**On Scepticism About Personal Identity Thought Experiments**

Word count 10 753

Abstract

Many philosophers have become sceptical of the use of thought experiments in theorising about personal identity. In large part this is due to work in experimental philosophy that appears to confirm long held philosophical suspicions that thought experiments elicit inconsistent judgements about personal identity, and hence judgements that are thought to be the product of cognitive biases. If so, these judgements appear to be useless at informing our theories of personal identity. Using the methods of experimental philosophy, we investigate whether people exhibit inconsistent judgements and, if they do, whether these judgements are likely to be the source of cognitive bias or, instead, sensitivity to some relevant factor. We do not find that people’s judgements are sensitive to any of the factors we investigate (relevant or irrelevant), nor that people have inconsistent judgements across cases. Rather, people’s judgements are best explained by them having a very minimal account of what it takes for a person to survive. Since this pattern of judgements is no reason to think that we are subject to cognitive bias, we see no reason, as things stand, to be sceptical of our judgements.

1. Introduction

It was once standard fare, when theorising about personal-identity, to appeal to thought experiments.[[1]](#footnote-2) This appeal has, over the last few decades, gradually declined, to be replaced with what we call *PI* *thought-experiment scepticism.* Of course, there are those who are, quite generally, sceptical of the role of thought experiments in philosophy.[[2]](#footnote-3) For instance, has been argued that the scenarios described by PI thought experiments are often too far removed from the actual world so that applying our concepts to such worlds is like throwing darts in the dark. (Wilkes 1988: ch.1; Meier 2022). But that scepticism has been especially potent in theorising about personal-identity, because *in* *addition* to all of the perfectly general concerns one might have about thought experiments, it has been thought that personal-identity thought experiments face an additional special sort of problem: namely that our judgements in those cases are strongly subject to cognitive bias.

Cognitive biases are usually taken to be systematic errors in thinking that occur when people process or interpret information, and which then affect decisions and judgements. It is not entirely uncontroversial whether cognitive biases are really errors in thinking, or whether, if they are, they are really problematic, at least much of the time, (see Gigerenzer 2010 for the view that they are not). We will assume, however, that if judgments about personal identity elicited from thought experiments are the product of cognitive biases, then this is at least *prima facie* reason to be suspicious of them.

The idea that our judgements about PI thought experiments are subject to some kind of cognitive bias owes its origin to Williams (1970).[[3]](#footnote-4) Williams presented readers with several thought experiments. In the first of these a person at one time is psychologically, but not bodily, continuous with a person at another time. In the second the person at one time is physically, but not psychologically, continuous with a person at another time. Williams predicted that this former thought experiment would elicit judgements consistent with the psychological continuity theory, and the latter would elicit judgements consistent with the bodily continuity theory. That is, he predicted that in one case people would make judgements that suggest they hold that psychological continuity is both necessary and sufficient for personal persistence, and in the other case they would make judgements that suggest they hold that bodily continuity is both necessary and sufficient for personal persistence. But if so, people make *incompatible judgements*, since if bodily continuity is sufficient for personal persistence, then psychological continuity cannot be necessary, and if psychological continuity is sufficient for persistence, then bodily continuity cannot be necessary. In turn, this is some reason to suppose that people’s judgements in such cases might be the product of cognitive bias (though there may be disagreement about which bias this is)[[4]](#footnote-5). If so, it seems reasonable to worry that these judgements are unreliable, and hence to entertain some PI thought-experiment scepticism.

This paper examines the question of whether we have reason to think that our judgements are indeed subject to some such bias. Thus, we focus on whether there is a *special* problem for PI thought experiments—cognitive bias—which is not often posed as a pressing problem for other thought experiments. This leaves open whether there are other more general reasons to be sceptical of thought experiments in general, including PI thought experiments:[[5]](#footnote-6) we make no attempt to address this issue.

Thus, we focus on Williams-style cases because they most clearly represent cases in which people *appear* to give *incompatible* judgements, and hence judgements that are fairly plausibly read as involving a cognitive bias.

To be clear then, we think it plausible (though by no means obviously the case) that if people’s judgements in these Williams-style thought experiments are the product of cognitive biases, then those biases might play a role in our judgments about other PI thought experiments. However, since other PI thought experiments have not been shown to elicit incompatible judgements, the best *reason* we have to posit the presence of cognitive biases at all, are Williams-style cases. If it turns out that people do not, for instance, make incompatible judgements in Williams-style cases, then not only do we have little reason to think that judgements in these cases are the product of cognitive biases, but we have no reason to think that judgements with regard to other PI though experiments are the products of such biases.

Thus, the present paper can been seen as mainly engaging in the *negative* project of experimental philosophy (Ludwig 2016: 386), where the use of empirical methods to analyse the patterns of how people respond to thought experiments aims at determine whether, and to which extent, philosophers could rely on thought experiments in their theorizing.

To determine whether our judgements about Williams’ cases are indeed subject to cognitive biases we need to ask two kinds of questions. First, do we, as Williams supposes, have incompatible judgements? Second, if we do, what explains our making those judgements? We empirically explore both questions.

We begin, (§2) by outlining relevant extant empirical research. We then formulate several more detailed hypotheses. In §3 we describe our methodologies for two experiments we ran. In §4 we discuss the upshot of these results for scepticism about PI thought-experiments.

1. Existing Research

The idea that people’s judgements are not stable across thought experiments is due to Williams (1970). We can distinguish two predictions made by Williams. First, we need to distinguish two kinds of *cases*, which we call *the psychological frame[[6]](#footnote-7)*  and *the bodily frame*. *[[7]](#footnote-8)*

*Psychological frame:* a case in which a person’s brain, along with their distinctive psychology (i.e., memories, intentions, values, character traits, etc.), is transferred from one body into a qualitatively identical new body, while the original body is destroyed.

B*odily frame:* a case in which someone’s brain is wiped of its distinctive psychology (i.e., memories, intentions, values, character traits, etc.) to prevent infection, after which the body will be subjected to several painful injections.

Then we need to distinguish four experimental conditions, each of which represents one way of manipulating a particular variable. These are:

*Psychological success condition*: a case where psychological continuity is preserved

*Psychological failure condition*: psychological continuity is *not* preserved

*Bodily success condition*: a case where bodily continuity is preserved.

*Bodily failure condition*: bodily continuity is *not* preserved.

With these in mind, Williams makes two predictions.

P1: That in the psychological frame people will judge that the individual survives in the psychological success condition but not in the psychological failure condition.

P2: That in the bodily frame people will judge that the individual survives in the bodily success condition but not in the bodily failure condition.

Some of these predictions appear already to have been confirmed. Blok, Newman, and Rips (2005) found evidence for P1. They found that in the psychological success condition people judge that the person survives, while they judge that they person does not survive in the psychological failure condition. Nichols and Bruno (2010) replicated this result, although it is worth noting that on a 10-point Likert scale (that ran from 0-9) the mean response was 5.95 when psychological continuity was preserved, and 4.09 when it was not. Although this difference is statistically significant, it does not show that people *overwhelmingly* judged that the individual survived in the psychological success condition and failed to survive in the lack of psychological failure condition. Nichols and Bruno (2010) found some evidence for P2. They found that in the bodily frame people tended to judge that the individual survived in the bodily success condition, although they did not test the bodily failure condition.

Hence there is some evidence that Williams’ predictions are vindicated. This is sufficient to suggest that people’s judgement are incompatible across these cases. In the psychological failure condition people judge that the individual does not survive, thus suggesting that psychological continuity is necessary for survival, but in the bodily success condition people judge that the individual survives despite there being no psychological continuity, hence suggesting that psychological continuity is not necessary for survival.

Our aim in what follows is twofold. First, we want to re-test P1, and to fully test P2. Let us call Williams’ predictions, jointly, the *incompatibility prediction.*

*The Incompatibility Prediction*: P1 and P2 are both true.

Second, on the assumption that the incompatibility prediction is vindicated we want to investigate the explanation of these incompatible judgements.

There are broadly two hypotheses about what could explain these incompatible judgements. It could be people’s judgements are sensitive to certain *irrelevant* features of the presentation of the cases, and hence are due to cognitive bias. This is the *cognitive bias hypothesis*.

