ONE—BUT NOT THE SAME[1]

John Schwenkler, Nick Byrd, Enoch Lambert, and Matthew Taylor

I

What kinds of changes can a human being undergo without thereby becoming someone different? Someone can lose an arm, for example, and after this change she will continue to exist, with physical characteristics that are different from those she had before. But what if you lose all your memories, or have your entire brain or body replaced, or suddenly acquire a radically different personality? Is the person who exists after changes like these a *different person entirely* than the one who existed before them?

Judgments about these matters are complicated by the fact that phrases like "same person" and "different person" have multiple uses in ordinary English. If, for example, your good friend has just returned from a life-changing semester abroad, you might say of her that

(1) She's not the same person I used to know.

But "same person" means something quite different in (1) than it would mean if, after you encountered John on Tuesday and his identical twin Joe the day after, their mother Alice were to tell you that

(2) The person you saw on Wednesday is not the same person you saw the day before.

[1] We are grateful especially to Josh Knobe, as well as to Randy Clarke, Shaun Nichols, David Rose, Nina Strohminger, and two referees with this journal, for valuable feedback and discussion. JS's research has been supported by an Academic Cross-Training Fellowship from the John F. Templeton Foundation, and compensation for experimental participants was provided by the Tufts University Center for Cognitive Studies. Author contributions were distributed as follows, according to the CRediT taxonomy (http://credit.niso.org): *Conceptualization:* EL, JS, MT; *Data curation:* JS; *Formal analysis:* NB, JS; *Funding acquisition:* EL, JS; *Investigation:* NB, JS; *Methodology:* NB, EL, JS, MT; *Project administration:* JS; *Visualization:* NB, JS; *Writing -- original draft:* JS; *Writing -- review & editing:* NB, EL, JS, MT.

If (1)-type and (2)-type ways of using "same person" can come apart from one another, then we need to get clear on which use is in play in any given case before we can understand what is being talked about.

How can we achieve this clarity? In real-life conversation, contextual cues usually suffice, and thus there is no prospect of misunderstanding what is meant in cases like the ones described just above. But philosophical contexts can make for ambiguity that may be harder to resolve. Suppose a man goes off to war, where he suffers a brain injury that corrupts his memory and changes his personality dramatically. And now let us ask: Is the man who returns home the same person as the man who left? In *a* sense of that phrase, we are quite prepared to say he is not. But what if that is not the sense of that phrase that interests us? What way do we have of asking whether the man returning home is the same person as the man who left, *in the same sense* of "same person" that Alice uses to say, in (2) above, that John is not the same person as his twin?

For those with training in philosophy, the task is fairly easy. In statements like (1), we say in philosophy, the word "same" expresses a concept of *qualitative* identity or difference—and in the context at issue they are often ways of characterizing similarity or difference in *personality* or self-understanding ("identity" in one colloquial sense), which is why we can also use this language to say of two close friends that they are "practically the same person".[2] By contrast, in a statement like (2) the concept expressed by the phrase "same person" is that of *numerical* identity or difference—the thing at stake in philosophical disputes over the rationality of concern for our future selves (Parfit 1984), the relevance of memory to personal continuity (Locke 1975; Reid 2002; Nichols 2017; Swinburne 2019, ch. 3), the possibility of surviving one's death (Baker 2005; Nichols et. al. 2018) or being "transferred" into an entirely different body (Williams 1970), and the proper characterization of transformative decisions (Paul 2014, 2020; Molouki and Bartels 2017; Molouki et al. 2020) and radical changes in personality or moral values (Searle 2005; Strohminger and Nichols 2015; Prinz and Nichols 2016; Earp, Latham, and

[2] We will return at the end of this paper to consider whether qualitative identity or difference is the *only* concept that can be expressed by the phrase "same person" in a statement like (1).

Tobia 2020). It's *this* concept, we say, i.e. that of being *numerically* the same person, that we mean to elicit in our question about the man returned from war.

But it is hard to take this approach in posing philosophical questions to those without prior mastery of this specialized terminology. As anyone who has tried teaching the topic of personal identity to beginning students can likely attest, questions of the form "Is this person *numerically* the same …?" always ring hollow at first, and even after their meaning has been explained a bit of Socratic probing reveals that opposing interpretations of the phrase "same person", frequently accompanied by a conception of "identity" as something that can be lost, acquired, altered, or discovered, are never very far from the surface. Whether the context is that of an examination, a casual conversation, or an attempt to discern commonsense "intuitions" about matters of philosophical interest, in order to know that *they* are thinking about the things *we* mean to be at issue, one needs to ensure that the distinctions being drawn are connected to the appropriate conceptual category.

