevant sense *attributable*. This is what we are concerned with when we ask whether it was Plum, Mustard or the monkey who killed Black. Here we figure out who “owns” a particular deed. The issue here is *attributability*.

The second relevant issue is whether it is fitting to *hold someone to account* for that particular deed. Note that “holding to account” can be understood differently: for example, it might be understood in terms of the reactive attitudes (e.g. Strawson, 1962) or in terms of a modification of relationships (e.g. Scanlon, 2008). I will remain neutral on that issue. Whether it is fitting to *hold someone to account* is what we are concerned with when we ask whether factors beyond Plum’s control turned him insane. This issue is *accountability*.

---

<sup>3</sup> Why assume that identity means numerical, rather than narrative identity here? First, I take this to be more plausibly what ENTAILMENT is about. Second, narrative identity plausibly presupposes personal identity, so even if one holds that moral responsibility presupposes narrative identity, ENTAILMENT understood in terms of numerical identity still needs to be true.

<sup>4</sup> The distinction and terminology goes back to Gary Watson (1996).

These issues are connected: It is *fitting* to hold someone to account for something, only if it is attributable to them. As Shoemaker highlights, this is a more basic claim than ENTAILMENT, namely

ACCOUNTABILITY & OWNERSHIP:<sup>5</sup> One can be accountable only for one's own actions.

Furthermore, with ACCOUNTABILITY & OWNERSHIP on the table, we can see that much of ENTAILMENT's plausibility seems to derive from it being an *interpretation* of one of the notions in ACCOUNTABILITY & OWNERSHIP, namely of what is involved in an action being "one's own." On this reading, ENTAILMENT says that *attributability* presupposes personal identity. The question is, consequently, whether ENTAILMENT so understood is true. So, let us now consider what speaks against ENTAILMENT's truth.

While problems for ENTAILMENT are already implicit in Parfit's arguments for the claim that identity isn't what matters, Shoemaker (2012) has recently articulated and discussed these problems more explicitly. Because Shoemaker provides the clearest, most elaborate and explicit formulation of the worries applying to ENTAILMENT, I focus on his argument.

The argument has two steps and starts with the following assumption: If ENTAILMENT is true, a statement of the following form should be plausible for some view about personal identity:

ATTRIBUTABILITY: What makes some past action one's own now is \_\_\_\_\_.

---

<sup>5</sup> Shoemaker (2012: 109) calls this claim "PLATITUDE." However, while the claim is, on first sight, quite plausible and accepted by many people (Shoemaker himself accepts it (2012: 122)), it is unclear whether it deserves to be called a *platitude*. After all, its plausibility depends very much on how we understand what it is for an action to be "one's own" and any reading is likely to be controversial. Hence, I've chosen a different label for this claim. I thank a referee for *Erkenntnis* for pushing me on this issue.

In the first step, Shoemaker argues that the major views on personal identity cannot plausibly complete ATTRIBUTABILITY. Shoemaker considers the biological view and the psychological view.

According to the *biological view*,  $x$  at  $t_1$  is the same person as  $y$  at some later time  $t_2$ , if and only if  $y$  is the same biological organism as  $x$  (proponents of the biological criterion are e.g. David DeGrazia (2005) and Eric Olson (1997)). However, there are plausible counter-examples to completions of ATTRIBUTABILITY with the biological view. For example, on the biological view, if Plum's brain is transplanted into Mustard's skull, Plum *wouldn't* be the person that is a combination of Plum's brain and Mustard's body, but Plum's brainless body (assuming it is kept alive). Hence, if Plum kills Black prior to the transplant, Black's murder should be attributable to the brainless body, not the person with Plum's brain. This seems false.

According to the *psychological view*,  $x$  at  $t_1$  is the same person as  $y$  at some later time  $t_2$ , if and only if  $y$  is uniquely psychologically continuous with  $x$ , where psychological continuity consists in overlapping chains of sufficient psychological connections (proponents of the psychological criterion are e.g. David Lewis (1983), Derek Parfit (1984) and Sydney Shoemaker (1984)). Again, though, this view faces counter-examples as a completion of ATTRIBUTABILITY. Consider a case of gradual fugue. Assume that, because of tumors that grow and wane, Plum gradually enters and exits a fugue state that lasts two years. During that period, the individual biologically continuous with Plum is called "Green." Green and Plum differ radically in their psychology. Assume that Green kills Black and that Plum has apparent memories of this, but rejects it as something done by someone else. However, because the change is gradual Black's murder should be attributable to Plum, if the psychological view were the correct completion of ATTRIBUTABILITY. This seems false.

This completes the argument's first step: the major views about personal identity have implausible implications regarding attributability. The second step strengthens this result via the fission and branch-line cases that speak against *any* completion of ATTRIBUTABILITY that rests on personal identity.

In the fission case, Plum splits in the middle and the halves re-grow their missing parts. These resulting halves, call them "Left-Plum" and "Right-Plum," are psychologically and biologically exactly similar to, and continuous with, Plum. Shoemaker argues that they cannot *both* be Plum, because identity is one-one, not one-many. So, what happens to Plum? According to Shoemaker, the only plausible claim on *any* account is that neither Left-Plum nor Right-Plum is Plum. Consequently, if Plum murdered Black, on no completion of ATTRIBUTABILITY that invokes personal identity is Black's murder attributable to Left-Plum or Right-Plum. According to Shoemaker, this is false, because Plum's actions are (at least partially) attributable to *both* Left-Plum *and* Right-Plum. If this is right, any account of personal identity has implausible implications regarding attributability.

Shoemaker's second counter-example is a branch-line case. Suppose that after Plum killed Black, scientists made Duplicate-Plum, a perfect duplicate of Plum at that time. Some days later, these scientists kill Plum, vaporize his body, and insert Duplicate-Plum in his place. Duplicate-Plum is never Plum. Hence, on no completion of ATTRIBUTABILITY that invokes personal identity is Black's murder attributable to Duplicate-Plum. Shoemaker thinks, however, that Black's murder is attributable to Duplicate-Plum.

This completes the second part of Shoemaker's argument. If the argument goes through, we should abandon ENTAILMENT. In what follows, however, I will argue that the argument doesn't go through, because at least two views about personal identity aren't touched by the problems raised for ENTAILMENT and these

problems only show something about the views Shoemaker considers, rather than about ENTAILMENT.