*The Cognitive Bias Hypothesis*: what explains why both P1 and P2 are true is that our judgements about these cases are the product of cognitive bias.

What do we mean when we talk of relevant/irrelevant features? As Meier (2022) notes, there are a great many ways in which features of a thought experiment might be relevant to people’s judgment about that case. Indeed, one say quite generally that a feature is relevant insofar as people’s judgements are sensitive to the presence or absence of that feature. This is not how we are using the term ‘relevant’. We suppose that a feature is *relevant* if there is some plausible account of why attending to that feature might matter in determining whether someone survives. So, both bodily and psychological continuity (or lack thereof) are relevant (even though psychological continuity theorists will not be sensitive to the presence or absence of bodily continuity); what experts say about survival is relevant, and so on, whereas the colour of the clothes the person-stages are wearing, and which order we describe the relationships between person-stages is not relevant. These latter are not relevant even if as a matter of fact people’s judgments are sensitive to them: if they are sensitive to such features then it seems plausible that those judgements may be the product of cognitive bias.

Alternatively, it could be that these judgements are sensitive to certain *relevant* features of the presentation of the cases, and hence are not due to cognitive bias. This is the *relevant information hypothesis.*

*The Relevant Information Hypothesis:* what explains why both P1 and P2 are true is that our judgements about these cases are the product of relevant information which varies across the cases.

There are then various versions of each hypothesis, depending on which features one supposes it is, to which our judgements are sensitive, and depending on whether one thinks those features are relevant or irrelevant. We explore a few of these more specific hypotheses about what these features might be.

Williams himself hypothesised that what drives the difference between our judgments in these two thought experiments is a difference between the first and third-person presentation of the cases. In his original formulation the psychological frame case is presented in the third person, and the bodily frame case is presented in the first person. This suggests that the cognitive bias in question might be the product of a framing effect. Framing effects occur when our judgement about some matter is influenced by the way in which the relevant facts (or questions) are presented and where our judgements should not be sensitive to differences in presentation/framing. That is, our judgements are tracking irrelevant information. Empirical evidence shows that our judgements across a range of areas are subject to framing effects.[[8]](#footnote-9) For instance, people are inclined to prefer a medical procedure that is described as having a 90% survival rate, over one that is described as having a 10% death rate, even though these are just two ways of describing the same set of facts about mortality. (Tversky & Kahneman (1981); Kahneman 2011: 367).

Williams suggested that the framing lies in presenting the cases in terms of first-person versus third-person, a detail that should be irrelevant to our judgements. Nichols and Bruno, however, found the same pattern of judgements even when the psychological and bodily frames are both presented in the same person (both first person and both third person). So this cannot be what explains our judgements, and we set aside this hypothesis

The first hypothesis we consider is the *pain* *difference-maker hypothesis*. This is the hypothesis that people’s judgements are sensitive to the presence of the painful shots in the bodily frame and their absence in the psychological frame.

*Pain Difference-Maker Hypothesis*: People’s judgements are sensitive to the presence of the painful shots in the bodily frame, and their absence in the psychological frame.

According to this hypothesis, at least part of the reason why people judge that the individual survives in the bodily success condition but not the psychological failure condition is because the former case features the presence of pain. We test this hypothesis in experiment 1.

Suppose that the pain-difference maker hypothesis is vindicated. What does that tell us about whether we have evidence in favour of the cognitive bias hypothesis or the relevant information hypothesis? Well, very little, since it tells us nothing about *why* people are sensitive to the presence or absence of the pain. Thus, in our second experiment we tested three further hypotheses, two of which are versions of the relevant information hypothesis.

The first of these is *the minimal psychological continuity hypothesis.* This hypothesis is hinted at by Nichols and Bruno(2010: 301). They asked bodily frame participants to explain why they agreed or disagreed with the statement ‘you will feel the pain’. In their self-reports, some explanations cited the perceived continuity or discontinuity of ‘experiences’/‘sense of feeling’. Nichols and Bruno’s idea is that in drawing attention to the fact that the individual post-operation will undergo pain the vignette implicitly suggests that there is some continuity of consciousness even if there is no continuity of memory or beliefs.

Even if the pain-difference hypothesis is not vindicated, however, the minimal psychological continuity hypothesis could play a role in explaining people’s judgements. It could be that people take some sort of minimal continuity to be sufficient for personal-persistence, and judge minimal continuity to be present in both the bodily and psychological success conditions, but not the psychological failure condition.

*Minimal Psychological Continuity Hypothesis*: People’s judgements are sensitive to whether they believe that minimal psychological continuity is present or not.

We take this to be a version of the relevant information hypothesis, because whether or not minimal continuity obtains (given that one thinks it sufficient for survival) is surely relevant.

The second hypothesis we consider is *the deference to experts hypothesis,* which is the hypothesis that people’s judgements are sensitive to what the experts in each vignette are represented (perhaps tacitly) as judging regarding whether psychological or bodily continuity are necessary or sufficient for survival. According to this hypothesis, part of the reason why people judge that the person survives in both the psychological and bodily success conditions is that in the former the experts—the doctors—are tacitly represented as judging that the presence of psychological continuity is sufficient for survival, and in the latter they are tacitly represented as judging that the presence of bodily continuity is sufficient for survival. (After all, in each case the doctors tell the patient that his/her ‘only chance for survival’ is the relevant procedure). This would also explain why people judge that the person does not survive in the psychological and bodily failure conditions, assuming that in fact this is what we find.

*Deference to Experts Hypothesis*: People’s judgements are sensitive to the beliefs of experts about what survival consists in.

This is also a version of the relevant information hypothesis. After all, facts about what the experts think about survival are relevant when making judgements.

Our third hypothesis is the *conceptual resources hypothesis*, inspired, again, by a suggestion of Nichols and Bruno (2010: 304). They suggest that the presence of the pain in the bodily frame makes a difference to our judgements because we imagine the post-operative person being subject to the pain, and we (rightly) conclude that *someone* is in pain. Due to limitations in our conceptual resources, when we ask ourselves whose pain it is, the only answer we can find is that it must be the pain of the same person who existed pre-operation.

Again, while this hypothesis is motivated by supposing the pain-difference hypothesis to be vindicated, the conceptual resources hypothesis could explain people’s tendency to judge that they survive in both the psychological and bodily success conditions because in each case people lack the conceptual resources to imagine that some ‘new’ person has suddenly come into existence. Since in both conditions it is clear that some person or other exists after the operation, they judge that in both cases the person survives. What remains unclear is how this could explain people’s tendency to judge that the individual does not survive in the psychological failure condition, assuming people do in fact have this tendency. Nevertheless, we take it to be worth testing.

*Conceptual Resources Hypothesis:* People’s judgements are sensitive to whether they have the conceptual resources to imagine that a ‘new’ person could suddenly come into existence.

This is not a version of the relevant information hypothesis. But nor is it really a version of the cognitive bias hypothesis. That is because if this hypothesis is right, then it does suggest that people’s judgements in these cases are not to be trusted, but it does not show that they are subject to some sort of pervasive cognitive *bias*; nor does it show that, given appropriate conceptual resources, their judgements would fail to be reliable.

We begin, in experiment 1, by attempting to replicate Nichols and Bruno’s findings, and determining whether there is any support for the pain-difference hypothesis—i.e., that the presence or absence of pain partly explains why subjects judge that the individual survives in the bodily success condition where there is psychological continuity, but not in the psychological failure condition where psychological continuity is not preserved.

**3 Methodology and Results**

**3.1 Experiment 1 Methodology**

*3.1.1 Participants*

1517 people participated in the study. Participants were U.S. residents, recruited and tested online using Amazon Mechanical Turk, and compensated $0.85 for approximately 7 minutes of their time. Given recent worries about the quality of data collected through MTurk, concerning both the quality of human responders and the presence of bots, we adopted a number of quality control measures.[[9]](#footnote-10) First, we used only those MTurk participants who have a HIT (task) approval rate of at least 95% and who have had their HITs (tasks) approved at least 1000 times. That means that all our participants had already successfully completed at least 1000 other studies, and received at least a 95% approval rating on these tasks, a standard that can be expected to eliminate most bots. Second, our study included both task instructions and attentional checks that doubled as comprehension checks. 784 participants had to be excluded for failing to follow task instructions. This means that they failed to answer the questions (165) or failed one of the attentional check/comprehension questions (619). The remaining sample was composed of 733 participants (303 female; 2 trans/non-binary; aged 20-83; mean age 40.30 (SD = 11.44)). Ethics approval for these studies was obtained from the [blanked] Human Research Ethics Committee. Informed consent was obtained from all participants prior to testing. The survey was conducted online using Qualtrics.