Here is a way of doing this, inspired by experience in the classroom and explored recently in the work of Vincent Descombes (2016). Outside philosophy, the most common way to express our concept of a person's numerical identity is through the connected use of proper names, definite descriptions, and the personal pronouns "I", "you", "he", and "she". Thus one might say, following Thomas Reid (2002, p. 276), of a certain man that *he* was flogged when a boy at school, for robbing an orchard; that later on *he* took a standard from the enemy in his first campaign; and that *he* was then made a general in advanced life. Perhaps this story does not settle whether the man was the "same person" in advanced life as in boyhood, where that phrase is used in the sense it has in (1) above. It does, however, commit us to thinking that there is *someone* who falls under all the descriptions here given, and that it is this (one, selfsame) person whom the entire story is about. In using a single pronoun to refer to the man under each of our three descriptions, we have thereby described him as ("numerically") identical across them.

As we indicated above, this proposal is relevant to the work of psychologists and philosophers who wish to probe commonsense intuitions about the conditions of personal identity over time (e.g. Nichols and Bruno 2010; Strohminger and Nichols 2014, 2015; Tobia 2015, 2016; Prinz and Nichols 2016; Molouki and Bartels 2017; Weaver and Turri

2018; Earp et al. 2019; Earp, Latham, and Tobia 2020). Such research has frequently identified surprising effects of *normative* considerations on identity judgments: for example, Tobia (2015) found that people were more inclined to say that someone had remained the same person if their moral character improved than if they became morally worse, while Strohminger and Nichols (2014) found that deterioration in a person's moral character had a larger influence on identity judgments than did other kinds of radical psychological changes such as extinction of their memories or desires. This work has been criticized, however, for failing to distinguish judgments of numerical identity from judgments of merely qualitative similarity and difference (Berniūnas and Dranseika 2016; Dranseika 2017; Starmans and Bloom 2018). In what follows we present the results of two pre registered studies[3] that together serve to validate this criticism, while also showing a way to address it.

II

In his (2015) study, Tobia investigated folk concepts of personal identity by presenting participants with one of two variants on the story of Phineas Gage:

> Phineas is extremely [kind / cruel]; he really enjoys [helping / harming] people. He is also employed as a railroad worker. One day at work, a railroad explosion causes a large iron spike to fly out and into his head, and he is immediately taken for emergency surgery. The doctors manage to remove the iron spike and their patient is fortunate to survive. However, in some ways this man after the accident is remarkably different from Phineas before the accident. Phineas before the accident was extremely [kind / cruel] and enjoyed [helping / harming] people, but the man after the accident is now extremely [cruel / kind]; he even enjoys [harming / helping] people. (Tobia 2015, p. 397)

[3] To view the pre-registration visit https://osf.io/pn5q6/?view_only=020591e7f1124eeaa51dfaf9e248bbdf.

Here, the text in brackets marks the difference between two conditions: one of moral IMPROVEMENT, in which Phineas changes from a cruel person who enjoys harming people to a kind person who enjoys helping them; and one of moral DETERIORATION, in which the change is from kindness to cruelty instead. Tobia's interest was in seeing whether these differences made a difference in judgments of the protagonist's diachronic identity. To this end, following the vignette Tobia's participants read a brief paragraph whose purpose was "[t]o engage participants with the relevant notion of *numerical* identity" (ibid.):

> Art and Bart disagree over what happened in this story. Art thinks that Phineas before the accident and the man after the accident are different in some respects but are still the same person. To Art, it seems like one person (Phineas) experienced some changes. Bart disagrees. He thinks that after the accident, the original man named Phineas does not exist anymore; the man after the accident is a different person. To Bart, it seems like one person died (Phineas before the accident), and it is really a different person entirely that exists after the accident (the man after the accident). (Tobia 2015, 397-398)

Participants were then asked to indicate whether they agreed more with Art or with Bart about what had happened. And these judgments were significantly affected by direction of change, as participants agreed more with Bart in the condition of moral DETERIORATION than that of moral IMPROVEMENT.

