



Cognitive Science 47 (2023) e13317

© 2023 The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13317

Do Obligations Follow the Mind or Body?

John Protzko,^{a,b} Kevin Tobia,^c Nina Strohminger,^d Jonathan W. Schooler^{a†}

^a*Department of Psychological & Brain Sciences, University of California, Santa Barbara*

^b*Department of Psychological Science, Central Connecticut State University*

^c*Georgetown University Law Center, Georgetown University*

^d*Legal Studies and Business Ethics, The Wharton School, The University of Pennsylvania*

Received 22 August 2022; received in revised form 15 May 2023; accepted 16 June 2023

Abstract

Do you persist as the same person over time because you keep the same mind or because you keep the same body? Philosophers have long investigated this question of personal identity with thought experiments. Cognitive scientists have joined this tradition by assessing lay intuitions about those cases. Much of this work has focused on judgments of identity continuity. But identity also has practical significance: obligations are tagged to one's identity over time. Understanding how someone persists as the same person over time could provide insight into how and why moral and legal obligations persist. In this paper, we investigate judgments of obligations in hypothetical cases where a person's mind and body diverge (e.g., brain transplant cases). We find a striking pattern of results: In assigning obligations in these identity test cases, people are divided among three groups: "body-followers," "mind-followers," and "splitters"—people who say that the obligation is split between the mind and the body. Across studies, responses are predicted by a variety of factors, including mind/body dualism, essentialism, education, and professional training. When we give this task to professional lawyers, accountants, and bankers, we find they are more inclined to rely on bodily continuity in tracking obligations. These findings reveal not only the heterogeneity of intuitions about identity but how these intuitions relate to the legal standing of an individual's obligations.

Keywords: Obligations; Identity; Responsibility

Correspondence should be sent to John Protzko, Central Connecticut State University, 1619 Stanley St, New Britain, CT 06053, USA. E-mail: protzko@gmail.com

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

1. Do obligations follow the mind or body?

Suppose someone owes you \$100, but then undergoes a large transformation. What happens to that \$100 debt? A common answer given in the philosophy of personal identity is that the latter person still owes you the debt, but only if the person is “the same person” as the earlier one. If the transformation is truly radical, we might think that the person who owed the debt is no longer around, and the latter person is no longer the one who owed you the debt.

What might constitute such a truly radical transformation is a central question of personal identity. What makes you the same person as your past and future selves? Philosophers have proposed two primary classes of answers to this question (see Shoemaker & Tobia, 2022). On the *mental continuity* view, identity persists because past and future selves share the same mind. On the *biological continuity* view, identity persistence results from the continuity of the body.

As the opening debt example suggests, questions about personal identity are connected to questions about inter-temporal practical concerns, such as obligations held over time. If certain changes replace one person with a new one, this may have implications for the obligations of the former person (which are not obviously owed by the new person). Our aim in this paper is to probe lay intuitions about identity by understanding what people believe happens to our obligations in cases of possible identity change.

Traditionally, philosophers evaluate answers to personal identity questions by conducting thought experiments. In John Locke’s (1689/1905) famous example of the Prince and the Cobbler, Locke invites us to imagine a prince, whose consciousness enters the body of a cobbler. He offers this thought experiment to elicit a shared intuition: The person with the prince’s consciousness and cobbler’s body is *the same person* as the prince. Is this intuition shared by laypeople?

There are many thought experiments about personal identity, some of which have been taken to support the mental continuity view and others that support the biological continuity view (Williams, 1970). Experimental studies have also been brought to bear on what properties affect lay attributions of personal identity (Nichols & Bruno, 2010; Strohminger & Nichols, 2014, 2015; Tobia, 2015; Turri & Weaver, 2018).

One debate that has emerged from these studies concerns the significance of lay participants’ judgments about these test cases. In many experiments, participants report their agreement with metaphysical non-identity statements like “John is no longer the same person” (Nichols & Bruno, 2010; Strohminger & Nichols, 2014, 2015; Tobia, 2015; Turri & Weaver, 2018). But some critics have suggested that laypeople do not really mean to endorse non-identity (Starmans & Bloom, 2018). Participants agreeing with these statements may be using hyperbole, intending to communicate something closer to “he’s not himself today” than “he’s now a metaphysically different person.”

A novel way of advancing this debate is to study lay views about obligations. In the philosophy of identity, obligations play a key role. For example, Locke was interested in identity as a metaphysical question, and also because identity is a forensic relation. Identity helps us keep track of obligations. Why does *he* (and not someone else) owe you money? Because we believe he is *the same person* as the man to whom you originally loaned money.

Here, we study laypeople's views about obligations, to provide new evidence about whether lay endorsement of "identity change" reflects mere hyperbole or a deeper commitment. If statements like "he is not the same person" are mere hyperbole, then laypeople's judgments about obligations in those cases should not be affected. But if those identity statements reflect a deeper commitment, then they should parallel judgments about obligations. Thus, what happens to obligations are another way of understanding identity relations.

We use a novel experimental paradigm to assess intuitions about cases in which mind and body diverge. This allows us to test whether participants judge that obligations (a) follow the mind, (b) follow the body, (c) are split between both mind and body, or (d) disappear (as happens with debt after estates are settled when someone dies). We also test the same options for inter-temporal obligations, allowing us to assess whether lay identity judgment carries commitments about moral and legal obligations.

Our investigation builds upon previous investigations by speaking to lay judgment of obligations, assessing whether laypeople evaluate obligations similarly to identity. We also increase the number of possible persons (continuers) to whom the obligation/identity can be assigned by introducing a novel brain transplant scenario. Finally, we extend the work to professional accountants, lawyers, and bank managers to evaluate how people with relevant professional experience regarding the continuance of obligations evaluate such questions about obligation transfer and the possible legal standing of identity.