*3.1.2* *Materials and Procedure*

The study was a 2x2x2 between-participants design. Participants were randomly assigned to one of eight conditions. These eight conditions reflected all possible combinations of frame (psychological/bodily), outcome of the procedure (success/failure) and presence of painful shots (pain/no-pain) Thus we have the following 8 conditions: Psychological No-pain Success; Psychological Pain Success; Psychological No-Pain Failure; Psychological Pain Failure; Bodily Pain Success; Bodily No-Pain Success; Bodily Pain Failure; Bodily No-Pain Failure.

Since Nichols and Bruno have shown that first-person/third-person framing makes no significant difference, we used first-person presentations throughout the experiments. Below are the two psychological and bodily frame vignettes in both the success/failure and pain/no-pain variants.

Text, letter

Description automatically generated

Text, letter

Description automatically generated

After reading the vignette participants responded to two attentional check/comprehension questions. The first asked “In this vignette, you were asked to imagine that you …”, to which participants could respond:

1. Had a brain infection
2. Had a body infection
3. Had a severe accident
4. Had cosmetic surgery

The second asked: “the procedure the doctors are discussing is expected to….” to which participants could respond:

1. Eliminate all of your personality traits, beliefs, values, and memories.
2. Eliminate none of your personality traits, beliefs, values, and memories.
3. Eliminate your personality traits and beliefs, but not your memories.
4. Eliminate your memories but not your beliefs or personality traits.

Participants who failed correctly to answer these questions were excluded from the study.

We next probed whether participants judged that they survived the procedure.

Participants in the no-pain conditions saw the following question:

“Do you survive the procedure?”

Participants in the pain conditions saw the following question:

“When the doctors administer the series of shots after the operation, will you feel the pain?”

And were given a forced choice response between “Yes” or “No”.[[10]](#footnote-11)

To vindicate Nichols and Bruno’s original findings, we should find that:

H1: Participants in the Psychological No-Pain Success and Bodily Pain Success conditions will judge that they survived the procedure.

In addition, we hypothesized that:

H2: Participants in the Psychological No-Pain Success condition will judge that they survive the procedure, while those in the Psychological No-Pain Failure condition will judge that they do not survive the procedure.

H3: Participants in the Bodily Pain Success condition will judge that they survive the procedure while those in the Bodily Pain Failure condition will judge that they do not.

H2 replicates the findings of Blok, Newman, and Rips (2005), and H3 extends these findings to the Bodily Pain cases.

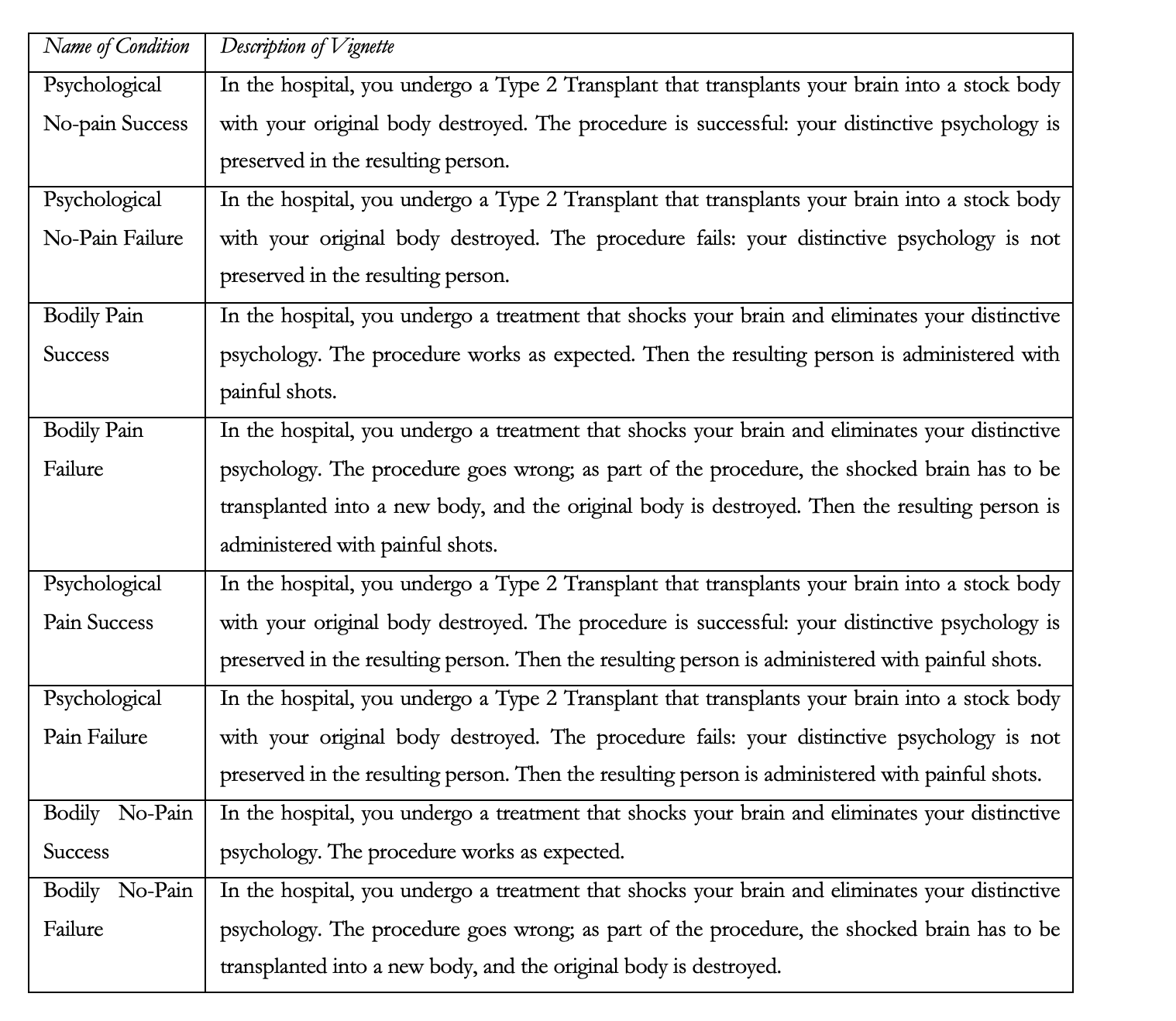
If the incompatibility prediction is correct, then all of H1, H2 and H3 should be vindicated.

If the pain difference-maker hypothesis is correct, then we should find that

H4: Participants in the Psychological Success conditions will be more inclined to judge that they survive the procedure in the Pain condition than in the No-Pain condition, and people in the Bodily Success conditions will be more inclined to judge that they survive the procedure in the Pain condition than in the No-Pain condition.

*3.1.3 Results*

Before reporting the specific test-statistics and details, let’s begin with a summary of our main findings. First, as a handy guide, below is a table that lists the 8 conditions that we tested, and reminds us of what was tested in that condition.



Consider, first, the three hypotheses that constitute the incompatibility prediction. H1 was vindicated. We found that people judged that they survived in both the Psychological No-pain Success condition and the Bodily Pain Success condition, replicating the original finding of Nichols and Bruno. H2 and H3 were not vindicated. While we found that more people judged that they survived in the Psychological No-Pain Success condition, compared to the failure condition (which shows that success/failure has an impact on people’s judgements) we found that a majority of people judged that they survived in both conditions. With regard to H3, not only did we find that a majority of people judged that they survived in both the Bodily Pain Success and Failure conditions, but there was no significant difference between the success and failure conditions.