## 2. The Non-Reductionist Alternative

Let me first highlight the strength of ENTAILMENT's initial credibility, by noting how radical it is to abandon it. Fundamentally, the issue this debate is concerned with is what factors determine whether someone is accountable for something. Many factors might be relevant here, e.g. whether it was an intentional action, done out of free will, etc. Denying ENTAILMENT amounts to denying that *personal identity* is such a factor. But note that this implies that, for example, for *you* to be accountable for something it isn't relevant that *you* did it. This is incredible and we'd need *very* good reasons to accept it. The crucial question is whether the problems raised by Shoemaker's argument provide such reasons. Let's take the argument's steps in turn.

The first step proceeds via counter-examples to particular views about personal identity. I won't dispute these examples. However, it is noteworthy that they don't merely speak against the biological and psychological views' ability to capture our intuitions about attributability. They *also* speak against their ability to capture our intuitions about personal identity.

Consider the counter-example to the biological view: after Plum kills Black, Plum's brain is transplanted into Mustard's skull. According to the biological view, the brainless body (if it is kept alive) is *Plum*, while the person with Plum's brain is *Mustard*. However, it isn't merely implausible that Black's murder is attributable to Plum's brainless body, rather than to the person with Plum's brain. It is *also* implausible that Plum *is* Plum's brainless body, rather than the person with Plum's brain.

Second, consider the counter-example to the psychological view: Plum gradually enters and exits a fugue state, in which the individual biologically continuous with him is called “Green.” Green, whose psychology differs radically from Plum’s, kills Black. According to the psychological view, Green *is* Plum, because even though their psychologies differ radically, the change from Plum to Green and back is gradual. However, it isn’t merely implausible that Black’s murder is attributable to Plum. It is *also* implausible that Plum *is* Green. While clearly not any gradual psychological change threatens personal identity, intuitively the change from Plum to Green and back *is* such a change.

So, here our intuitions about *both* attributability *and* personal identity deviate from these views. Furthermore, these intuitions go hand in hand: our intuitive judgements about attributability follow our judgements about personal identity in these cases, just in the way ENTAILMENT suggests. That is, our intuitions about these cases suggest that actions are attributable to Plum if and only if *Plum* performed them. Hence, that neither of the considered views captures our intuitions about attributability, doesn’t speak against ENTAILMENT. Rather, the thought-experiments *support* ENTAILMENT. Another good explanation for the failure of these views is, consequently, that ENTAILMENT is true, but none of these proposals captures personal identity.

Of course, if the prominent proposals about personal identity cannot capture these intuitions, this might undermine those intuitions. This, however, is correct only if Shoemaker’s picture of the kinds of views about personal identity we could endorse is exhaustive. But, he actually only considers *reductionist* views.

Reductionist views about personal identity give accounts of personal identity in *impersonal* terms, terms that don’t employ, in one form or another, the concept PERSONAL IDENTITY. If you will, reductionists hold that an ontology fully describable

without mentioning personal identity can fully account for facts about personal identity. Personal identity facts aren't *sui generis*.

The biological and the psychological views are the most prominent forms of *reductionism*. But, if *these* views face counter-examples as accounts of personal identity and of attributability, this tells us only that *reductionist* views are problematic.<sup>6</sup> This is evidence against ENTAILMENT, only if reductionist views are the only candidates for accounting for personal identity.

This can be challenged. Obvious alternatives are *non-reductionist* views about personal identity, according to which we cannot fully account for personal identity in impersonal terms (contemporary non-reductionists are E.J. Lowe (2012), Geoffrey Madell (1981, 2015), Trenton Merricks (1999), and Richard Swinburne (1984)): to fully account for personal identity we need to enrich our ontology with *sui generis* personal identity facts. This is often seen as the view that we are Cartesian Egos or souls, *featureless* entities existing separately from minds and bodies. However, it will be helpful to quickly look at Geoffrey Madell's (2015) view, who rejects this interpretation.

On Madell's account, the core issue regarding personal identity is the clash between our first-person and an objective, third-personal perspective. The crucial question, so Madell holds, is whether impersonal criteria can settle what part of reality impersonally described is uniquely *me* (which arises for any self). According to Madell, we can adequately capture that some part of impersonally described reality is *me*, only by introducing irreducible "indexical" facts into our ontology. Specifically, "I" is a primitive indexical, which conveys irreducible indexical facts about whether some impersonally described thing is *mine*. On Madell's view, there can be no *criteria*

---

<sup>6</sup> Strictly speaking, of course, the counter-examples only show that the specific versions of reductionism considered here are problematic. However, given that these are the most prominent, and widely considered to be best, versions of reductionism (and, in fact, are often taken to exhaust the space of plausible reductionist options), I take it that this constitutes good evidence for the view that reductionist views in general are problematic.

for personal identity, because facts about something being mine are primitive and so the connection between the self and the properties it has at a time and over time are *contingent*.

What can we say about personal identity on Madell's account? On his view, I exist over time if and only if there is a series of conscious experiences over time that are uniquely mine (and, of course, you persist if there is a series of conscious experiences over time which are uniquely yours (See Madell, 2015: 107 and 130)). However, there is no impersonal criterion *what* series of experiences this is. Note, furthermore, that on this view no separately existing Egos are required for personal identity. In fact, the view explicitly speaks against this, because if a series of experiences being *mine* is irreducible, it also isn't reducible to ownership of Egos.

For non-reductionists like Madell it isn't surprising that reductionist views have counter-intuitive implications, because for any description of the world in impersonal terms, the facts about personal identity can differ. An impersonal description of the world leaves something out, namely the persistence of selves, only describable in terms of the relevant primitive indexical facts. Using examples similar to Shoemaker's supports this. Hence, the intuitions that are problematic for reductionists actually support non-reductionism. It should be clear that non-reductionism faces no such counter-examples.

What about the connection to attributability? Insofar as our intuitions about personal identity and about attributability go hand in hand (as they do in the counter-examples), non-reductionism captures and accounts for the connection. Further support derives from the *prima facie* implausibility that I could be accountable for something *I* didn't do. Non-reductionism could capture this connection, because on this account ATTRIBUTABILITY is to be completed as

ATTRIBUTABILITY<sub>NON-RED</sub>: What makes some past action mine now is that I performed it.