Study 1 is a preregistered conceptual replication of previous work on identity change and tests whether obligations may, for some people, follow the physical body and not "identity" or the mind. In fact, it is possible that there is not one unified way people think about obligations, but different people systematically use different criteria when deciding between obligations following the mind or the body. This possibility is explored in Study 2a, while in Study 2b, we test which individual differences may predict variance in such beliefs (see also SOM). Finally, we test these cases on professionals who have experience in the transfer of obligations of persons over time, including accountants, estate lawyers, and bank managers.

2. Study 1

Study 1 first begins by replicating previous work on when people believe someone would lose their identity. We do this to ensure the basic phenomenon can be built upon reliably, testing whether people believe someone who has lost their sense or morality is interpreted as "no longer the same person" (Strohming & Nichols, 2014). We extend this work by asking what people believe happens to the obligations the person had; if someone is "no longer the same person," is the person free from previous obligations?

3. Methods

Participants were 289 Americans recruited from Amazon Mechanical Turk (mTurk) with a 95% approval rating or higher, who had completed no more than 500 studies. They also

could not have completed one of our previous studies on this topic. The procedures were preregistered at <https://osf.io/cpzqtq>.

Participants began by reading the following information about a person named “J—.” No actual name was given to help ensure participants would not interpret information about the individual from their name. Whether the vignette was about a man or a woman was randomly assigned between participants. Full wording for all scenarios is available at <https://osf.io/j9q6e/>.

Please consider the following hypothetical scenario about a woman named J—: Note that this information comes from a list of information about J—. We are only showing you a random subset of information about J—.

Ever since J— was born, it was clear that there was something distinctive about her personality. She sometimes did bad things to other people, but deep down in her very essence, she was a fundamentally good person. At the very core of her being, she had a profound compassion for other people and a genuine concern for their well-being.

Last Monday, J— went to the bank and took out a loan for \$10,000. As J— was driving home from the bank, she was going faster than normal. She was pulled over and was given a speeding ticket.

As a manipulation check, participants then rated J—’s true self as 1 = *fundamentally evil* to 7 = *fundamentally good*. Since all participants are told the person is fundamentally good, we consider anything below a 5 a failure of attention. Thirty-six percent of the sample failed this attention check. As we did not preregister whether we would drop these participants for the analysis, however, we kept those participants in the study. We did, however, preregister an exploratory moderation test. This moderation was not statistically significant ($p = .402$).

On the next page, participants read the following story (adapted from Strohminger & Nichols, 2014):

One day as she is walking down the street, J— sustains a severe head injury from construction equipment accidentally being dropped on her head. Her 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 them 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.

In the memory loss condition, participants learn that the individual has forgotten how to swim:

They discover that the transplant recipient thinks and acts the same way as before the accident, except she has forgotten how to swim.

While in the morality loss condition participants learn the individual has lost their moral conscience:

They discover that the transplant recipient thinks and acts the same way as before the accident, except she has lost her moral conscience—she is no longer capable of judging right from wrong or being moved by the suffering of others.

On the next page, half of the participants were asked the following question about personal continuity:

Please rate which of the following you believe to be true about the transplant recipient: After the transplant recipient wakes up, the transplant recipient is : [*a completely different person than before the transplant = 4, a somewhat different person than before the transplant = 3, somewhat the same person as before the transplant = 2, completely the same person as before the transplant = 1*; participants did not see any numbers].

Only half of the participants were asked this question to prevent the possibility that being asked the question would itself influence people's beliefs on the obligation question. On the next page, all participants were asked whether the transplant recipient has to pay the speeding ticket and pay back the bank loan: "*Does the transplant recipient have to pay the speeding ticket from last Monday? [yes/no].*" "*Does the transplant recipient have to pay back the \$10,000 loan to the bank? [yes/no].*"

4. Results

We find that loss of moral conscience leads a person to be judged as having changed more than when they lose a type of memory, consistent with prior research (Strohming & Nichols, 2014). For morality loss versus memory loss, there was a 2.8 log odds decrease in belief that J— was the same person (95% CI = -3.62 to -1.97). Treating the data as continuous instead of ordered-categorical leads to the same conclusion ($\beta = -0.59$, $p < .001$; Gomila, 2021). Having replicated the basic finding that losing one's moral sense leads to an especially large disruption of personal identity, we test whether this changed person is also less responsible for the earlier person's obligations.

Eighty-seven percent of participants believed the individual still had to pay back a loan, and 92% of participants believed the individual still had to pay a speeding ticket. Neither of these decisions was altered by the type of mental faculty that was lost (both b 's_{probit} (287) < 0.05 , both $ps > .77$), nor moderated by whether participants were asked about identity change

before answering the obligations question (both $ps > .1$). It seems that people can endorse the belief that someone has become *a completely different person* without believing that this eliminates all of their prior obligations.

5. Study 2a

One way to explain the dissociation between identity and obligation in Study 1 is that people are interpreting the question about someone becoming a *different person* as referring to the degree of features changed (*qualitative identity* change) rather than whether these are two separate people (*numerical identity* change; Dranseika, Nichols, & Strohminger, 2023; Starmans & Bloom, 2018).

Alternatively, people might believe that someone's numerical identity has changed while also believing that obligations remain intact—because the physical body is what carries the obligation. Our next series of studies investigate this possibility. In this study (and Studies 2b and 3), we present participants with scenarios involving two plausible continuers, one continuer that shares the original person's mind and another that shares the original person's body.

5.1. Methods

Participants were 300 Americans recruited from mTurk. This study was preregistered prior to data collection at <https://osf.io/6jrph>.