Last, we hypothesised that participants in the Psychological Success conditions will be more inclined to judge that they survive in the Pain Condition than in the No-Pain condition, and participants in the Bodily Success conditions will be more inclined to judge that they survive in the Pain condition than in the No-Pain condition. Neither component of this hypothesis was vindicated. Instead, we found that more people judged that they would survive in Psychological/Bodily Pain conditions (relative to No-pain conditions) only when the procedure *fails*. That is, as suggested earlier, the presence of pain might counteract the effect that procedure failure has on some people’s judgments (keeping in mind the majority think they survive in all cases).

Table 1 summarises the descriptive results for survival judgments. The %Yes column represents the proportion of participants who judge that they would survive the procedure or would feel pain after the procedure (and so survived). The %No column represents the proportion of participants who judge that they would *not* survive the procedure or would *not* feel pain after the procedure (and so did *not* survive). Separate one-way Chi-squared tests were conducted for each condition and the results of those tests show that the significant majority of participants across all conditions tested judged that they would survive the procedure.

Table

Description automatically generated

To test whether there was any association between condition and survival judgment a Chi-square test of homogeneity was conducted. The result of that test showed that there was an association, χ2(7, *N* = 733) = 52.246, *p* < .001. While this test reveals that there is an association between conditions and survival judgments it does not reveal where the association is located. In order to examine that we performed a series of follow-up tests guided by our remaining hypotheses.[[11]](#footnote-12)

First, we looked at the effect of procedure success or failure on people’s survival judgments. Somewhat consistent with Nichols and Bruno we found an effect of procedure success and failure on people’s survival judgments in the Psychological No-pain conditions. While the majority judged that they survived in both the success and failure conditions, significantly more people did so in the success variant, χ2(1, *N* = 165) = 19.129, *p* < .001. However, we found no such evidence of an effect between success and failure in the Bodily Pain conditions, χ2(1, *N* = 207) = 0.019, *p* = .892.

Interestingly, we found the same pattern when we looked at the difference between success and failure in the Psychological Pain and Bodily No-pain conditions. That is, we found no evidence of an effect of success/failure in the Psychological Pain conditions, χ2(1, *N* = 163) = 1.451, *p* = .228. But we did find an effect of success/failure in the Bodily No-pain conditions, χ2(1, *N* = 198) = 23.359, *p* <.001.

Second, holding fixed procedure success, we looked at the effect of pain on people’s survival judgments. Contrary to our prediction we found no evidence of a difference between the Psychological No-pain condition and Psychological Pain condition, χ2(1, *N* = 176) = 7.173, *p* = .007. Likewise, contrary to our prediction we found no evidence of a difference between the Bodily Pain condition and Bodily No-pain condition, χ2(1, *N* = 210) = 2.835, *p* = .092. That is, there is no evidence that pain is associated with more people judging that they survive in success conditions.

Interestingly, this was not the case when we looked at failure conditions. More people judge that they survive the Psychological Pain condition than the Psychological No-pain condition when the procedure fails, χ2(1, *N* = 152) = 8.650, *p* = .003. Likewise, more people judge that they survive the Bodily Pain condition than the Bodily No-pain condition, χ2(1, *N* = 195) = 12.089, *p* < .001. Thus, it appears as though the presence of pain counteracts the effect that procedure failure would ordinarily have on some people’s judgments.

**3.2 Experiment 2(a) (b) and (c) Methodology**

*3.2.1 Participants*

2311 people participated in the study. Participants were U.S. residents, recruited and tested online using Amazon Mechanical Turk, and compensated $0.85 for approximately 7 minutes of their time. Again, we used only those MTurk participants who have a HIT (task) approval rate of at least 95% and who have had their HITs (tasks) approved at least 1000 times and included both task instructions and attentional checks that doubled as comprehension checks. 1202 participants had to be excluded for failing to follow task instructions. This means that they failed to answer the questions (196) or failed one of the attentional check/comprehension questions (1006). The remaining sample was composed of 1109 participants (526 female; 5 trans/non-binary; aged 19-89, mean age 40.84 (SD = 12.57)). Ethics approval for these studies was obtained from the [blanked] Human Research Ethics Committee. Informed consent was obtained from all participants prior to testing. The survey was conducted online using Qualtrics.

*3.2.2* *Materials and Procedure*

Experiment 2(a) tested the minimal psychological continuity hypothesis; experiment 2(b) tested the conceptual resources hypothesis and experiment 2(c) tested the deference to experts hypothesis.

In experiment 1, the majority of participants responded that they survived even in failure conditions where neither psychological continuity nor bodily continuity obtained, which seems to indicate that survival judgements are sensitive to something other than psychological/bodily continuity. What else could survive consist in? A salient candidate, as Nichols and Bruno suggested, is the continuity of conscious experience. If this is correct, then it would be no surprise that positive survival judgements prevailed – none of our vignettes in experiment 1 explicitly stated that the stream of consciousness was terminated. We hypothesize that minimal psychological continuity is sufficient for survival. Moreover, the minimal psychological continuity hypothesis seems to neatly account for the finding that the presence of pain was associated with positive survival judgements in failure conditions but not in success conditions: The presence of pain might prompt people to perceive a continuing stream of consciousness; and this only made a difference in failure conditions because in success conditions, psychological or bodily continuity was much more salient than minimal psychological continuity.

In experiment 2(a) participants were split into one of four conditions in which they saw minimally modified variants of the Bodily Pain Success, Bodily Pain Failure, Psychological No-pain Success, and Psychological No-pain Failure vignettes. We have bolded the additions to the amended sentences, below. The Bodily Pain vignettes were amended as follows:

“Unfortunately, this procedure will permanently eliminate your distinctive mental states including your thoughts, memories, personality traits, values, likes and dislikes, **and it will result in there being *no* stream of consciousness that connects your mental states before the procedure, and any mental states after the procedure.”**

The failure condition was amended as follows:

“Something goes wrong. The doctors succeed in shocking your brain as expected, and in doing so they wipe all of your memories, beliefs, personality traits, **and there is *no* stream of consciousness that connects your mental states before the procedure, and any mental states after the procedure.”**

We call the resulting two conditions Bodily (Pain) *Discontinuity* Success, Bodily (Pain) *Discontinuity* Failure.[[12]](#footnote-13)

Participants then saw two comprehension questions. The first was the same as in experiment 1. The second asked “the procedure the doctors are discussing is expected to….” to which participants could respond:

(a) Eliminate none of your personality traits, beliefs, values, and memories, and preserve your stream of consciousness.

(b) Eliminate none of your personality traits, beliefs, values, and memories, but terminate your stream of consciousness.

 (c) Eliminate all of your personality traits, beliefs, values, and memories, but preserve your stream of consciousness.

(d) Eliminate all of your personality traits, beliefs, values, and memories, and terminate your stream of consciousness.

Participants who passed the comprehension check then responded to the following question by giving a forced choice response between “Yes” or “No”.

“When the doctors administer the series of shots after the operation, will you feel the pain?”

The Psychological No-pain conditions were amended in the following manner:

“The transplant will then preserve all of the memories, values, and beliefs of your brain before the transplant **and there will be a stream of consciousness that connects your mental states before the procedure, and any mental states after the procedure.”[[13]](#footnote-14)**

The success condition was amended as follows:

“Fortunately, the procedure is successful and when the doctors scan the brain of the transplant recipient they note that all of your memories, values, and beliefs are the same as they were in the brain before the operation, **and there is a stream of consciousness that connects your mental states before the procedure, and any mental states after the procedure**.”

Accordingly, the failure condition was amended as follows:

“Unfortunately, something goes wrong in the procedure, and when the doctors scan the brain of the transplant recipient they note that none of your distinctive mental states including your thoughts, beliefs, memories, personality traits, values, likes and dislikes are the same as they were in the brain before the operation **and there is *no* stream of consciousness that connects your mental states before the procedure, and any mental states after the procedure.”**

We call the former the Psychological (No-pain) *Discontinuity* Success condition, and the latter the Psychological (No-pain) *Discontinuity* Failure condition.

Participants then saw the same comprehension questions as in the Bodily Pain Discontinuity conditions.Participants who passed the comprehension check were then presented with the following probe question:

“Do you survive the operation?”

And were given a forced choice response between “Yes” or “No”.