Consequently, we've identified a view about personal identity that provides a plausible (yet uninformative) completion of ATTRIBUTABILITY *and* explains why other views fail. The first part of Shoemaker's argument fails. Of course, this only partially defuses the argument. After all, the second part is supposed to speak against *all* completions of ATTRIBUTABILITY. Hence, maybe it also troubles non-reductionism.

Let's start with the branch-line case. Here Plum kills Black but is later killed by scientists who insert Duplicate-Plum in Plum's place. Duplicate-Plum isn't Plum, but Shoemaker argues that Black's murder should still be attributable to Duplicate-Plum:

"The pull of this intuition has its most compelling source in a vivid representation of the phenomenology of the duplicate. He will come into existence full-blown thinking that he is [Plum], and he will (quasi-)remember [Plum's] actions, will delight in thinking that he's gotten away with the immorality, will carry out [Plum's] intention to celebrate the immorality, and so forth. Once we imaginatively project ourselves into his mental terrain and appreciate that he is psychologically exactly similar to the *real* [Plum], it becomes extraordinarily difficult to resist the verdict that [Plum's] actions are properly attributable to him as well." (D. Shoemaker, 2012: 120)

However, I neither see this intuition's pull nor how vivid representation of Duplicate-Plum's phenomenology helps (and Shoemaker offers no further grounds for this intuition (see 2012: 120)). It just shows that Duplicate-Plum *mistakenly thinks* that he is Plum. But nothing follows from *this*. Note, for example, that Duplicate-Plum's thoughts don't differ from *yours* if someone brainwashed you into having a

mental life similar to Plum. If Shoemaker's claim about Black's murder being attributable to Duplicate-Plum was correct, it should equally be attributable to you in that case. But surely murders wouldn't suddenly become attributable to people if they are brainwashed to have the psychological life of a murderer.

Second, imagine what would happen if we unveiled to Duplicate-Plum that he isn't Plum. It seems that the discovery that it wasn't him, but someone else who committed Black's murder (as well as anything else Duplicate-Plum thinks he remembers) would enormously impact Duplicate-Plum's thoughts about these things. Of course, this discovery would likely disturb Duplicate-Plum, given that he'll discover that most of the important things he thought about his past are false. More importantly, though, it also seems that Duplicate-Plum would disavow any ownership of Plum's actions. Of course, given Duplicate-Plum's psychology, it will often be hard for him to disassociate himself from Plum's life. This, however, doesn't change the fact that from the moment of discovery on Duplicate-Plum will, at least in his reflective moments, think of Plum's actions not as *his* actions and as something people cannot legitimately hold him accountable for. This means that he won't think of them as attributable to himself.<sup>7</sup>

Of course, some things *are* wrong with Duplicate-Plum. For example, he takes joy in the (mistaken) thought that he killed a human being. So, maybe he requires psychotherapy or needs to be isolated from the general population. None of this,

---

<sup>7</sup> Interestingly, support for this claim comes from various science fiction sources. For example, in Stanislaw Lem's (1970) novel *Solaris* protagonist Kris Kelvin meets a "visitor" who is a mental duplicate of his deceased wife Rhexya. Upon discovering that she is a duplicate of Rhexya, this visitor ceases to believe herself to be Rhexya. Rather than continuing to think of Rhexya's actions as her *own*, she becomes disassociated from them to the point that she attempts to kill herself. To use another example, in the movie *Blade Runner* (Scott, 1982), the character Rachael discovers that she is in fact a replicant who has been turned into a mental duplicate of Eldon Tyrell's niece. Upon discovering this fact, Rachael also, visibly, becomes disassociated from the content of her apparent memories, no longer thinking of these actions as her own. These kinds of stories and how they play out lend support to the intuitive plausibility of my claims about Duplicate-Plum's situation and how personal identity matters to attributability. I thank a referee for *Erkenntnis* for drawing my attention to this way in which Lem's story support my claims and making me think further how these kinds of themes figure and are played out in fiction.

however, establishes that Duplicate-Plum *owns* Plum's actions. The reasons we have for doing these things aren't of the *right* kind to make it fitting to hold Duplicate-Plum *accountable*. Again, the same reasons apply to anyone who *delusionally* thinks she is Plum, for whom there is no reason to think Plum's deeds attributable to her. Duplicate-Plum, though, is in the exact same situation, so Plum's deed shouldn't be attributable to him either. Hence, I find neither Shoemaker's intuition, nor the support he offers for that intuition compelling. In fact, if we think carefully about Duplicate-Plum's situation, the case seems to me to offer compelling support for the opposite intuition. Hence, the branch-line case is no counter-example to ENTAILMENT.

What about the fission case? Here Plum physically splits in the middle and the two halves – Left-Plum and Right-Plum – immediately re-grow their missing part. According to Shoemaker, it is implausible that *both* are Plum and so the only plausible conclusion is that neither of them is. Hence, Plum wouldn't survive fission. Nevertheless, so Shoemaker argues, Plum's deeds are attributable to both Left-Plum and Right-Plum. Hence, attributability doesn't presuppose personal identity. Is this convincing?

Assume we concede that neither Left-Plum nor Right-Plum is Plum. We should then reject that Plum's deeds are attributable to Left-Plum and Right-Plum for the reasons mentioned in Duplicate-Plum's case. If neither Left-Plum nor Right-Plum is Plum, they are in the same situation as Duplicate-Plum: each *falsely* thinks that he is Plum, which explains their appropriating stances towards Plum's actions, as long as they remain ignorant of this thought's falsehood. If this fact were revealed to them, though, their reaction would be like Duplicate-Plum's: from the moment of discovery on Left-Plum and Right-Plum will stop thinking of Plum's actions as *their* actions and as something people cannot legitimately hold them to account for.

Construing the case parallel to Duplicate-Plum's case seems *prima facie* plausible, if we stipulate that neither is Plum. There is a crucial difference between the fission case and the branch-line case, however, that might make us more inclined to think of Plum's deeds as attributable to them. Both Left-Plum and Right-Plum plausibly *would be* Plum, if there weren't two of them. Suppose Plum had the ability to shed and regrow one half of his body. Here, it seems plausible that the resulting person would be Plum. But, the only difference between this case and the fission case is that both halves survive in the latter. Hence, everything required for Plum to persist is there in the fission case, except that it is doubled. Why, though, should the *doubling* matter? And, if Plum's deeds are still attributable to the person after the shedding of one half, why shouldn't Plum's deeds be attributable to *both* persons after fission, even though they aren't Plum?