All participants filled out an attention check, which was administered to increase participants' seriousness and data quality (see Oppenheimer, Meyvis, & Davidenko, 2009). No participants were removed based on the attention check, consistent with our preregistration plan. On the next page, participants read a vignette that described either a man or a woman (as with Study 1). For full study materials see <https://osf.io/cywz2/>. Participants were randomly assigned to one of three conditions, designed to measure three types of obligation: loan, charity, or tax penalty. In the loan condition participants read the following:

Please consider the following hypothetical scenario about a woman named Jane.

Jane took out a \$2000 loan from her bank. As she is driving away from the bank, she is speeding down the highway. She gets into a terrible car accident with Anne, who was also speeding. Both women fall into a coma, the doctors believe that neither woman will survive. The doctors decide to try an experimental procedure, they attempt a brain transplant between Jane and Anne. Since neither woman has any family and they were about to die, the doctors continue with the surgery. They transplant Jane's brain into Anne's head, and Anne's brain into Jane's head. The transplant is a success. What is more, after the surgery, both women wake up from their comas. Once the women wake up, the doctors test all physiological responses and determine that both transplant recipients are functioning well. There are no negative side effects from the operation.

In the charity condition, participants read the same but learned the person had promised \$2000 to a local charity. In the tax penalty condition, participants read the same but learned the person was fined \$2000 for cheating on their taxes. All participants were asked the following:

Both Jane and Anne have their bank accounts at the same bank. The bank needs to make the withdrawal to [have the \$2000 loan paid back/donate the \$2000 promised to the charity/pay the IRS \$2000].

Before the transplant, Jane had bank account #J09876. Before the transplant, Anne had bank account #A54321. Now Jane's body has Anne's brain in it and Anne's body has Jane's body in it.

Which account should the bank take the money out of and how much? You may choose any amount from either, both, or neither bank account.

Participants gave their responses on two sliders, one labeled “#J09876” and the other “#A54321.” Both sliders went from \$0–\$2000 and were presented in randomized order between subjects.

5.2. Analysis plan

Following our preregistration plan, we calculated participant responses by dividing the amount given to the body by the total amount assigned by the participant across the two bank accounts. So if a participant assigned the full obligation to the body, their score would be 1; if they split the difference between the two, their score would be 0.5; if they took \$1500 from the body's account and \$500 from the mind's account, they would be scored as 0.75; if all money came from the mind's account they would be scored as 0. Responses that took no money from either account were given no value as they evaporated the obligation; a new variable (denoting whether the obligation was evaporated) was created and they were scored as a 1, with all those who suggested any amount of obligation scored as 0.

As these data use proportions, we use a beta distribution for analysis. Because we *a priori* predicted there would be an inflation of responses fully following the body (i.e., an excess of 1s) and an inflation of responses fully following the mind (i.e., an excess of 0s), we use the zero-one inflated beta estimator in our analyses (Buis, 2010).

5.3. Results

First, we found no evidence that the pattern of results differed by whether it was a loan, a promise to charity, or a tax penalty (all $ps > .08$, see <https://osf.io/sqsef/wiki/>). Therefore, we collapsed across these conditions and report people's beliefs about the tracking of obligations.

The data overwhelmingly were clustered on the option that all money should come from the bank account of the body that made the obligation; put simply of people that did not negate the obligation: 129/257 participants (50%) believed the money should come out of the bank account tied to the body that incurred it; 53/257 participants (21%) believed it should come

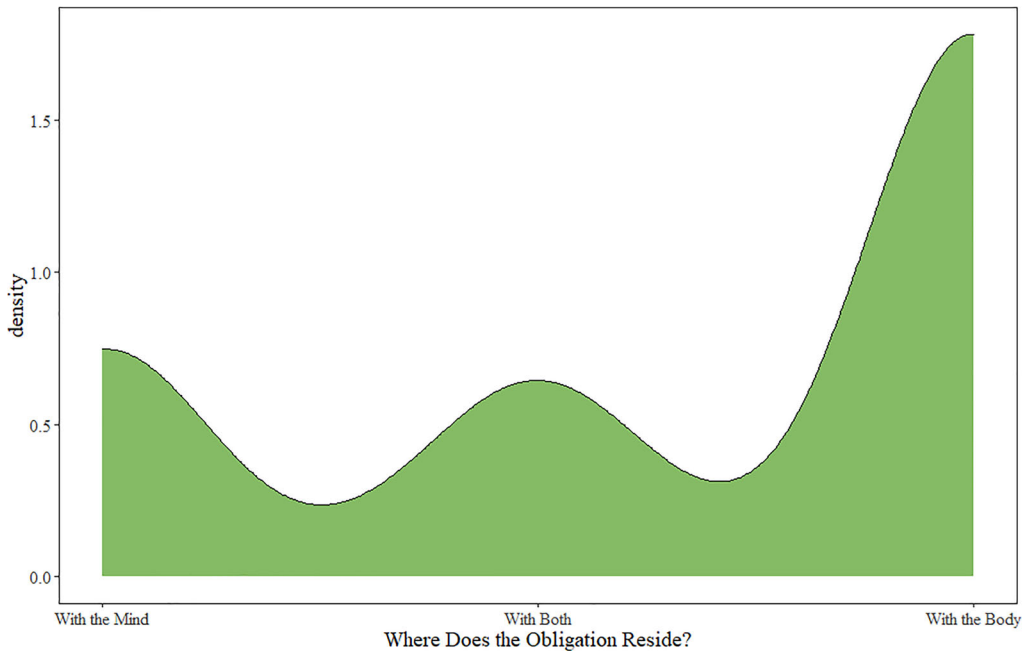


Fig 1. Density plot of responses to whether a financial obligation incurred before someone had their brain swapped with someone else should come from the mind of the person who incurred the fine (which is now in a new body), be shared between the two people, or come from the bank account of the body that incurred it (with a new mind in it).

from the mind that incurred the obligation, and the rest split the difference, with the majority (23/257, 9%) splitting it equally. Thus, participants largely agreed that the obligation rested with the body, not with the mind or somewhere in between ($b_{ZOIB_1-inflate} = 0.54$, $p < .001$, 95% CI = 0.26 to 0.83; see Fig. 1).