But did Tobia's clarificatory paragraph have the intended effect? His disputants' repeated talk of "same person" and "different person" in expressing their views provides some reason to doubt this, given the common use of these phrases to express concepts of qualitative identity and the naturalness of considering the dispute in this light. And while the character of Bart uses phrases like "does not exist" and "different person entirely" in an attempt to clarify the stakes, it's possible to read this talk as metaphor or hyperbole. A person reading the paragraph above could very well think that Bart is right in *some* important sense in what he says about what happened in the story, without believing that the man after the accident is a different person than the man before it in the same sense as, in our sentence (2) above, Joe is said to be a different person than his identical twin.

To explore this matter, we presented participants (n = 301: 44% male, mean year of birth = 1996), recruited from the online platform Prolific, with a vignette that was identical to Tobia's save for its opening sentence:

> Phineas grew up in Brooklyn. He is extremely [kind / cruel] ...

Following this vignette, our participants read a clarificatory paragraph identical to Tobia's, summarizing the imaginary debate between Art and Bart. As in Tobia's study, the bracketed text above marks the differences between two conditions, one of moral DETERIORATION and one of moral IMPROVEMENT. Following the vignette and clarificatory paragraph, all participants responded to each of the following prompts:

(A) Please indicate whether you agree more with Art or with Bart about what happened in this story. *[7-point scale ranging from "I agree strongly with Art" to "I agree strongly with Bart".]*

(B) Other than Art and Bart and the doctors, how many people are described in this story? *[Forced-choice between "1", "2", "3", and "More than 3".]*

(C) Please indicate your agreement with the following statements about **the person who exists after the accident**:

  (i) He was born in Brooklyn. *[7-point scale ranging from "Strongly agree" to "Strongly disagree".]*

  (ii) He was originally extremely [kind / cruel], and then was in an accident that made him become extremely [cruel / kind]. *[7-point scale ranging from "Strongly agree" to "Strongly disagree".]*

The last three prompts were intended to engage a concept of numerical identity in the manner outlined in our opening section: the question in (B) did so by inviting participants simply to *count* the number of people described in the vignette, and the statements under (C) did so by inviting them to consider whether there was a basis for attributing Phineas's upbringing, and the entire series of moral characteristics described in the vignette, to the person who existed after the accident. By contrast, prompt (A) replicates the test question

from Tobia's original experiment. We predicted that we would observe an effect of direction of change (IMPROVEMENT vs. DETERIORATION) on responses to (A), but not on responses to (B) and (C).[4]

In this study our three prompts were always displayed in the order shown above, while the order of statements (i) and (ii) under (C) was randomized. Each prompt was shown on a separate page with the vignette and clarificatory paragraph above it, and participants were not able to return to a previous page of the survey once they had completed it. Following an a priori decision, we excluded 33 participants from the final analysis for choosing "3" or "More than 3" as the answer to (B), on the grounds that the only sensible responses to this vignette were to think either that the person after the accident was numerically the same person as Phineas, in which case the answer would be "1", or that he was numerically a different person, in which case the answer would be "2".[5] As a result of this exclusion, prompt (B) was effectively a binary choice question, and so we analyzed it according to the proportion of participants who had answered "1" rather than "2".

The results of the study were in line with our predictions. A two-way analysis of variance revealed a highly significant main effect of direction of change on responses to prompt (A), such that participants agreed more with Bart in the condition of moral DETERIORATION (M = 3.32, SD = 1.87) than in the condition of moral IMPROVEMENT (M = 2.65, SD = 1.54).[6] However, we found no effect of direction of change on responses to (B) and (C), which were essentially identical between our two conditions. Regardless of whether the person who existed after the accident had an improved or worsened moral

[4] We realized only after completing the study that there was an error in our wording of the statement in (Ci): while the vignette begins by saying that Phineas *grew up* in Brooklyn, this statement says instead that he was *born* there. A post hoc analysis revealed that overall agreement with the statement in (Ci), M = 2.21, SD = 1.53, was significantly lower than overall agreement with the statement in (Cii), M = 1.61, SD = 0.94: $t(267) < .001$. However, this difference is immaterial to our argument, since as we explain just below there was no effect of direction of change on agreement with either of the two statements, and this last thing is what our predictions all concerned. A similar point applies to prompt (B) in our Study 2.

[5] Notably, this exclusion did not affect our results: a post hoc analysis including all participants who finished the survey found a significant effect of direction of change on ratings of (A), $F(1,289) = 12.2$, $p < .0001$, but no significant effect of condition on ratings of the other prompts: for (B), $X^2(3,291) = 1.052$, $p = .789$; for (Ci), $F(1,289) = 0.216$, $p = .643$; for (Cii), $F(1,289) = 0.014$, p = .903.