The quick answer is, of course, that neither of them is Plum. However, there's also a different way to respond to these worries. There are four possibilities regarding what happens with Plum: First, he is neither of the resulting people. Second, Plum is Left-Plum, but not Right-Plum. Third, Plum is Right-Plum, but not Left-Plum. Fourth, Plum is *both* Right-Plum *and* Left-Plum. What I want to argue now is that independently of the option we take, our intuitions about attributability go along with it and that non-reductionism captures this. I've already dealt with option one, so let's turn to the other options.

Of course, Shoemaker's take on these options is that only the first is plausible. He reasons as follow:<sup>8</sup> First, Plum can't be either Left-Plum or Right-Plum, because there is no relevant difference between them: whatever relation holds between Plum and Left-Plum *also* holds between Plum and Right-Plum. Hence, on no view about personal identity can it be plausible that Plum is Left-Plum, but not Right-Plum or

---

<sup>8</sup> D. Shoemaker, 2012b: 118. The reasoning follows Parfit's (1984: 255-259) discussion of fission.

*vice versa*. This rules out options two and three. But, numerical identity is one-one and not one-many, so Plum cannot be *both* Left-Plum and Right-Plum. This rules out option four. Hence, only option one is feasible. Combined with the reasoning above about it being implausible that the doubling matters for attributability we arrive at the conclusion that fission undermines ENTAILMENT.

However, this reasoning is faulty. First, the assumption that there is no a relevant difference between Left-Plum and Right-Plum holds only for reductionist views. After all, even if Left-Plum and Right-Plum are both psychologically and physiologically similar there could still be the relevant difference between them that either Left-Plum's or Right-Plum's conscious experiences are Plum's. Hence, a non-reductionist view *could* capture how it could be that Plum is Left-Plum, but not Right-Plum or *vice versa*. The inference from Left-Plum's and Right-Plum's similarity to the claim that either none or both of them is Plum isn't justified.

Note furthermore what would happen if someone with privileged access to these indexical facts were to tell Left-Plum and Right-Plum that Left-Plum, but not Right-Plum is Plum. Here Right-Plum's situation would become exactly like the situation of Duplicate-Plum, while Left-Plum's situation would be like that of Plum. This suggests, again, that attributability follows personal identity.

So, the first step of Shoemaker's reasoning can be questioned if we accept a non-reductionist view and if we do so, we see that our intuitions about attributability will go hand in hand with our intuitions about personal identity. However, this isn't the end of the line, because the second step of Shoemaker's reasoning can also be questioned.

Of course, numerical identity is one-one. However, this is no reason to think that Plum couldn't be *both* Left-Plum *and* Right-Plum. Here is why. Parfit (1984: 246-248) presents the following case: He is trying to answer a physics question in ten

minutes. He realizes that there are two viable paths to an answer, but is unsure which is more likely to succeed. Fortunately, in this thought-experiment, Parfit can divide and remerge his consciousness at will. So, he divides his consciousness, letting each half tackle one of the ways to answer the question. After ten minutes, consciousness is reunited. Parfit points out that this case is best described as one in which one person has two streams of consciousness that are hers, not as one where someone is temporarily replaced by two different people.

Imagine now that Plum is like Parfit in that case, except that he can divide and merge his consciousness *and* body at will. He faces two tasks each demanding his undivided physical and mental attention. He has ten minutes to complete both tasks, but realizes that he can only finish one of them in that time. So, he decides to fission and merge once the tasks are completed. Here Plum divides both his consciousness, as well as his body for a time. Again, though, it seems that this case is most plausibly described as one in which a person temporarily has a divided consciousness in a separated body, *not* as one person being replaced temporarily by two different people. *Both* fission products would be Plum. Of course, this description is a little misleading because it suggests that two persons are identical with one. This isn't how the case should be understood, however. These aren't three different people, but just one with, temporarily, two disassociated streams of consciousness in two different bodies. The mistaken reading is suggested by the assumption that different streams of consciousness are sufficient for different persons. However, it isn't the existence of different streams of consciousness that results in there being different people, but these streams *belonging* to different people. This can be easily cashed out, for example, with Madell's non-reductionism.

If this is an acceptable description of this case, though, it opens the possibility that even in the original fission case the relation between Plum and the fission

products needn't be one-many. After all, we can now say that what we have here is just one person with two permanently disassociated streams of consciousness in two different bodies. Again, this can easily be cashed out with Madell's non-reductionism. Note, furthermore, that if we endorse this view about fission, there is no difficulty in holding on to Shoemaker's intuition that Plum's deeds are attributable to *both* fission products, while holding onto ENTAILMENT. Hence, a non-reductionist view has the resources to preserve ENTAILMENT, while accepting Shoemaker's intuitions about attributability. It seems, consequently, that fission isn't universally troubling for all views about personal identity when it comes to preserving ENTAILMENT, but only for all *reductionist* views.

Of course, at this point one might want to respond that there is a difference between the fission case and the case I described above. In fission, the halves are not just *permanently* separated. Plum *cannot* reunite his halves. And this, so the response goes, means that the fission case cannot be described as a case of one person with two permanently disassociated streams of consciousness in two different bodies.<sup>9</sup> There are a couple of things to be said about this response.

To start, we must be clear how to understand the response. If the response is that it is *impossible* to describe the case in this way, then this is false. If we accept, for example, Madell's non-reductionism, it is possible to describe the case in this way (see Madell's (1981: 127-134 and 2015: 114-115) own discussion of the fission case, in which he argues at length that this description is a distinct possibility). On Madell's view, what matters is to whom streams of consciousness and bodies *belong* and it does not seem impossible for disassociated streams of consciousness and separate bodies belonging to the same person, even if they cannot be reunited.