Thus, we have our first evidence that for some obligations there is indeed an inclination to believe the obligation is tagged to the body that incurred it, regardless of whose mind is inside. This can help explain the findings in Study 1, where participants indicated that mental changes could render someone a *different person* without vitiating their previous obligation. The body, rather than the mind, appears to be the nexus of the obligation, especially when the obligation is a monetary one (as in Study 1).

The results of this study are striking, in that a significant number of people believe the obligation resides entirely with the body that incurred it regardless of whose mind is inside (hereafter “body-followers”). One possibility could be participants did not understand or have an opinion about the scenario and were chosen randomly. Another concern is that participants do not believe the mind is going with the brain, so they may believe that while a brain transplant has occurred, the “mind” is somehow still with the body (e.g., via a soul). We replicate and test possible alternate explanations in the next study.

6. Study 2b

6.1. Methods

Study 2b was a replication of Study 2a in a larger, more representative sample using additional stimuli and measures to rule out alternative explanations.

Participants were 1500 American adults, drawn in a stratified way with unequal probabilities of selection so that those who complete the survey will resemble the nation's adult population (according to the most recently available Current Population Survey, conducted by the U.S. Census Bureau) in terms of gender, age, education, ethnicity, race, region, and income. This sampling was handled by the company CriticalMix.

Participants first filled out the same attention check as in Study 2a on separate pages (for full instructions see the SOM at <https://osf.io/2vgsa/>). Participants read a vignette that was gender-matched, as we predicted gender could moderate the results, so we also randomly assigned people to conditions within gender (see Athey & Imbens, 2017). Here, is a sample vignette:

John was recently found guilty of cheating on his taxes. He was fined \$2000 and given a 20-month prison sentence, to be served starting 1 month from now.

As he is driving away from the courthouse, he is speeding down the highway. He gets into a terrible car accident with Adam, who was also speeding.

Both men fall into a coma, the doctors believe that neither man will survive.

The doctors decide to try an experimental procedure, they attempt a brain transplant between John and Adam. Since neither man's family can be contacted and they were about to die, the doctors continue with the surgery.

They transplant John's brain into Adam's head, and Adam's brain into John's head. The transplant is a success. What is more, after the surgery, both men wake up from their comas.

Once the men wake up, the doctors test all physiological responses and determine that both transplant recipients are functioning well. There are no negative side effects from the operation.

Participants were then asked four questions, each on a separate page, in randomized order. One was on who has to pay back a fine, one was on who has to spend time in prison, one was on who gets to drive the car the accident was in, and one was on who gets to sleep with the spouse of one of the individuals. Here, is the question about prison time:

The guilty verdict was 20 months in prison. The judge needs to decide who will spend what time in jail.

Now John's body has Adam's brain in it and Adam's body has John's brain in it.

Who should spend how much time in jail? Should it be John's body with Adam's brain in it? Should it be Adam's body with John's brain in it? You may choose any amount of months below from either, both, or neither.

Participants gave their responses on two sliding scales ranging from 0 to 20. One anchor was labeled "John's body with Adam's brain in it" and the other "Adam's body with John's brain in it." The wording of the other three questions may be found at <https://osf.io/2vgsa/>.

Finally, participants filled out two comprehension check questions. These comprehension checks were to confirm that participants understood the scenario and also serve to check that participants indeed believe the mind is going with the brain (rather than the body): "*Who knows the PIN code to the bank account that used to belong to [John/Jane]?* [unnumbered: 0 = John's body with Adam's brain in it/Jane's body with Anne's brain in it; 1 = Adam's body with John's brain in it/Anne's body with Jane's brain in it]. *Who would remember [John/Jane] spouse's birthday?* [unnumbered: 0 = John's body with Adam's brain in it/Jane's body with Anne's brain in it; 1 = Adam's body with John's brain in it/Anne's body with Jane's brain in it]." These procedures, analysis plan (including code), and sample size for this study were preregistered at <https://osf.io/t9vdf>.

6.2. Results

Seventy-seven percent of participants answered both comprehension check questions correctly, suggesting people largely understand that a brain switch also implies a mind switch. Our analyses include only those who passed both comprehension checks, but additional analyses show the results do not meaningfully change when including all participants.

The results largely replicated those in Study 2a. Across all four Dependent Variables, there was a strong trimodal pattern with people split between whether the obligation follows the mind, follows the body, or is split between the two. Among those who passed both attention checks, the first scenario people saw (due to inconsistent order effects, see <https://osf.io/9e8tj/wiki/home/> for more details), we saw notably fewer body-followers in this replication than the first study (see Table 1).

Thus, overall, we still saw the trimodal pattern of obligations following the mind, the body, or being split between the two in a large sample drawn to resemble the U.S. adult population. While in Study 2b, ensuring that participants understood the scenario decreased the number of people who believed the obligation rests *entirely* with the mind, compared to Study 2a, which in no way eliminated it. This suggests that people systematically differ in their intuitions about whether obligations track the mind or the body. The question now becomes: Which individual-level factors predict this difference?

Table 1

Percent of participants who chose the obligation fully followed the mind or fully followed the body

Obligation	% Fully Agreeing the Mind in the New Body	% Fully Agreeing the Body with the New Mind
Who Pays the Fine?	35%	26%
Who Goes to Jail?	41%	11%
Who Drives the Car?	25%	13%
Who Sleeps with the Spouse?	26%	21%

With that possibility in mind, we ran yet another replication of this brain swap and obligations paradigm, testing intuitions about identity and obligations, and also measuring several potential individual difference factors that may predict responding.