If the minimal psychological continuity hypothesis is correct, then we should expect to find that:

H5: Participants in the Psychological Discontinuity Failure condition will judge that they do *not* survive the procedure.

H6: Participants in the Bodily Discontinuity Failure condition will judge that they do *not* survive the procedure.

By making these two hypotheses, we do not mean to be hypothesizing that minimal psychological continuity is necessary for survival. If these two hypotheses are vindicated, it may be the case that minimal psychological continuity is indeed necessary. But it may not be, for it may be that when either of the “thicker” continuity (psychological or bodily continuity) *is* present, the “thicker” continuity itself is sufficient for survival; it’s only when thicker continuities are absent that minimal psychological continuity is responsible for survival judgments.

Further, if in experiment 1 some people supposed that there was minimal continuity even in the Psychological and Bodily Failure conditions, and they took this to be sufficient for survival, then we should find the following:

H7: Participants in the Psychological Discontinuity Failure condition will be *less* inclined to judge that they survive the procedure than in the Psychological No-Pain Failure condition.

H8: Participants in the Bodily Discontinuity Failure condition will be *less* inclined to judge that they survive the procedure than in the Bodily Pain Failure condition.

One might also wonder whether minimal psychological continuity is not only sufficient, but also necessary for survival. If it is, then we should find:

H9: Participants in the Bodily Discontinuity Success condition will be *less* inclined to judge that they survive the procedure than in the Bodily Pain Success condition.

In experiment 2(b) we developed amended versions of the Bodily Pain Success and Failure conditions, and the Psychological No-pain Success and Failure conditions. These vignettes were amended to afford people certain conceptual resources that would allow them to make sense of the idea that the post-operative person is not the same person as the pre-operative person. We call these ‘nudging’ vignettes. The Nudging Bodily (Pain) Success/Failure vignettes are as follows:

Text, letter

Description automatically generated

After reading the vignette, participants responded to the same comprehension questions as in experiment 1. They were then presented with the following probe questions:

1. When the doctors administer the series of shots after the operation, will Computer Name feel the pain?
2. Are you Computer Name?

And were given a forced choice response between “Yes” or “No”.

Similarly, the Nudging Psychological (No-pain) Success/Failure vignettes added more conceptual resources to the original Psychological No-pain Success/Failure vignettes:

Text, letter

Description automatically generated

After reading the vignette, participants responded to the same comprehension questions as in experiment 1. They were then presented with the following probe questions:

(1) Does Computer Name wake up from the procedure?

(2) Are you Computer Name?

And were given a forced choice response between “Yes” or “No”.

According to the conceptual resources hypothesis, people judge that they survive in both the psychological and bodily success conditions because they lack the conceptual resources to imagine that some ‘new’ person has suddenly come into existence. Indeed, this could also explain why they tend to judge that the person survives even in the failure conditions. If in fact people are really psychological continuity theorists, but their lack of conceptual resources is impeding their judgements, then we should find that a majority of participants will judge that they survive in the Nudging Psychological Success condition, but not the Nudging Bodily Success condition. We should find the reverse if in fact people are really bodily continuity theorists. Since this is precisely what is at issue, we made no predictions in this regard, but it is worth bearing this in mind in what follows.

However, if the conceptual resources hypothesis is correct we should find that people are less inclined to judge that the person survives in the Nudging Failure conditions compared to the Failure conditions:

H10: Participants in the Nudging Psychological Failure condition will be more inclined to judge that they do *not* survive the procedure than in the Psychological No-pain Failure condition.

H11: Participants in the Nudging Bodily Failure condition will be more inclined to judge that they do *not* survive the procedure than in the Bodily Pain Failure condition.

In experiment 2(c) we tested the deference to experts hypothesis, according to which participants’ judgements are sensitive to the attitudes of the experts in the cases. Participants were assigned to one of four conditions: Scientist Psychological (No-Pain) Success, Scientist Psychological (No-pain)[[14]](#footnote-15) Failure, Scientist Bodily (Pain) Success and Scientist Bodily (Pain) [[15]](#footnote-16) Failure. These vignettes removed mention of doctors, who might be expected to be both experts on survival and to only perform procedures that they think are consistent with the possibility of survival. Instead we introduce scientists, who have no interest in whether the person survives or not.

Scientist Psychological No-Pain Success/Failure

Text, letter

Description automatically generated

Scientist Bodily Pain Success/Failure

Text, letter

Description automatically generated

After reading the vignette, participants responded to two attentional check/comprehension questions. The first is the same as the previous experiment. The second asked “The procedure the scientists performed...” to which participants could respond:

1. Eliminated all of your personality traits, beliefs, values, and memories.
2. Eliminate none of your personality traits, beliefs, values, and memories.
3. Eliminate your personality traits, beliefs, and values, but not your memories.
4. Eliminate your memories, but not your personality traits, beliefs, and values.

Participants were then presented with one of the following two questions depending on the condition to which they were assigned:

(1) Do you survive the procedure? (Psychological Conditions)

(2) When the doctors administer the series of shots after the operation, will you feel the pain? (Bodily Conditions)

And were given a forced choice response between “Yes” or “No”.

If the deference to experts hypothesis is correct, then we should find that:

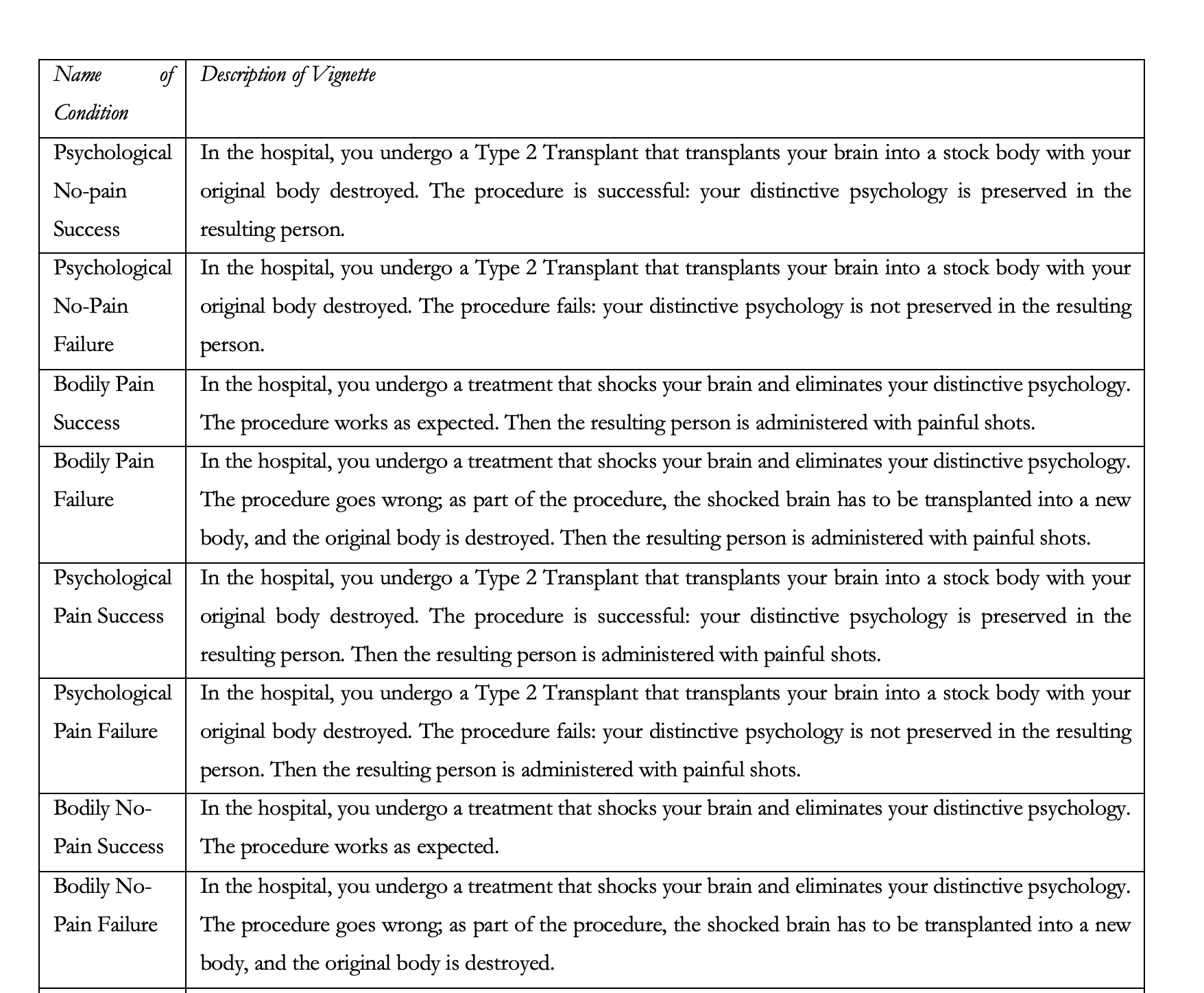
H12: Participants in the Scientist Psychological Success condition will be more inclined to judge that they do *not* survive the procedure than in the Psychological No-pain Success condition.