Of course, the response might also be that it is intuitively not plausible to describe the fission case that way. To this, I have the following three replies. First, it

---

<sup>9</sup> I thank an anonymous referee for *Erkenntnis* for pushing me on this point.

should be clear that the fission case is a very unusual situation and one where it is unlikely that we will have definitive and clear intuitions about what description fits the case best. Hence, it is not likely that there will be conclusive evidence for or against one particular verdict based only on our intuitions. In the end, it will likely have to be more theoretical considerations that should push us towards one or the other description for this particular case. But, the fact that the non-reductionist description preserves ENTAILMENT (if we believe that Plum's deeds are attributable to both halves) is one such consideration and, hence, supports that description.

Second, it is not clear whether the fact that Plum *cannot* reunite his halves makes a decisive difference to the plausibility of this description. It should first be noted, that the fact that the division is *permanent* does not seem to matter. In the modified fission case, we can imagine for any period of time that Plum decides to fission for exactly that amount of time. Maybe Plum wants to travel around the world, but also complete a book. So, he decides to fission for several years and merge again once these tasks are completed. It still seems plausible to think of this case as one where Plum has disassociated streams of conscious in a separated body for several years. But, even if the halves *never* reunite, this then seems a plausible description. However, for any case in which the halves *cannot* reunite, we can imagine a parallel case in which the halves could reunite, but never do. If we think that the description of one person with two permanently disassociated streams of consciousness in two different bodies is apt in the latter kinds of cases, it seems then hard to see, though, why it should not be apt in the former cases. More importantly, there is some evidence that, intuitively, the fact that the halves cannot reunite does not matter for the issue whether or not they are halves of the same person. Parfit's original inspiration for his divided mind case, were cases in which people had formed two separate spheres of consciousness as the result of the connections between their

brain hemispheres being cut. However, even though these two disassociated streams of consciousness *could not* ever be reunited, Parfit (1984: 245/246) naturally described these cases as cases of *one* person with two disassociated streams of consciousness. And, this description does not seem outlandish. If that is the case, though, thinking of the fission case along the lines of one person with two permanently disassociated streams of consciousness in two different bodies should be a feasible description of the case, even if the halves cannot reunite.

Third, even if the description was not intuitively plausible (and, as I said, this is far from clear), this, actually, does not matter for the reply to Shoemaker's argument. After all it is a possible description of the case that doesn't violate the principle that identity is one-one and *can* preserve the judgement that Plum's deeds are attributable to both of Plum's halves, while holding on to ENTAILMENT. Contrary to what Shoemaker claims it is, hence, possible, *for the non-reductionist*, to hold on to the verdict that Plum's deed are attributable to both *and* to ENTAILMENT, without violating that identity is one-one. Note also that if the description is ultimately not plausible, one of the other options for the fission case will be the one we should prefer. But then, as my discussion further above suggests, it also does no longer stay intuitively plausible that Plum's deeds are attributable to both halves and ENTAILMENT is, again, not threatened.

It seems, consequently, that the non-reductionist view can deal with the problems for ENTAILMENT. And, not only this: I take my discussion to show that a plausible diagnosis of the way how our intuitions about *both* personal identity *and* attributability diverge from reductionism while at same time falling in line with non-reductionism to be that *reductionism* just misses something crucial to our concept PERSONAL IDENTITY. This suggest that it isn't ENTAILMENT we should reject, but that we need to accept that PERSONAL IDENTITY is a non-reductionist concept that

plays a special role in determining ownership of actions – a role impersonal concepts can't play. This, of course, suggests Shoemaker's results are unsurprising if we restrict our attention to reductionist views about personal identity, because such views *per definitionem* miss a central aspect of the concept.

### 3. The Non-Representationalist Alternative

Of course, there might be good reasons why Shoemaker doesn't even consider non-reductionism in his discussion of ENTAILMENT. And, indeed there are such reasons, because non-reductionism is generally a neglected view in the contemporary debate that few people accept.<sup>10</sup> This is so for good reasons, because non-reductionism itself has serious problems.

#### 3.1. Troubles for Non-Reductionism

One major worry about non-reductionism, of course, is that it is questionable whether and to what extent the view fits with the picture of the world as provided by science. Even bracketing such concerns, though, there are other concerns, one of which is particularly worth highlighting in the context of this paper.<sup>11</sup>

On non-reductionist views, the personal identity facts are modally distinct from the impersonal facts, including facts about biological or psychological continuity. But, we rely on these latter facts as *evidence* for judgements about personal identity. Non-reductionist views, however, cannot explain why these facts would provide such evidence. On non-reductionist views, infinitely many different people might, for all we know, occupy the psychological and physiological space we

---

<sup>10</sup> With the notable exception of those mentioned further above. It is telling that in the *Stanford Encyclopedia* entry for personal identity non-reductionism doesn't even have its own section (Olson, 2015).

<sup>11</sup> This line was famously pressed by John Locke and Immanuel Kant. In the contemporary debate, the argument has been pressed e.g. by Parfit (1984: 223/224).

associate with one person over time. Hence, if non-reductionism were true, what we normally use as evidence for personal identity cannot play this role. But, it doesn't seem as if anything else could play this role. Consequently, non-reductionism rules out knowledge of personal identity.

Of course, this doesn't mean that non-reductionism *couldn't* be true. It also doesn't speak against ENTAILMENT or non-reductionism preserving ENTAILMENT. It means, however, that non-reductionism implies that personal identity cannot play the practical role we assign it: the practical concerns associated with PERSONAL IDENTITY and the crucial role it plays in our practice are defective, because they presuppose something that couldn't play the role assigned to it. Hence, even if non-reductionism rescues the conceptual connection between MORAL RESPONSIBILITY and PERSONAL IDENTITY, this victory is hollow. All we've shown is that there is a conceptual connection that shows this conceptual framework and the practice based on it to be defective, which undermines the legitimacy of ascriptions of personal identity *and* of moral responsibility.

Of course, it is plausible to take these implications as evidence *against* non-reductionism. After all, non-reductionism offers an interpretation of our ordinary practice governed by the concept PERSONAL IDENTITY. However, other things being equal, an interpretation of that practice is preferable that leaves it legitimate to one that entails it is systematically and radically mistaken. Given that non-reductionism *would* have this consequence we, consequently, shouldn't accept it unless for good *independent* reasons.