7. Study 3

7.1. Methods

Participants were 1500 American adults drawn in the same manner as in Study 2b to match the demographics of the U.S. adult population with the added restriction that anyone who participated in Study 2b could not take part in this study.

Study 3 was similar to Studies 2a and 2b in that it portrayed a brain–body swap and measured intuitions about obligations following the swap. To further ensure participants understood the scenario, we included “before” and “after” drawing depicting the swap to help participants understand that brains had been moved between two bodies.

Participants were randomly assigned to fill out the essentialist beliefs scale (Horne & Cimpian, 2019) and a scale measuring the participants’ belief that the mind and body are distinct entities (mind–body dualism; Nadelhoffer, Shepard, Nahmias, Sripada, & Ross, 2014) either before or after responding to the swap scenario. The purpose of the mind–body dualism scale was to see if the belief that the brain is separable from the mind was related to the reason, as we have had a large number of “body followers.” The purpose of the essentialism scale, which at the time of study design was the only scientifically validated scale to measure individual differences in the degree people believe in essences (see Horne & Cimpian, 2019), was to measure if the reason we observed so many “body followers” was because people believe that people have an “essence” (see Starmans & Bloom, 2018) that stays with the body and is not encapsulated in the mind. An example item is “People who receive heart transplants from violent criminals could become a little more violent themselves” (see Horne & Cimpian, 2019, <https://osf.io/3zxgy/wiki/home/> for complete items).

Half of the participants read a scenario that was described as a “brain swap” (as in the previous studies). The other half learned about the same case, though it was described as a “body swap”—where the brains stayed put and the bodies were moved. This condition was

added to test the role of framing what is the originating (vs. transferred) part of the person in what people assign priority to.

In addition to collecting responses on who has to pay the fine, we added another type of obligation. Participants were told one of the individuals had recently purchased a 30-day gym membership and some days were left. This was to further ensure our results were not idiosyncratic to the specific obligations we have investigated so far. Half the participants answered who had to pay the fine, and the other half answered who had access to the gym membership.

In the end, education level was recorded and coded as years of education (following Rietveld et al., 2014). The sample size, procedures, and analysis plan including the code for this study were preregistered at <https://osf.io/yjcps>.

7.2. Results

We found no consistent effect of referring to the procedure as a body swap versus a brain swap (for analysis, see <https://osf.io/uxgck/wiki/>). This suggests people were not deciding that the obligation followed the body simply because the body stayed put while the brains were moved. We therefore collapsed this factor for all analyses, consistent with our preregistration plan.

There was a tendency for more educated people to be more likely to dismiss the fine ($b = 0.005$, $p = .038$, 95% CI = < 0.001 to 0.01) though this was not replicated in the gym-membership obligation ($b = -0.002$, $p = .261$). None of the other individual differences significantly predicted who would dismiss the obligation entirely. Thus, the remaining analyses focus on where people think the obligation resides if they think the obligation remains.

For the fine, the more someone believes in mind/body dualism the more likely they were to believe the obligation tracks the body ($b = 0.04$, $p = .035$, 95% CI = 0.003 to 0.08). In addition, the more someone believes in essences, the more likely they were to believe the obligation tracked the body ($b = 0.15$, $p = .001$, 95% CI = 0.06 to 0.25). Likewise, the more someone believed in essences, the less likely they believed the obligation stays with the mind ($b = -0.2$, $p < .001$, 95% CI = -0.25 to -0.14). No other relationships were significant (see full output at <https://osf.io/uxgck/wiki/>). Concerning who gets to use the gym membership, we again found mind-followers were the least likely to believe in essences ($b = -0.13$, $p < .001$, 95% CI = -0.18 to -0.08). We also found that the best-educated people were the most likely to be mind-followers ($b = 0.01$, $p = .045$, 95% CI = < 0.001 to 0.02 ; see Fig. 2). This offers one possible explanation of why the idea of an obligation tracking the physical body has received comparatively less support among philosophers (Bourget & Chalmers, 2014); the higher an education one has, the more likely one is to believe obligations track the mind that incurred them.

As an exploratory analysis requested by one reviewer, we found that the more years of education someone had, the less they believed in essences ($r_S(1498) = -.15$, $p < .001$). Furthermore, years of education were unrelated to participants' beliefs in mind-body dualism; people tended to endorse mind-body dualism ($M = 5.11$ on a 7-point scale) regardless of their level of education ($r_S(1498) = -.02$, $p = .385$).