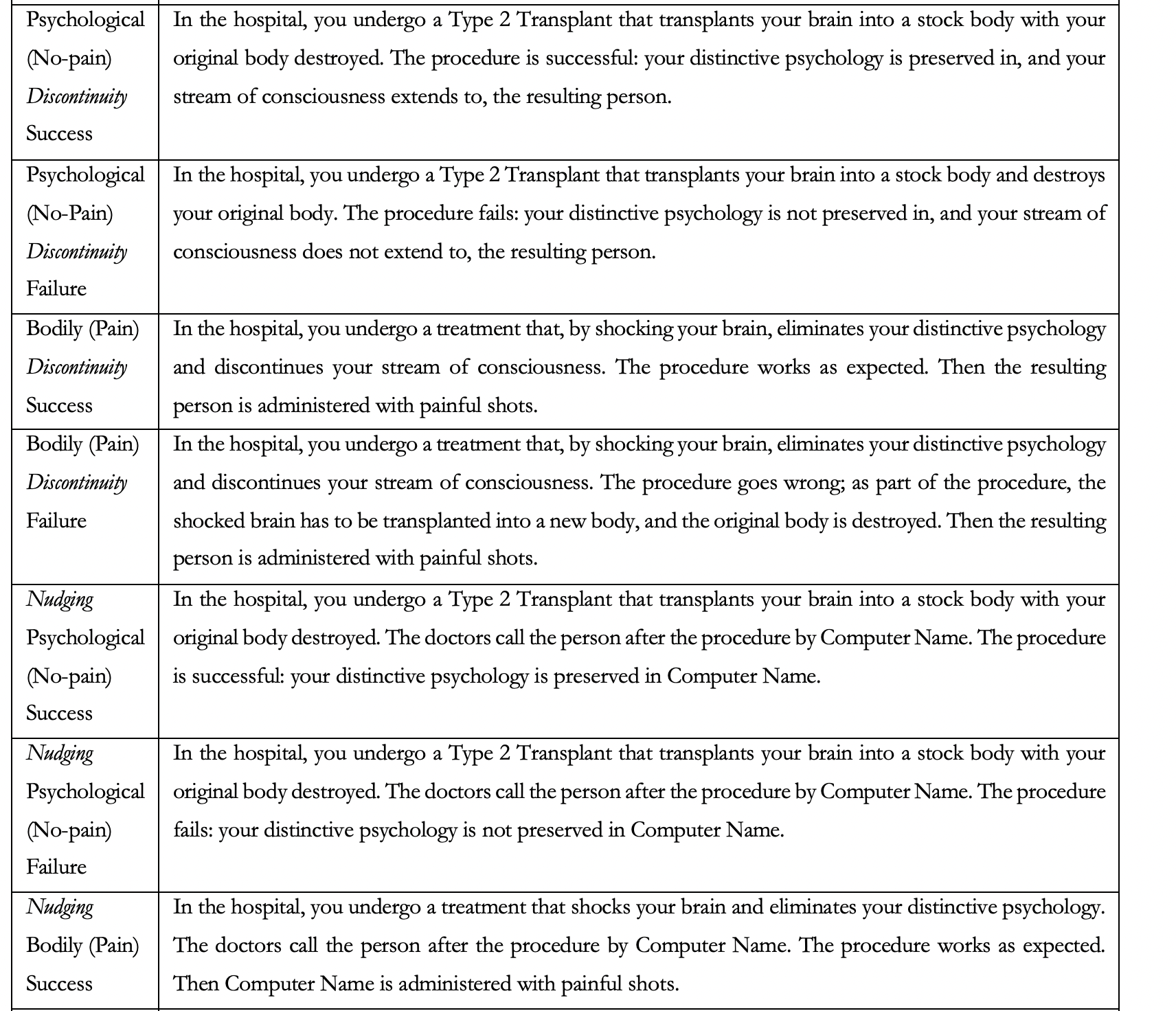
H13: Participants in the Scientist Bodily Success condition will be more inclined to judge that they do *not* survive the procedure than in the Bodily Pain Success condition.

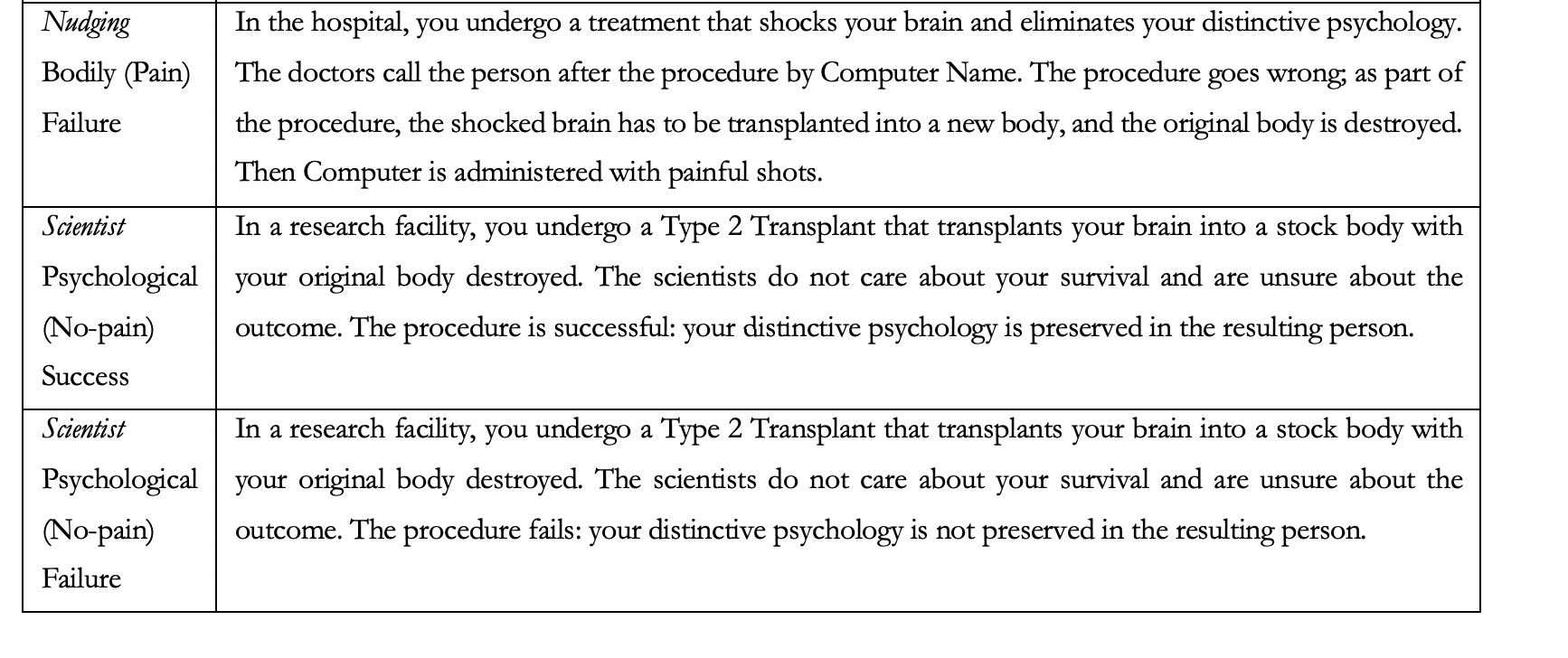
*3.5.3 Results*

Before reporting the specific test-statistics and details, let’s begin with a summary of our main findings.