[6] $F(1,266) = 10.31$, $p = .002$, $d = .39$. For comparison, in the corresponding experiment in Tobia (2015), the mean response on an identical 7-point scale (1 = Strongly Agree With Art; 7 = Strongly Agree with Bart) was 3.26 (SD = 1.91) in the IMPROVEMENT condition, and 2.61 (SD = 1.67) in the DETERIORATION condition.

character, participants agreed overwhelmingly that our vignette described a stretch in the life of one person,[7] born in Brooklyn,[8] and that *he* had been in an accident that led to a radical change in *his* moral character.[9] In neither case, then, did our participants view the person after the accident as a numerically different person from the original Phineas.

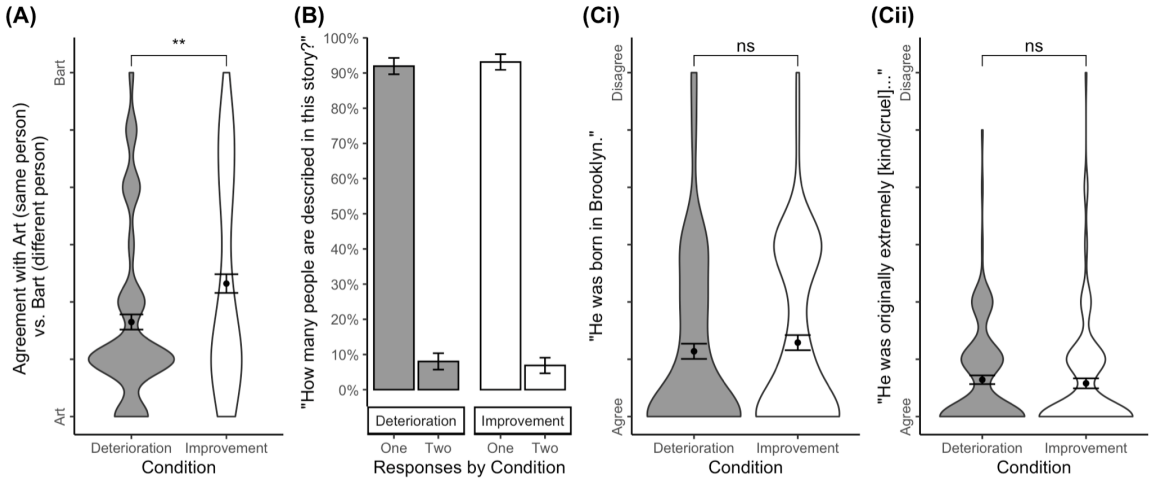

Figure 1. Summary of responses to prompts (A), (B), (Ci), and (Cii) in Experiment 1 with standard error bars (final $n = 268$).

The results of this experiment are visualized in Figure 1. While they replicated Tobia's (2015) finding of an effect of direction of change on agreement with the disputants in his clarificatory paragraph, they also support the hypothesis that this paragraph failed to ensure that his participants employed a concept of purely numerical identity in evaluating the positions described, as no such effect was found in responses to our prompts (B) and (C). Radical changes in moral character, whether for the better or for the worse, did not lead participants to question the *numerical* identity of the person after the accident with Phineas before it, as expressed in responses to the latter prompts. Across both conditions, our participants consistently judged that there was just one person, other than Art and Bart and the doctors, who was described in this vignette, and

[7] Prompt (B): 93.1% of participants answered "1" in the condition of moral IMPROVEMENT and 92.0% answered "1" in the condition of moral DETERIORATION: $X^2(1,268) = 0.016$, $p = .898$.

[8] Prompt (Ci): in the condition of moral IMPROVEMENT, $M = 2.29$, $SD = 1.52$; in the condition of moral DETERIORATION, $M = 2.14$, $SD = 1.55$: $F(1,266) = 0.653$, $p = .42$.

[9] Prompt (Cii): in the condition of moral IMPROVEMENT, $M = 1.58$, $SD = 0.99$; in the condition of moral DETERIORATION, $M = 1.64$, $SD = 0.89$: $F(1,266) = 0.293$, $p = .589$.

they consistently described the person after the accident as having the same birthplace as Phineas, and as having been changed from having one kind of moral character to having another.