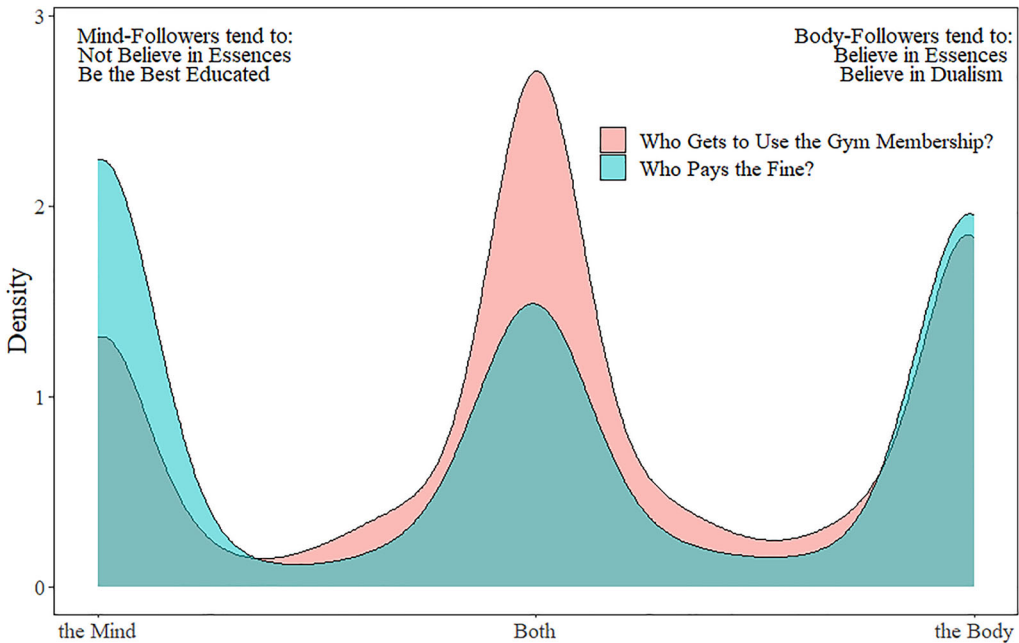


Fig 2. Where obligations are thought to reside, represented as the proportion answering body, mind, or both. Distribution as a density plot of choices in a brain swap scenario about whether the obligations follow the body (rightmost in the graph), obligations follow the mind (leftmost in the graph), obligations are split between the two (middle peak) or somewhere in between. People who believed the obligation went entirely with the mind (leftmost) tend to have a lower belief in essences and are better educated. People who believed the obligation went entirely with the body (rightmost) tended to believe in dualism and essences.

Thus, Study 3 revealed some important new insights about identity intuitions. The variety of views about obligations observed in previous studies is not an error or random response but is systematically related to how individuals think about essences, mind/body dualism, and to some extent their education.

The role of education in generating such intuitions gives us some hint that the lay populace may have very different intuitions than those whose training requires a nuanced understanding of obligations. To investigate this, Study 4 measured identity-swap intuitions among professionals with expertise in tracking obligations.

8. Study 4

While the results of Study 3 showed the highest educated people were most likely to be mind-followers, professionals with training and experience related to inter-temporal obligations may see things rather differently. This provides the first use of such professional groups in these types of scenarios.

8.1. Methods

Study 4 was a direct replication of Study 3, with the following changes. We included a new within-subjects obligation: “Which later person has to fulfill a work contract that the earlier person previously signed?” We also included measures of how confident people were in their decisions of who would have to pay back the obligation and a direct question about the identity of the individuals. Finally, we ran the study simultaneously in a naïve sample and in a sample of professionals who have experience tracking obligations; these professionals were bank managers, lawyers, and accountants.

Lay participants were 679 mTurkers from the United States with a 95% approval rating or higher and having participants in no more than 500 previous studies nor in one of our previous studies on this topic. Professional participants were 576 professionals who were lawyers (probate and estate administration, healthcare/health law, tax law, or banking law), accountants (estate planning, medical or hospital accounting, income tax preparation and planning (for others), or bank accounting), and bank managers.

Professionals were first asked about their relative expertise, and those not fulfilling one of our relevant categories were excluded from participating in the study. Participants received the fine-dependent measure from Study 3 and the work contract-dependent measure in randomized order (for full materials, see <https://osf.io/aevwt/wiki/home/>). For the work contract obligation, participants read:

Before the accident, John had signed an employment contract stating that John would work at a new company for 40 h each week. Adam did not have any employment contract with that company.

Once both patients recover from surgery, for how many hours a week is each responsible to work for the new company, according to the employment contract? You may choose any amount of hours from either, both, or neither of them.

Response options were two sliding scales, presented in random order, with anchors at one end labeled John’s body with Adam’s brain in it and the other anchor labeled Adam’s body with John’s brain in it. After the scenarios and questions, participants were asked the level of confidence they had with their answer on a 5-point scale (following Mastrandrea et al., 2010).

For the explicit identity judgments, participants were asked (following Weaver & Turri, 2018):

John’s body with Adam’s brain in it is wheeled into the West Recovery Room, while

Adam’s body with John’s brain in it is wheeled into the East Recovery Room.

Where is John?

Participants could select either in the West Recovery Room, in the East Recovery Room, in both Recovery Rooms, or in neither Recovery Room. Response options were presented in random order.

Participants responding “in both Recovery rooms” were asked:

To what extent is John in the West Recovery Room and to what extent is John in the East Recovery Room?

You can select any percent in each room but the total must equal 100% [John is in the East Recovery Room, John is in the West Recovery room, presented in random order].

Participants were also asked about their confidence in their answers to the explicit identity question. Finally, we asked participants about their highest level of education. The sample size, procedures, and analysis plan (including code) for this study were preregistered at <https://osf.io/tzk8w>.

8.2. Results

Professionals were less likely to believe in essences than lay participants ($\beta = -0.13$, $p < .001$, 95% CI = -0.19 to -0.07). Professionals also tended to believe more in mind/body dualism, though the effect was small and statistically non-significant ($\beta = 0.05$, $p = .06$). Based on the results of Study 3, as our professional groups were better educated and less likely to believe in essences, we might expect these professionals to be more likely to believe the obligation tracks the mind.

8.2.1. Who should pay the fine?

Professionals showed largely the same trimodal pattern as laypeople. Professionals were more likely to assign the fine to the body than the general population and less likely to split the difference between the two accounts ($b_{ZOIB} = 0.43$, $p = .007$, 95% CI = 0.12 to 0.74). They were no more likely to say the obligation evaporated ($b_{logit} = 0.06$, $p = .892$). These results do not change when conditioning on the belief in mind/body dualism (i.e., adding dualism as a covariate).

8.2.2. Who gets to work for the company?

Concerning who should work for the company, the results were largely the same. Both groups had a trimodal pattern, with professionals more likely to say the body of the person who signed the work contract should fulfill the obligation—even with someone else’s brain inside of it ($b_{ZOIB} = 0.38$, $p = .021$, 95% CI = 0.06 to 0.7). These results likewise do not change when conditioning on beliefs in mind/body dualism.