Before we do so, here is a reminder about the conditions we are testing in these two experiments.







In experiment 2(a) we predicted that participants would judge that they do not survive the procedure in the Psychological Discontinuity Failure condition (H5) or the Bodily Discontinuity Failure condition (H6). We found that the majority of participants across *all* Discontinuity conditions judged that they survive the procedure.

We also predicted that participants would be less inclined to judge that they survive the procedure in the Psychological Discontinuity Failure condition than in the Psychological No-pain Failure condition (H7), and in the Bodily Pain Discontinuity Failure condition than in the Bodily Pain Failure condition (H8). We also wondered whether less participants would judge that they survive the procedure in the Bodily Pain Discontinuity Success condition than in the Bodily Pain Success condition (H9).

Once again, none of these hypotheses were vindicated. There were no differences in participant’s survival judgments between the Discontinuity conditions we tested and the original conditions.

In experiment 2(b) we predicted that participants would be less inclined to judge that they survive in the Nudging Psychological Failure condition than in the Psychological No-pain Failure condition (H10), and in the Bodily Nudging Failure condition than in the Bodily Pain Failure condition (H11). While H10 was not vindicated (there was no evidence of a difference), H11 was vindicated. Fewer participants judged that they would survive in the Bodily Nudging Failure condition than in the Bodily Pain Failure condition.

Finally, in experiment 2(c) we predicted that participants would be less inclined to judge that they survive in the Scientist Psychological Success condition than in the Psychological No-pain Success condition (H12), and in the Scientist Bodily Success condition than in the Bodily Pain Success condition (H13). Neither hypothesis was vindicated. There were no differences between the Scientist conditions and the original conditions.

Table 2 below summarises the descriptive results for survival judgments for experiment 2(a). Once again, the %Yes column represents the proportion of participants who judge that they would survive the procedure or would feel pain after the procedure (and so survived). The %No column represents the proportion of participants who judge that they would *not* survive the procedure or would *not* feel pain after the procedure (and so did *not* survive). Separate one-way Chi-squared tests were conducted for each condition and the results of those tests show that the significant majority of participants across all conditions tested judged that they would survive the procedure.

Table

Description automatically generated

To test whether there was any association between condition and survival judgment a Chi-square test of homogeneity was conducted. Because we were interested in testing whether there was any effect of Discontinuity on participant’s survival judgements, we also included data from the original Psychological and Bodily Frame conditions. The result of that test showed that there was an association, χ2(7, *N* = 747) = 67.578, *p* < .001. To reveal where the association is located, we performed a series of follow-up tests guided by our hypotheses.[[16]](#footnote-17)

First, we looked at the effect of procedure success or failure on people’s survival judgements in discontinuity conditions. While the majority of participants judged that they survived in both the success and failure variants, in the Psychological Frame, significantly more people did so in the success variant relative to the failure variant, χ2(1, *N* = 164) = 32.754, *p* < .001. There was no evidence of a difference between success and failure variants in the Bodily Frame, χ2(1, *N* = 211) = .249, *p* = .618.

Holding fixed procedure success or failure, we looked at the effect of Discontinuity on people’s survival judgements. We found no evidence of a difference between the Psychological Discontinuity Success condition and original Psychological No-pain Success condition, χ2(1, *N* = 177) = 1.447, *p* = .229, and between the Bodily Discontinuity Success condition and original Bodily Pain Success condition, χ2(1, *N* = 203) = .239, *p* = .625. Similarly, we found no evidence of a difference between the Psychological Discontinuity Failure condition and its original variant, χ2(1, *N* = 152) = .332, *p* = .565, and between the Bodily Pain Failure condition and its original variant, χ2(1, *N* = 215) = .020, *p* = .888.

Table 3 summarises the descriptive results for survival judgments for experiment 2(b). Separate one-way Chi-squared tests were conducted for each condition and the results of those tests show that the significant majority of participants across all conditions tested judged that they would survive the procedure.

Table

Description automatically generated

To test whether there was any association between condition and survival judgment a Chi-square test of homogeneity was conducted. Once again, we included data of the original Psychological and Bodily Frame conditions to see whether there was any effect of Nudging on survival judgments. The result of that test showed that there was an association, χ2(7, *N* = 764) = 35.226, *p* < .001. Follow-up tests were performed to identify where the significant association is located.[[17]](#footnote-18)

First, we looked at the effect of procedure success or failure on people’s survival judgements in Nudging Conditions. There was no evidence of a difference between success and failure variants in the Psychological Frame, χ2(1, *N* = 168) = 2.546, *p* < .111. Similarly, there was no evidence of a difference between success and failure variants in the Bodily Frame, χ2(1, *N* = 224) = 3.199, *p* = .074.

Holding fixed procedure success, we looked at the effect of Nudging on people’s survival judgements. We found no evidence of a difference between the Psychological Nudging Success condition and its original variant, χ2(1, *N* = 169) = 3.685, *p* = .055, and between the Bodily Nudging Success condition and its original variant, χ2(1, *N* = 222) = 1.719, *p* = .190. Similarly, we found no evidence of a difference between the Psychological Nudging Failure condition and its original variant, χ2(1, *N* = 164) = 1.181, *p* = .277. However, participants were significantly less inclined to judge that they would survive in the Bodily Nudging Failure condition than the original Bodily Pain Failure condition, χ2(1, *N* = 209) = 9.462, *p* = .002.

Finally, we were interested in exploring people’s judgments to being asked whether they are Computer Name. The majority of participants across all the Nudging conditions that we tested judged that they were Computer Name (62.5%; χ2(1, *N* = 392) = 24.500, *p* < .001). To test whether there was an association between judging that you are Computer Name and judging that you survive the procedure, we ran a chi-square test of independence. The result of that test showed that there was a significant association, χ2(1, *N* = 392) = 12.016, *p* < .001. Participants that judged they were Computer Name were significantly more inclined to judge that they would survive the procedure.

Table 4 summarises the descriptive results for survival judgments for experiment 2(c). Separate one-way Chi-squared tests were conducted for each condition and the results of those tests show once again that the significant majority of participants across all conditions tested judged that they would survive the procedure.

Table

Description automatically generated

To test whether there was any association between condition and survival judgment a Chi-square test of homogeneity was conducted, including data of the original Psychological and Bodily Frame conditions. The result of that test showed that there was an association, χ2(7, *N* = 714) = 39.031, *p* < .001. Follow-up tests were conducted to locate the source of the association.[[18]](#footnote-19)

First, we looked at the effect of procedure success or failure on people’s survival judgements in Scientist conditions. We found no evidence of a difference between success and failure variants in the Psychological Frame, χ2(1, *N* = 152) = 3.786, *p* .052. Similarly, we found no evidence of a difference between success and failure variants in the Bodily Frame, χ2(1, *N* = 190) = 1.387, *p* = .239.

Finally, holding fixed procedure success, we looked at the effect of Scientists on people’s survival judgements. We found no evidence of a difference between the Scientist Psychological Success condition and the original Psychological No-pain Success condition, χ2(1, *N* = 166) = 1.709, *p* = .191, and between the Scientist Bodily Success condition and the original Bodily Pain Success condition, χ2(1, *N* = 203) = 2.340, *p* = .126. Similarly, we found no evidence of a difference between the Scientist Psychological Failure condition and the original Psychological No-pain Failure condition, χ2(1, *N* = 151) = 1.442, *p* = .230, and between the Scientist Bodily Failure condition and the original Bodily Pain Failure condition, χ2(1, *N* = 194) = .039, *p* = .844.