III

Our second experiment extended these findings by applying the same methodology within a slightly different paradigm. In Study 1 of their (2014) paper, Strohminger and Nichols presented participants with the following scenario:

> Jim is an accountant living in Chicago. One day, he sustains a severe head injury from a car accident. His only chance for survival is participation in an advanced medical experiment called a Type 2 transplant procedure. It is the year 2049 and scientists are able to grow different parts of the brain if they become damaged. A stock of brain tissue is kept cryogenically frozen to be used as spare parts in the event of an emergency. In a Type 2 transplant procedure, a team of doctors removes the damaged parts of the brain and carefully replaces it with the stock brain tissue. The damaged brain tissue is destroyed after it has been removed. After the operation, all the right neural connections between the old brain and the replacement brain tissue have been made. The doctors test all physiological responses and determine that the patient is alive and functioning. The doctors scan the brain of the transplant recipient and run some standard psychological tests.

Strohminger and Nichols's vignette then concluded with one of several possible descriptions of the condition of the transplant recipient after the surgery. Here we will focus on the following three:

UNCHANGED They [*viz.*, the doctors] discover that the transplant recipient thinks and acts the same way as before the accident.

> MORALITY They [*viz.*, the doctors] discover that the transplant recipient has lost his moral conscience—he is no longer capable of judging right from wrong, or being moved by the suffering of others. Aside from this, he thinks and acts the same way as before the accident.
>
> MEMORY They [*viz.*, the doctors] discover that the transplant recipient has lost his memories—he can no longer remember anything that happened before the accident. Aside from this, he thinks and acts the same way as before the accident.

Following the vignette, participants in this study indicated their agreement with the statement that "After the surgery, the transplant recipient is still Jim". And Strohminger and Nichols found significant effects of condition on agreement with this statement, such that mean agreement with it was significantly lower in the MORALITY condition than the MEMORY condition, and significantly lower in both these conditions than in the condition in which Jim was psychologically UNCHANGED.

Yet like the phrase "same person", the description of someone as being or not being "still so-and-so" does not unambiguously express a concept of numerical personal identity. The question arises, therefore, whether participants who denied that the person after the transplant was ("still") Jim meant the same sort of thing one might mean in saying, for example, that the chair of one's department is not ("still") Professor Marques. To explore this question we presented a second group of participants (n = 183: 45% male, mean year of birth = 1996), recruited from the online platform Prolific, with a vignette that began as follows:

> Jim is an accountant who grew up in Chicago. One day, he sustains a severe head injury ...

The vignette continued in the same way as the original vignette used by Strohminger and Nichols, and concluded in one of three ways:

UNCHANGED They [*viz.*, the doctors] discover that the transplant recipient thinks and acts the same way as Jim did before the accident.

MORALITY They [*viz.*, the doctors] discover that the transplant recipient has no moral conscience—he is not capable of judging right from wrong, or being moved by the suffering of others. Aside from this, he thinks and acts the same way as Jim did before the accident.

MEMORY They [*viz.*, the doctors] discover that the transplant recipient has no memories—he cannot longer remember anything that happened before the accident. Aside from this, he thinks and acts the same way as Jim did before the accident.

The small differences between our MORAL and MEMORY conditions and those used by Strohminger and Nichols served only to counter any implication that the transplant recipient was numerically the same person as the original Jim. While their transplant recipient had *lost* his moral conscience or memories, was *no longer* capable of judging right from wrong or remembering anything that had happened to him, and otherwise was just as *he* had been before the accident (where all these descriptions imply that the recipient had existed beforehand as a person with similar or different characteristics), our patient simply didn't *have* a moral conscience or memories, simply was *not* capable of judging morally or remembering the past, and was otherwise the same as *Jim* had previously been. As such, these vignettes were better suited than the originals to elicit unbiased judgments of the numerical non-identity of Jim and the transplant recipient.

After reading the vignette, the participants in our experiment responded to the following prompts:

(A) Please indicate your agreement with the following statement about **the transplant recipient after the surgery**: He is still Jim. *[7-point scale ranging from "Strongly agree" to "Strongly disagree".]*

(B) Please indicate your agreement with the following statement about **the transplant recipient after the surgery**: He was born in Chicago. *[7-point scale ranging from "Strongly agree" to "Strongly disagree".]*

(C) Other than the doctors, how many people are described in the story you just read? *[Forced-choice between "1", "2", "3", and "More than 3".]*

Each participant viewed all three prompts, with (A) and (B) displayed first in a counterbalanced order, followed by (C). As with Experiment 1, we followed an a priori decision to exclude 12 participants for choosing "3" or "More than 3" as the answer to (C).[10] As a result of this exclusion, prompt (C) was effectively a binary choice question, and so we analyzed it according to the proportion of participants who had chosen "1" rather than "2". We predicted that we would observe an effect of direction of condition (MORAL vs. MEMORY vs. UNCHANGED) in responses to (A), but not in responses to (B) or (C).