8.2.3. Explicit identity judgment

Explicit identity judgments followed the same pattern as obligation judgments. Overall, professionals were more likely to say that John was in the room with his body (40% in

professionals vs. 32% in laity, $b = 0.59$, $p = .001$, 95% CI = 0.25 to 0.92). Given that there was a strong covariance between people's judgments of identity and obligations for both paying the fine ($r_S = .45$, $p < .001$) and who should fulfill the work contract ($r_S = .44$, $p < .001$), this should be unsurprising. Furthermore, the covariance of identity judgments and obligations was the same in both groups (not moderated by professional status; $p = .125$ for pay and $p = .122$ for work).

Compared with laypeople, lawyers, accountants, and bank managers were more likely to endorse the view that the body, not the mind, is where identity and obligations reside. There were no differences among the three different professional groups (see <https://osf.io/cw5qg/wiki/>).

We have now seen that in a hypothetical brain swap case, a substantial number of people believe the obligation would stay with the physical body that incurred it, not in the mind. Furthermore, among those in professional careers directly relating to the tracking of obligations, the tendency to track the obligation with the body was enhanced, despite being better educated and less likely to believe in essences (both of which would predict tracking the obligation to the mind).

9. General discussion

Across a series of studies, we found that the relationship between identity and obligation varies substantially depending on both the number of continuers and the individual decision-makers psychological traits. In Study 1, changes in identity largely failed to impact assessments of obligation. In Studies 2–4, we found a more complex pattern of results. While perceptions of obligation largely followed people's assessments of identity (as in Study 4), their assessment of the continuation of both varied markedly. Furthermore, the majority of participants understood that when brains were switched between two bodies, the thoughts and memories also transferred with the brain.

Some people believe obligations and identity judgments follow the body, others believe they follow the mind, and still others split obligations between the mind-continuer and body-continuer. Despite strong philosophical arguments favoring identity and obligations tracking the mind (e.g., Locke, 1689/1905), lay judgments are notably more varied, and in our studies, many people believed obligations and identity actually tended to follow the body with someone else's mind inside it. Study 2b demonstrates that this was not a function of participants misunderstanding the scenario. Studies 3 and 4 find that the tendency to believe obligations and identity follow the body is predicted by specific beliefs about essences, mind/body dualism, and education levels.

Whether one is a mind-follower, body-follower, or splitter was predicted by several psychological traits, suggesting that participants' decisions were not arbitrary. Furthermore, the use of comprehension checks did not moderate the results, so the variety of assigning obligations was not due to participants not understanding the scenarios. We found physical essentialism and mind/body dualism predict body-following, while the best-educated participants are

more likely mind-followers and the least educated are more likely splitters. The professional experts were more likely to be body-followers.

Essentialism predicted the belief that obligations track the body. This may seem mysterious, until we consider that much of essentialism has to do with tracking physical (if invisible) properties. A sample item from the Beliefs in Essentialism scale, for example, is: “*People who receive heart transplants from violent criminals could become a little more violent themselves*” (Horne & Cimpian, 2019). If someone believes that essences are physically real in this way, it makes sense that they would also believe that obligations and identity go with the body. Yet the measurement of the belief in essences is not just relegated to the acceptability of receiving an organ transplant from someone with undesirable traits. Items on the scale also include concerns over the authenticity of artwork and feelings of discomfort interacting with items owned or used by people with undesirable traits. That all these items comprise a single factor suggests they measure more than discomfort or authenticity but likely a belief in essences. More work, however, should be done to ensure the inferences drawn from this scale about the belief in essences are highly valid and, especially, whether other measures and conceptualizations of essences would produce the same results. Finally, more research is needed to gain conceptual clarity on whether essences are a physical substance and/or something that transcends the physical could help elucidate folk belief in essences.

Consideration of specific items in the Mind/body Dualism Scale (Nadelhoffer et al., 2014) similarly offers insight into its relationship with the continuity of obligation in this study. Items like *human action can only be understood in terms of our souls and minds and not just in terms of our brains* indicate that for mind/body dualists, a person is not reducible to their brain. Accordingly, for mind/body dualists, though the brain may change, something else remains in the body that maintains both identity and obligations. This is especially important as all participants used passed both attention check items and therefore understood that the mind traveled with the brain into the new body.

The results here provide evidence of the relationship between intuitions about numerical identity and practical obligations. Participants—whether mind-followers, body-followers, or splitters—relied on consistent internal criteria in assigning obligations and attributing identity. In brain-swap cases, for example, mind-followers were inclined to judge the person with the same mind as the same person *and* the person with the continuous mind inherits the prior obligations. Body-followers, meanwhile, believe the person with the same body is the same person and retains those obligations and identity, regardless of whose mind is inside.

9.1. Limitations and future directions

This series of studies has attempted to shed light on questions commonly used to study personal identity—questions like “Is John the same person?”—by asking about *obligations*. Of course, one might now raise concerns about these new questions. For example, Study 2a asks: “Which account should the bank take the money out of and how much? You may choose any amount from either, both, or neither bank account.” We take “should” in this question to be expressing a deontic modal about *obligation*. There may be other possible ways to interpret this statement.¹ For example, perhaps some participants understood the question predictively

or empirically: From which account *will* the bank take the money? We find this predictive reading an unusual way for participants to understand “should” in Studies 2a and 2b, although it is possible that this reading is more accessible in other studies (e.g., Study 3).

Another concern about the questions relates to their ability to identify a sense of obligation that is perfectly screened off from unrelated pragmatic concerns. Consider for example the question in Study 2a, which is repeated in the previous paragraph. Perhaps a participant reasoned, “the mind of the person who took out the loan is now in a different body, and that person should be held responsible; however, it would be practically difficult to establish that the original person’s mind is there, so perhaps we should just collect the loan from the person with the original body.” Or perhaps a participant reasoned, “the person with the same mind should be held responsible, but adopting such a policy would lead to bad social consequences, so we should just collect the loan from the person with the same body.”