However, if Shoemaker's arguments against ENTAILMENT are correct, reductionism about personal identity cannot preserve the legitimacy of certain concerns regarding personal identity either. Following the proposed methodological principle, this means that a view that *severs* the connection between MORAL

RESPONSIBILITY and PERSONAL IDENTITY provides the most plausible interpretation of our practice, because it leaves the legitimacy of ascriptions of identity and of ascriptions of personal identity intact. And in this way, we've abandoned ENTAILMENT after all. While this result seems straightforward, however, it is far from being so. After all denying ENTAILMENT *also* has a cost: It forces us to give up a very intuitive conceptual connection. So, *all* of the options are problematic and we are at most choosing the lesser evil. This raises the question, whether there isn't a better way to resolve this conflict. I think there is.

### 3.2. Non-Representationalism About Personal Identity

Both reductionism and non-reductionisms are instances of a view about how we account for the nature of the concept PERSONAL IDENTITY. This is *representationalism* about personal identity, the view that we account for the nature of that concept in terms of what it *represents*, i.e. what personal identity *consists in*. On this view, questions about metaphysics have priority when it comes to accounting for the distinctive nature of talk and thought about personal identity.

Representationalism is a very prominent methodological approach not only for personal identity, but for most domains of philosophy. This approach, however, can be questioned. A prominent example for an approach that does so is meta-ethical expressivism, i.e. *expressivism* about *normative* thought and discourse (prominent expressivists are Simon Blackburn (1998) and Allan Gibbard (2003)). According to expressivists, normative thought and discourse don't represent some part of reality, but play a different non-representational, action-guiding role.<sup>12</sup> Views like

---

<sup>12</sup> Note that we need to be careful to distinguish two ways in which a state might be called "representational." First, in the sense that it has propositional content. In this sense, a state is representational in that it has *representational* content. Second, in the sense that it has a certain sort of direction of fit, namely a mind-to-world direction of fit. In this sense, a state is representational in that it *represents the world as being a certain way*. The non-representationalist in the sense to be understood in this paper, does not deny that

expressivism, which question the representationalist account, suggest that to account for the relevant concept we should, instead, invoke some non-representational function played by those concepts. I will call such views forms of “non-representationalism.”

A common motivation for non-representationalism lies in a dilemma representationalism faces. Representationalists about any concept can be reductionists or non-reductionists. The problem, however, is that in many philosophically interesting domains reductionists seem unable to account for distinctive characteristics of the concept in question. For example, a common complaint about reductionism about normative thought and discourse is that it cannot account for the familiar “openness” of normative questions vis-à-vis non-normative issues hinted at by the Open Question Argument. This is the first horn of the dilemma, which suggests that the relevant concept is a *non-reductionist* one. The problem with non-reductionist approaches, however, is that they come with dubious metaphysics and epistemology. A familiar charge against meta-ethical non-reductionism, for example, is that the view cannot explain how we get access to the irreducible normative facts. It can only postulate a faculty – “intuition” – that grants us such access, which seems to highlight the problem more than solve it. Such metaphysical and epistemological worries suggest that if our concepts are to be understood as non-reductionism suggests, we have to see these concepts and our practice involving them as defective, because there is good reason to think that the things represented by those concepts don’t exist. Non-reductionism, consequently, threatens to lead to an error theory.

Non-representationalists suggest that we should evade this dilemma, by rejecting representationalism for the problematic concept and, instead, account for

---

judgements of the relevant sort are representational in the first sense, only that they are representational in the second sense.

the concept in terms a non-representational function. This opens up space for an approach that can respect the non-reductionist nature of the concept without non-reductionism's problems. Of course, such an account would still have to be fleshed out. At the very least, however, non-representationalism offers a way to preserve the legitimacy of our practice, to account for the non-reductionist nature of the concepts in question, while avoiding metaphysical and epistemological conundrums. These are significant theoretical virtues.

It seems to me that there is a strong resemblance between the dialectic surrounding ENTAILMENT and this dilemma that motivates non-representationalism. On the one hand, reductionists about personal identity have to deny the plausible connection articulated by ENTAILMENT. Hence, reductionists miss something distinctive about PERSONAL IDENTITY. This suggests that PERSONAL IDENTITY is a non-reductionist concept. But, while non-reductionism preserves ENTAILMENT, it faces epistemological (and probably also metaphysical) problems that undermine the legitimacy of the practice invoking PERSONAL IDENTITY. Given this dialectic, a natural move is to explore non-representationalism about PERSONAL IDENTITY.

How do we develop such an account? In the context of this paper I can do so only in a restricted manner. Recall that the *general* issue we are concerned with here, reformulated in a more theory-neutral manner, is whether judgements about *personal identity* play a crucial role in settling questions about accountability. While a full account of PERSONAL IDENTITY would probably have to capture a wider range of roles (e.g. the governance of patterns of self-concern), here I will, consequently, only suggest an account that focuses on that role.

For some  $\phi$  (be it an action, mental state, or mental event) to be properly attributable to someone is for that individual to stand in a relation to  $\phi$  that makes her the *owner* or *author* of that action in the sense that we can properly say the action

was *hers*. So, when I think of some being A as the *owner* of some  $\phi$ , I think of A as having some kind of status vis-à-vis  $\phi$ . Specifically, my thinking so makes A the target of a range of attitudes and assessments of mine: It makes me disposed to hold *that* being accountable for  $\phi$ . I won't here take a stand on what exactly "holding accountable" involves, though I am going to assume that *whatever it is*, it's a stance crucially characterized by certain attitudinal and behavioral patterns formed in response to certain further judgements about  $\phi$  and A. Hence, to think that  $\phi$  is attributable to some being, just is to be disposed to hold it accountable for  $\phi$ , depending on certain of one's further judgements about  $\phi$  and the being.

Following this, my suggestion is to think of PERSONAL IDENTITY as playing a central role in *making judgements about certain questions about attributability over time*. Specifically, PERSONAL IDENTITY functions to allow us to make judgements about which being is the *inheritor* of the attributability of a particular set of actions, mental states, and mental events, i.e. who to be disposed to hold to account for that set of actions, mental states, and mental events. The crucial question is how to cash this out in non-representational terms. My suggestion is to use an expressivist framework to do so, because expressivism's theoretical commitments and problems are already relatively well-known, and, hence, provide a useful and well understood starting point to develop a non-representationalist view in unfamiliar territory. Before I present the account, though, I want to emphasize that as an *expressivist* view, the account will only be an account about what it is to make *judgements* about personal identity and not about what personal identity *is*. The account shouldn't, for example, be conflated with a form of *conventionalism* about personal identity, according to which the conditions for the persistence of persons are determined by certain practical attitudes. With these remarks out of the way, let me sketch how an account of PERSONAL IDENTITY might be developed in an expressivist framework. The view that I will use

is Allan Gibbard's (2003).