The results of this experiment were in line with our predictions. A three-way analysis of variance revealed a highly significant main effect of condition on responses to (A), such that participants agreed more with the statement in (A) in the UNCHANGED condition ($M = 1.93$, $SD = 1.07$) than the condition of erased MEMORY ($M = 2.73$, $SD = 1.56$), and agreed with it least of all in the condition of MORAL impairment ($M = 3.22$, $SD = 1.81$).[11] Planned pairwise analyses revealed that responses to (A) were significantly different between the UNCHANGED condition and the MORAL condition[12] and between the

[10] Once again, this exclusion did not affect our results: a post hoc analysis including all participants who finished the survey found a significant main effect of direction of change on responses to (A), $F(2,180) = 10.45$, $p < .0001$, but no significant effect of condition on responses to the other prompts: for (B), $F(2,180) = 0.811$, p = .446; for (C), $X^2(6,183) = 4.140$, $p = .658$.

[11] $F(2,168) = 10.7$, $p < .001$. For comparison, in the corresponding experiment in Strohminger and Nichols (2014), the mean responses on an identical 7-point scale (1 = Strongly Agree; 7 = Strongly Disagree) were 2.34 ($SD = 1.43$) in the UNCHANGED condition, 3.68 ($SD = 1.72$) in the MEMORY condition, and 4.77 ($SD = 2.03$) in the MORALITY condition. We thank Nina Strohminger for providing us with these statistics.

[12] $F(1,114) = 21.68$, $p < .001$, $d = .86$

UNCHANGED condition and the MEMORY condition,[13] while the difference in responses to (A) between the MORAL condition and the MEMORY condition was statistically nonsignificant[14] but trended toward statistical significance in the same direction as Strohminger and Nichols's original (2014) finding.[15] By contrast, we found no effect of condition on responses to (B) and (C), which were essentially identical between our three conditions. Regardless of the psychological state of the transplant recipient after the surgery, participants agreed overwhelmingly that our vignette described a stretch in the life of one person,[16] born in Chicago.[17] In neither case, then, was there any evidence that our participants viewed the person after the accident as numerically different from the original Jim, nor did these judgments differ between our three conditions.

These results are visualized in Figure 2. While they largely replicated Strohminger and Nichols's (2014, Study 1) finding of the effect of condition on participants' evaluation of whether the transplant recipient after the surgery was "still Jim", they also provide strong evidence that participants were not interpreting this statement as expressing a concept of purely numerical identity, since our participants responded to prompts (B) and (C) in the same way when the transplant recipient was psychologically unchanged as when he had lost his memory or moral conscience. Even when the recipient of the Type II transplant had no memories or moral conscience after the surgery, participants consistently read the vignette as describing a stretch in the life of just one person, born in Chicago and then, in some cases, significantly transformed in the wake of an accident.

---

[13] $F(1,110) = 10.01$, $p = .002$, $d = .60$

[14] $F(1,112) = 2.414$, $p = .123$, $d = .29$

[15] Since our primary interest was not in the pattern of responses to question (A) we were not troubled by this lack of a statistically significant difference, but will pause to emphasize that failure to replicate a previously observed effect with a *p*-value of less than .05 does *not* amount to a non-replication of that earlier finding. While our observed effect size of $d = .29$ for the comparison between the MORAL and MEMORY conditions was smaller than the effect size of $d = .58$ observed by Strohminger and Nichols it was still not negligible, and there is no reason to see one or the other of these as the "true" size of the effect in question. (We thank Nina Strohminger for providing us with this last statistic.) Once there are further replications of these experiments, researchers should examine how many samples' confidence intervals fall outside the range of the population effect size, rather than how many *p*-values are below a given threshold of statistical significance (Cumming 2014).

[16] Statement (C): 86.0% of participants answered "1" in the UNCHANGED condition, 93.2% answered "1" in the MORALITY condition, and 90.9% answered "1" in the MEMORY condition: $X^2(2,171) = 1.77$, $p = .413$.

[17] Statement (B): in the UNCHANGED condition, $M = 1.23$, $SD = 0.80$; in the MORALITY condition, $M = 1.14$, $SD = 0.55$; in the MEMORY condition, $M = 1.22$, $SD = 0.69$: $F(2,168) = 0.282$, $p = .754$.