Call these interpretations “mixed prudential” readings. In such a reading, the “should” questions are not only capturing participants’ basic judgment about people’s inter-temporal obligations but also participants’ considerations about unrelated factors, like which choices would be easy or beneficial to enact as social or legal policies.

A first response to this concern is to recall that the tri-modal pattern emerged across diverse stimuli—fines, prison sentencing, car ownership, and romantic relations. It is not clear what uniform “mixed prudential” reading could plausibly explain this pattern across these different contexts. Of course, perhaps one could elaborate four (or more) different pragmatic effects that implicate each of the different scenarios, which happen to result in a consistent pattern. The current paper cannot rule out all such possibilities, but future work could elaborate in more detail on how that result could emerge and empirically evaluate those new hypotheses. These findings help inform debates about how to interpret studies where participants insist that someone is “no longer the same person”: Are those statements hyperbole, or do they reflect a deeper metaphysical evaluation? Our data suggest a complicated answer. Judgments of obligations depend on some core philosophical positions of the people being asked. Both the extent to which one believes in essences and the extent one believes in mind/body dualism predict how one thinks about identity and obligations. Furthermore, the more educated one is, the more likely one is to believe identity and obligations reside with the mind, not the body. Therefore, it is not as simple as “laypeople believe/do not believe X,” we instead must look at *which* people believe *what*. This will undoubtedly complicate drawing broad inferences from what *all* people supposedly believe or endorse.

Finally, professionals with more experience in tracking obligations differed from other highly educated participants in seeing obligation as following the body. For lawyers, bankers, and accountants, it may be useful to rely on a concrete and easily identifiable objective criterion like the body in assigning identity and obligations. Moreover, it may be useful to rely on a criterion that admits *external* identification and assessment, for the purposes of enforcement. The body may be a more useful criterion than the mind for this purpose.² Therefore, our professionals’ work experiences of enforcing the transfer of obligations may have been more likely to endorse the more enforceable option.

10. Conclusion

Contrary to Locke, both professionals and many laypeople do not unanimously endorse a straightforward mind-criterion of identity and obligations. In brain transplant cases, with participants understanding that the mind travels with the brain into the new body—for many, identity and obligation follow the body, no matter whose mind is inside it.

Acknowledgments

The authors would like to thank Christina Starmans for her helpful comments and discussions. Thanks go to Madeline Gross for her help with the drawings used in Studies 3 and 4, and John Scully and Bill Scully for their help in identifying relevant professional groups. Protzko was supported by a grant from the Fetzer Franklin Fund.

Open Research Badges



This article has earned Open Data and Open Materials badges. Data and materials are available at <https://osf.io/sqsef>.

Notes

- 1 Thanks to a reviewer for these suggestions.
- 2 We thank a reviewer for this point.

References

- Athey, S., & Imbens, G. W. (2017). The econometrics of randomized experiments. In A. V. Banerjee & E. Duflo (Eds.), *Handbook of economic field experiments* (Vol. 1, pp. 73–140). Amsterdam: North-Holland.
- Bourget, D., & Chalmers, D. J. (2014). What do philosophers believe? *Philosophical studies*, 170, 465–500.
- Buis, M. L. (2010). *ZOIB: Stata module to fit a zero-one inflated beta distribution by maximum likelihood*. Statistical Software Components S457156, Boston College Department of Economics (revised 08 Aug 2012).
- Dranseika, V., Nichols, S., & Strohminger, N. (2023). Which kind of sameness? Disambiguating two senses of personal identity with a novel linguistic task. *Cognition*, 238, 105545.
- Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, 150(4), 700–709. <https://doi.org/10.1037/xge0000920>
- Horne, Z., & Cimpian, A. (2019). Intuitions about personal identity are linked to essentialist thinking in adults and children. *Cognition*, 191, 103981.
- Locke, J. (1905). *Locke's essay concerning human understanding* (No. 58). Chicago, IL: Open court Publishing Company.
- Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., Plattner, G.-K., Yohe, G. W., Zwiers, F. W. (2010). Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. Available at: https://pure.mpg.de/rest/items/item_2147184/component/file_2147185/content. Accessed April 28, 2020.

- Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. T. (2014). The free will inventory: Measuring beliefs about agency and responsibility. *Consciousness and Cognition*, 25, 27–41.
- Nichols, S., & Bruno, M. (2010). Intuitions about personal identity: An empirical study. *Philosophical Psychology*, 23(3), 293–312.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45, 867–872.
- Rietveld, C. A., Conley, D., Eriksson, N., Esko, T., Medland, S. E., Vinkhuyzen, A. A., ... & Social Science Genetics Association Consortium. (2014). Replicability and robustness of genome-wide-association studies for behavioral traits. *Psychological science*, 25(11), 1975–1986.
- Shoemaker, D., & Tobia, K. (2022). Personal Identity. In M. Vargas & J. M. Doris (Eds.), *The Oxford Handbook of Moral Psychology*, Oxford Handbooks (online edn, Oxford Academic), <https://doi.org/10.1093/oxfordhb/9780198871712.013.28>
- Starmans, C., & Bloom, P. (2018). Nothing personal: What psychologists get wrong about identity. *Trends in Cognitive Sciences*, 22(7), 566–568.
- Strohming, N., & Nichols, S. (2014). The essential moral self. *Cognition*, 131(1), 159–171.
- Strohming, N., & Nichols, S. (2015). Neurodegeneration and identity. *Psychological Science*, 26(9), 1469–1479.
- Tobia, K. P. (2015). Personal identity and the Phineas Gage effect. *Analysis*, 75(3), 396–405.
- Weaver, S., & Turri, J. (2018). 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(2), 161–180.