A central notion to Gibbard's expressivism is the notion of a *plan*. In the ordinary sense, *plans* are, roughly, mental states that govern and organize one's intentions over time. So, for example, if I plan to go shopping tomorrow, this plan will organize my intentions for that day, by e.g. preventing me to intend to do things that will make my going shopping impossible, by making me form intentions that are means to my going shopping (e.g. figuring out where the closest supermarket is, etc.), etc. Plans are, thereby, dispositional states that can be helpfully characterized by a function from circumstances to intentions ruled out.

It isn't plans in the ordinary sense that play a center role in Gibbard's account, however, but contingency "plans" in a broader sense. To understand what this means it is important to make the following observation: what is *distinctive* about plans is that they structure, control, govern, and organize intentions in distinctive ways. In fact, it is because of this *regulating role* that ordinary plans are non-representational states, *not* because they regulate intentions specifically: Ordinary plans aren't non-representational because they regulate intentions, but because their functional role doesn't primarily involve representing the world as being a certain way, but rather of regulating our mental economy in response to certain beliefs.

What Gibbard points out is that we needn't assume that only plans in the *ordinary* sense have this feature. There could be mental states that play such a role, but that play this role not (just) for intentions but (also) for other mental states, such as, for example, beliefs or emotional states. Plans in the ordinary sense just provide a *model* for thinking about mental states that have a particular feature, namely to regulate the formation of other mental states in a particular way. In particular, they provide us with a way to make legitimate the idea that states can play a non-representational role, even if that role has no particular connection to action or

intention. Gibbard (a little confusingly and misleadingly) uses the term “plan” for this more general class of mental states of which plans in the ordinary sense are a member. For convenience, I will follow him in this use in what follows.

Coming back to accountability, attributability, and personal identity, how do we cash out my above suggestions in terms of plans? Let’s start with judgements about accountability. Such judgements are *plans* to hold some being to account for some particular  $\phi$ . This will be a plan that structures, controls, governs, and organizes sets of intentions, but also sets of emotional states, such as the reactive attitudes, depending on one’s favored account of what it is to hold someone to account. Of course, this is only a rough sketch, but giving a more substantial account would be beyond the purposes of this paper. It should suffice for our purposes, though.

Judgements about attributability, on the other hand, are plans that play a structuring, controlling, governing, and organizing role regarding the plans that constitute judgements about accountability. Specifically, to think that  $\phi$  is attributable to some being, just is to plan to hold *it* accountable for  $\phi$ , depending on one’s further judgements about  $\phi$  and the being, such as e.g. whether one judges that the being had sufficient control over and knowledge of  $\phi$ . On the other hand, thinking that  $\phi$  *isn’t* attributable to some being, is a plan that blocks one from having any kind of disposition of this kind.<sup>13</sup>

Lastly, we turn to PERSONAL IDENTITY. Take a being A that one thinks of as a person at time  $t_1$  and a being B’s that one thinks of as a person at time  $t_2$ . To think of

---

<sup>13</sup> Note that while judgements about accountability and attributability are related on this view, they are also importantly distinct. For example, it is possible, e.g. for a hard determinist, to judge  $\phi$  attributable to a person, while judging that the person isn’t accountable for  $\phi$ . This is just a case where one doesn’t plan to hold the person accountable, because one doesn’t judge her to have control, but also plans to hold her accountable, if one comes to judge that she has control, relevant knowledge, etc. In this way, the account captures how attributability is prior to accountability: you can’t judge someone accountable for an action without judging that action to be attributable to her, but you can judge the other way around.

A and B as the *same* person is to settle that those actions, mental states, and mental events attributable to A are attributable to B. Specifically, to think of A and B as the *same* person is to think of B as the *inheritor* of the attributability for a particular set of actions, mental states, and mental events. This judgement is the plan that disposes one to judge that the specific set of actions, mental states, and mental events that would be attributable to A is a proper subset of the specific set of actions, mental states, and mental events attributable to B. This plan, in a sense, is the judgement who inherits – in terms of attributability – the specific set of actions, mental states, and mental events that would be attributable to A (or in the other direction, what set B is the inheritor of).

This puts a non-representational account of PERSONAL IDENTITY on the table. Of course, these are just the bare bones of the beginnings of such an account. For this paper's purposes, however, it should suffice. Recall that Shoemaker has argued that ENTAILMENT needs to be abandoned, because no plausible account of personal identity can capture it. I've argued that this holds only for *reductionist* accounts and that the problems such accounts face actually *support* ENTAILMENT rather than undermine it, because in all relevant cases our intuitions about personal identity go hand in hand with our intuitions about attributability. Furthermore, I've argued that a non-reductionist account, *can* capture those intuitions and, hence, the conceptual connection given by ENTAILMENT. This seems to suggest that ENTAILMENT *does* capture a conceptual connection between personal identity and attributability *and* that personal identity is a non-reductionist concept. The problem with non-reductionism, however, is that it comes with epistemological worries that threaten to undermine the legitimacy of the practice governed by PERSONAL IDENTITY, and consequently, our practice involving attributions of responsibility.

However, the non-representationalist account I've suggested also captures this

connection: On this account, the very role of PERSONAL IDENTITY is to allow us to make judgements about attributability. Hence, this account is on an equal footing to non-reductionism, regarding the phenomena that trouble reductionists. As a non-representationalist account – and a version of expressivism in particular, which is often explicitly motivated on the basis that it does not face the epistemological worries of a non-reductionist account (see e.g. Blackburn, 1998: 86, Lenman, 2007, Ridge, 2014: 70) –, though, it doesn't incur the epistemological worries of non-reductionism and, so, doesn't threaten the legitimacy of our practice governed by PERSONAL IDENTITY. Let me explain, briefly and roughly, why this is so.<sup>14</sup>

On a non-representationalist account of PERSONAL IDENTITY, questions about personal identity are questions of a different kind than they would be on a representationalist account, such as e.g. non-reductionism. For example, on the account I've offered, questions about personal identity are *planning* questions, i.e. questions about whom to hold account for what. For the non-representationalist, whether personal identity holds is, hence, a distinct question from questions about e.g. biological or psychological continuity, just as it would be on a non-reductionist account. Indeed, because the former is a *planning* question, judgements about biological or psychological continuity do not, by themselves, conclusively settle questions about personal identity, as those judgements do not conclusively settle, by themselves, the planning question whom to hold account for what.