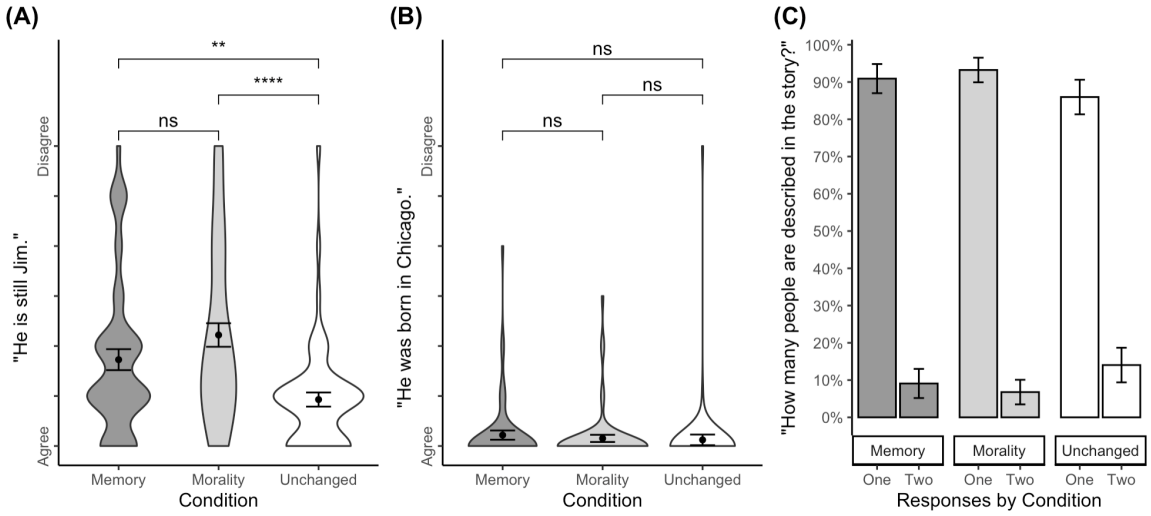

Figure 2. Summary of responses to prompts (A), (B), and (C) in Experiment 2 with standard error bars (final n = 171).

IV

If phrases like "the same person", "the same man", and "still so-and-so" don't always engage a concept of numerical identity, then what other concept or concepts do we use these words to express? To this point we have followed Dranseika (2017; cf. Berniūnas and Dranseika 2016) and Starmans and Bloom (2018) in assuming that it is the concept of *qualitative* psychological similarity or sameness in personality, but there are other views on offer. For example, a large body of work (e.g. Newman, Bloom, and Knobe 2013; Newman, De Freitas, and Knobe 2015; Strohminger, Knobe, and Newman 2017; De Frietas et. al., 2018) has analyzed the concept of a "true self" as a moralized conception of a person's individual essence, while Descombes (2016) links ordinary talk of "who I am" to the concept of *identity* as something that is defined by a person's own self-understanding, and Knobe (in prep.) suggests that talk of who a person is may express a "dual character concept" that can be employed either descriptively or in a normatively laden way, as when we speak of who someone is *truly* or *ultimately*. A yet further possibility is that which of these concepts, and perhaps some other ones too, come

into play is a function of conversational context. Nothing in our findings helps to decide between these different views.

Our findings do, however, suggest a promising strategy for the experimental study of how philosophically important concepts are employed by people without formal philosophical training. As we noted above, in philosophy we use phrases like "numerical identity" and "qualitative identity" in a somewhat artificial way, in order thereby to disambiguate between the different meanings a phrase like "same person" can have in ordinary language. But we cannot easily disambiguate things in *this* way when we wish to investigate how these concepts are understood by non-philosophers: for a question like "Is the man after the accident *numerically the same* as the man before?" cannot be posed to such a person without first explicating the meaning of the italicized phrase.

The clarificatory paragraph in Tobia (2015) was an attempt to provide this sort of explication. As we have shown, however, the paragraph simply did not have the intended effect. Yet our results suggest that the best way to address this deficiency is not by attempting further explication of the terms of an imaginary philosophical debate. What's required instead is to probe participants' judgments with a greater variety of measures, informed by reflective sensitivity to the different ways that philosophically important concepts find expression in ordinary language. We hope that further work will extend this methodology further, by developing diverse measures and rigorously assessing their validity and reliability, then employing these in connection with a wider range of cases in the literature.

REFERENCES

Baker, L. R. 2005. Death and the afterlife. In *The Oxford Handbook of Philosophy of Religion*, ed. W. J. Wainwright. Oxford University Press.

Berniūnas, R., and V. Dranseika. 2016. Folk concepts of person *and* identity: a response to Nichols and Bruno. *Philosophical Psychology* 29: 96-122.

Cumming, G. 2014. The new statistics: why and how. *Psychological Science* 25: 7–29.

De Freitas, J., Sarkissian, H., Newman, G.E., Grossman, I., De Brigard, F., Luco, A., and J. Knobe. 2018. Consistent belief in a true self in misanthropes and three interdependent cultures. *Cognitive Science* 42: 134-160

Descombes, V. 2016. *Puzzling Identities*, trans. Stephen Adam Schwartz. Harvard University Press.

Dranseika, V. 2017. On the ambiguity of 'the same person'. *AJOB Neuroscience* 8: 184-186.

Earp, B. D., Skorburg, J. A., Everett, J. A. C., and J. Savulescu. 2019. Addiction, identity, morality. *AJOB Empirical Bioethics* 10: 136-153.

Earp, B. D., Latham, S. R., and K. P. Tobia. 2020. Personal transformation and advanced directives: an experimental bioethics approach. *The American Journal of Bioethics* 20: 72-75.

Knobe, J. In prep. Personal identity and dual character concepts.

Locke, J. 1975. *An Essay Concerning Human Understanding*, ed. P. H. Nidditch. Oxford University Press.

Molouki, S., and D. M. Bartels. 2017. Personal change and the continuity of the self. *Cognitive Psychology* 93: 1-17.

Molouki, S., Cheng, S. Y., Urminsky, O., and D. M. Bartels. 2020. How personal theories of the self shape beliefs about personal continuity and transformative experience. In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, ed. E. Lambert and J. Schwenkler. Oxford University Press.

Newman, G. E., Bloom, P., and J. Knobe. 2013. Value judgments and the true self. *Personality and Social Psychology Bulletin* 20: 1-14.

Newman, G. E., De Freitas, J., and J. Knobe. 2015. Beliefs about the true self explain asymmetries based on moral judgment. *Cognitive Science* 39: 96-125.

Nichols, S. 2017. Memory and personal identity. In *The Routledge Handbook of Philosophy of Memory*, ed. S. Bernecker and K. Michaelian. Routledge.

Nichols, S., and M. Bruno. 2010. Intuitions about personal identity: an empirical study. *Philosophical Psychology* 23: 293-312.

Nichols, S., Strohminger, N., Rai, A., and J. Garfield. 2018. Death and the self. *Cognitive Science* 42: 314-332.

Parfit, D. 1970. *Reasons and Persons*. Oxford University Press.

Paul, L. A. 2014. *Transformative Experience*. Oxford University Press.

Paul, L. A.. 2020. Who Will I Become? In *Becoming Someone New: Essays on Transformative Experience, Choice, and Change*, ed. E. Lambert and J. Schwenkler. Oxford University Press.

Prinz, J., and S. Nichols. 2016. Diachronic identity and the moral self. In *The Routledge Handbook of Philosophy of the Social Mind*, ed. J. Kiverstein,

Reid, T. 2002. *Essays on the Intellectual Powers of Man*, ed. D.R. Brookes. Pennsylvania State University Press.

Searle, J. 2005. The self as a problem in philosophy and neurobiology. In *The Lost Self: Pathologies of the Brain and Identity*, ed. T.E. Feinberg and J.P. Keenan. Oxford University Press.

Starmans, C., and P. Bloom. 2018. Nothing personal: what psychologists get wrong about identity. *Trends in Cognitive Science* 22: 566-568.

Strohminger, N., Knobe, J., and S. Nichols. 2017. The true self: a psychological concept distinct from the self. *Perspectives on Psychological Science* 12: 551-560.

Strohminger, N., and S. Nichols. 2014. The essential moral self. *Cognition* 131: 159-171.

Strohminger, N., and S. Nichols. Neurodegeneration and identity. *Psychological Science* 26: 1468-1479.

Swinburne, R. 2019. *Are We Bodies or Souls?* Oxford University Press.

Tobia, K. 2015. Personal identity and the Phineas Gage effect. *Analysis*, 75: 396-405.

Tobia, K. 2016. Personal identity, direction of change, and neuroethics. *Neuroethics* 9: 37-43.

Weaver, S., and J. Turri. 2019. Personal identity and persisting as many. *Oxford Studies in Experimental Philosophy* 2: 213-242.

Williams, B. 1970. The self and the future. *The Philosophical Review* 79: 161-180.