The reason why these questions are distinct, though, is quite different from the reason offered by a non-reductionist and not one that would give much credibility to skepticism about questions concerning personal identity or make it mysterious why e.g. facts about biological or psychological continuity serve as *evidence* for judgements about personal identity. Indeed, it is pretty clear that judgements about e.g. biological or psychological continuity will play a crucial role in settling questions about personal

---

<sup>14</sup> I thank an anonymous referee for *Erkenntnis* for asking me to elaborate.

identity, if those are planning questions: they offer the kinds of considerations that we will have to rely on in settling whom to hold to account for what. Hence, the non-representationalist can explain why these kinds of facts serve as *evidence* for judgements about personal identity, while upholding that, ultimately, whether personal identity holds is distinct from these matters. Of course, let me also highlight that none of this means that a non-representationalist about personal identity will deny that questions about personal identity have objectively true answers, hold that anything goes by way of forming judgements about personal identity or think that our own plans actually conclusively settle questions about personal identity, rather than being our beliefs on these questions. Instead, the sensible non-representationalist will be a quasi-realist who upholds the view that judgements about personal identity can be objectively true, that they can constitute knowledge and so on, by offering a non-representationalist friendly treatment of these notions (see e.g. Blackburn, 1998, Gibbard, 2003, Ridge, 2014, Sinclair, 2006 amongst others). However, as has been suggested by (or on behalf of) quasi-realists, adding these sorts of features to a discourse does not, actually, expand the epistemological burdens of the account.<sup>15</sup>

Non-representationalism about PERSONAL IDENTITY doesn't, hence, incur the epistemological worries of non-reductionism. This doesn't just recommend this view for further investigation, but also shows that the connection postulated by ENTAILMENT can be preserved without threatening the legitimacy and practical relevance of questions about personal identity.

## Bibliography

---

<sup>15</sup> Of course, in meta-ethics, there is a debate about the question whether a successful meta-ethical quasi-realist account faces the problems that is faced by non-reductionist account. This challenge has prominently been raised by Sharon Street (2011). For convincing replies to this worry, see Dreier, 2012 and Golub, 2017.

- Blackburn, S. (1998). *Ruling Passions*. Oxford: Oxford University Press.
- DeGrazia, D. (2005). *Human Identity and Bioethics*. Cambridge: Cambridge University Press.
- Dreier, J. (2012). Quasi-Realism and the Problem of Unexplained Coincidence. *Analytic Philosophy*, 53(3), 269–287. <https://doi.org/10.1111/j.2153-960X.2012.00567.x>
- Gibbard, A. (2003). *Thinking How to Live*. Cambridge MA: Harvard University Press.
- Golub, C. (2017). Expressivism and the Reliability Challenge. *Ethical Theory and Moral Practice*, 20(4), 797–811. <https://doi.org/10.1007/s10677-017-9794-1>
- Lem, S. (1970). *Solaris*. New York: Walker.
- Lenman, J. (2007). Expressivism and Epistemology: What is Moral Inquiry? *Aristotelian Society Supplementary Volume*, 81, 63–81.
- Lewis, D. (1983). Survival and Identity. *The Identities of Persons*, 1, 55–78. <https://doi.org/10.1093/0195032047.001.0001>
- Locke, J. (1694). *An Essay Concerning Human Understanding*. (P. Nidditch, Ed.). Oxford: Clarendon Press.
- Lowe, E. J. (2012). The Probable Simplicity of Personal Identity. In G. Gasser & M. Stefan (Eds.), *Personal Identity: Simple or Complex?* Cambridge: Cambridge University Press.
- Madell, G. (1981). *The Identity of the Self*. Edinburgh: Edinburgh University Press.
- Madell, G. (2015). *The Essence of the Self. In Defense of the Simple View of Personal Identity*. New York: Routledge.
- Merricks, T. (1999). Endurance, Psychological Continuity, and the Importance of Personal Identity. *Philosophy and Phenomenological Research*, 49(4)(4), 983–997.
- Olson, E. (1997). *The human animal: Personal identity without psychology*. *Philosophy of mind series*. <https://doi.org/10.2307/2653504>

- Olson, E. (2015). Personal Identity. In R. Martin & J. Barresi (Eds.), *Stanford Encyclopedia of Philosophy*. Oxford: Stanford University. Retrieved from <https://plato.stanford.edu/entries/identity-personal/>
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Ridge, M. (2014). *Impassioned Belief*. Oxford: Oxford University Press.
- Scanlon, T. M. (2008). *Moral Dimensions. Permissibility, Meaning, Blame*. Cambridge MA: Harvard University Press.
- Scott, R. (1982). *Blade Runner*. United States of America: Warner Bros.
- Shoemaker, D. (2012). Responsibility Without Identity. *The Harvard Review of Philosophy*, 18, 109–132.
- Shoemaker, S. (1984). Personal Identity: A Materialist's Account. In R. Swinburne & S. Shoemaker (Eds.), *Personal Identity*. Oxford: Blackwell.
- Sinclair, N. (2006). The Moral Belief Problem. *Ratio*, 19(June), 249–260.  
<https://doi.org/10.1111/j.1467-9329.2006.00323.x>
- Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the Aristotelian Society*, 48, 1–25.
- Street, S. (2011). Mind-Independence without the Mystery: Why Quasi-Realists Can't Have it Both Ways. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics. Vol. 6* (pp. 1–32). Oxford: Oxford University Press.
- Swinburne, R. (1984). Personal Identity: The Dualist Theory. In R. Swinburne & S. Shoemaker (Eds.), *Personal Identity* (pp. 1–66). Oxford: Blackwell.
- Watson, G. (1996). Two Faces of Responsibility. *Philosophical Topics*, 24(2), 227–248.