4. Discussion

Before we consider the implications of these results, it is worth reflecting on some limitations of these studies. First, one might worry that the vignettes were *too* cognitively demanding for non-philosophers to understand, and hence the results may not indicate much about people’s judgments about personal identity. To address this concern, we included multiple attention checks and comprehension check questions. This served two purposes. First, it weeded out bots, and people selecting answers at random and without thought to quickly receive payment (Ahler, Roush, and Sood 2020). This is something that needs to be especially guarded against when running online studies and can result in large numbers of participant exclusions. Still, it is worth noting that the numbers of participants excluded by our studies are roughly comparable to those of Bruno and Nichols (2010). When they ran a single comprehension check question on students in an introductory philosophy class at the University of Arizona, they reported that 33% and 44% of students counted as failing to adequately understand the Pain frame thought experiments. Further, the numbers of participants excluded here are consistent with other recent online studies such as Latham, Miller, Tarsney and Tierney (2021a, 2021b). Second, the comprehension checks enabled us to achieve a sample composed of people who understood the vignettes. The numbers of people who comprehend each condition in our studies are often more than double the numbers reported by Nichols and Bruno (2010), and Blok, Newman, and Rips (2005), both of whom report significant personal identity findings. Thus, our studies are more than adequately powered to detect the effects which they report.

Still even if the current results are informative, one might worry that they are only *narrowly* so because the remaining sample is not representative of the general population. This is possible and something to bear in mind when evaluating the results, but we see little reason to suppose that’s true. *Perhaps* people that pass all the comprehension questions are more thoughtful, reflective, or intelligent, than those who did not. But even if it is so (and there is no reason to think it is) there is no reason to think that such people are less subject to cognitive biases than those who did not. In fact, Bruno and Nichols (2010) found the exact same patterning of results to the Pain frame thought experiments between the full sample and just those who passed a comprehension check. Most importantly though, as we will soon discuss, the current results, while different in certain respects, replicate some of the key previous empirical findings on people’s personal identity judgments. The fact that this was done for the first time with a sample *not* composed of undergraduate students undermines the thought that this sample and results is (especially) unrepresentative.

With this in mind, let’s first consider the incompatibility prediction. That prediction was not confirmed. People do not both judge that they *do* survive in the psychological continuity success condition, but *not* the failure condition (suggesting that they take psychological continuity to be necessary and sufficient for survival) while *also* judging that they *do* survive in the bodily success condition but *not* in the bodily failure condition (suggesting that they think that psychological continuity is neither necessary nor sufficient for survival). Rather, across all four case we find that a majority of people judge that they survive.

This result is somewhat different from that of Nichols and Bruno (2010) and Blok, Newman, and Rips (2005). Of course, none of those studies investigated the Bodily Failure condition. And our results are consistent with both studies with regard to the Bodily Success and Psychological Success conditions. So, where our results differ, lies with the Psychological Failure condition. Like both of these studies (and as we discuss shortly) we found that fewer people judged that they survived in the Psychological Failure condition compared to the Success condition. Unlike these studies, however, we found that a majority of people judge that they survived in both conditions. This is perhaps not as surprising as one might think. As we noted earlier in the paper, Nichols and Bruno found that the mean response in the Psychological Failure condition was over 4, when the midpoint of their Likert scale was 4.5. Hence either most, or all, of the participants in the Psychological Failure condition were only weakly of the view that they did not survive in that condition, or else people were split between those who thought they did survive, and those who thought they did not. Indeed, it could even be that a majority of people judged that they did survive (but made a weak such judgement) while a minority were strongly of the view that they did not, consistent with a mean of just over 4. Moreover, later in their study Nichols and Bruno explicitly asked participants which they thought was most important for survival, psychological or bodily continuity. Although Nichols and Bruno report that a majority of people said that it was psychological continuity, in fact their results show that the majority in question is not statistically significantly different from a 50/50 split. That is, there result is consistent with participants being evenly split between thinking that bodily continuity mattered and thinking that psychological continuity mattered. This fits well with our results.

Returning to the pattern of judgments we found, then, we can notice that this pattern is not inconsistent and does not, on its face, seem to call out for an explanation that appeals to cognitive bias. Rather, what our experiments show is that people have very robust responses to the scenarios presented. Across all the cases, people judge that the individual survives. This suggests, first, that people are neither psychological nor bodily continuity theorists, but rather, have some other view about the persistence of persons. For instance, perhaps people have a ‘featureless soul’ view, on which the persistence of persons consists in the persistence of a bare or featureless soul: [[19]](#footnote-20) that is, a soul that can continue to exist despite there being no psychological or bodily continuity. That would explain why people judge that the individual survives in both the psychological frame and the bodily frame, and in both the success and failure conditions.

Or perhaps people have some other very minimal notion of survival, on which psychological nor bodily are both sufficient, but neither is necessary, for survival.

Notably, either of these hypotheses is consistent with people being sensitive to the sorts of factors to which philosophers have attended in the debate about personal-identity. There are two ways in which this could be so.

First, although it might be that survival is an all or nothing affair, and that people have a very minimal notion of what is required for survival, that they nevertheless take there to be *better* and *worse* surviving, where perhaps this maps onto important ethical and prudential features of cases. Perhaps, for instance, someone who survives, but in a worse way, is less to blame for earlier actions or, prospectively, has fewer reasons to act in the interests of their future selves.

Second, perhaps in fact survival is not an all or nothing notion at all (something that Braddon-Mitchell and Miller 2020 suggest), but nevertheless people will judge that someone survives, or does not, (outright) when prompted, as long as the degree of survival in question is high/low enough. Then perhaps although people would judge that the individual survives to a *different* degree across the four cases, they nevertheless take the view that in all four cases there is a sufficiently high degree of survival to warrant judging that the person simply survives.

Notably, either of these hypotheses is consistent with some people being sensitive to whether psychological and/or bodily continuity is preserved or not. For it could be that some people take the presence/absence of these kinds of continuity to be *evidence* for the persistence of the same soul, or the existence of whatever minimal feature they suppose is required for survival, or is relevant to the degree to which the person survives.

In light of this, consider the fact that in experiment 1 and 2 (a) we found that failure in the psychological condition makes a significant difference to people’s judgements. Fewer people judge that they survive in the failure condition that in the success condition. This could be explained by a minority of people taking the *absence* of psychological continuity to be evidence that the ‘minimal thing that is necessary for survival’ (such as the soul) has ceased to persist, and hence evidence of a failure of personal persistence. What of bodily continuity? Well overall, we found that people’s judgements were no different in the bodily continuity *success* condition as compared to the *failure* condition. However, when we break this down we see that this is only true in the Pain condition. (Similarly, success/failure did not make a difference in the Psychological Pain condition in experiment 1.) In the No-pain condition we found that fewer people judged that they survived in the failure condition compared to the success condition. This suggests that the presence of pain plays a role in ‘undermining’ the effect of failure in the bodily condition. Setting aside people in the pain condition, then, it is tempting to conclude that participants take both bodily and psychological continuity to be evidence for the existence of the minimal thing necessary for survival.

Here is one suggestion we have regarding what all of this tells us about people’s judgements about survival and identity: people have multiple concepts they employ in making such judgements. It might be that people employ three concepts. First, they employ a basic concept of survival, which is really quite minimal (such as a concept that is satisfied by a featureless soul). We will call this *simple survival.* Second, they employ a concept that is more closely connected to the neo-Lockean conception of personal-identity. And perhaps this latter concept is connected to questions about moral responsibility and prudence. Something *like* this suggestion is made by Schechtman (1996) Olson (1997) and DeGrazia (2005). They recommend we distinguish between a metaphysical sense of personal persistence, and a practical sense that is relevant to the normative domain. We will call this *normative survival.* Then there is a concept that is more closely connected to bodily continuity; a concept that people use in ordinary life in order to track persons over time. Call this *practical survival.* We can call this *the three concepts hypothesis.*

Note that we are not hypothesizing that non-philosophers have in mind three sets of clearly defined necessary and sufficient conditions (or something close) corresponding to the three concepts. Rather, we take that what’s required for having a concept involves only minimal articulatable definitions when prompted, mental representations of the prototypical exemplifications of the concept, and relatively stable dispositions to apply the concept across actual and counterfactual scenarios. And unless there is reason to believe, for example, that our everyday PI-involving thoughts and language (such as judgments about responsibility and compensation resting on PI-judgments), though exhibiting relatively stable patterns, actually result from massive coincidence, it would be safe to assume that non-philosophers do have concepts of PI.[[20]](#footnote-21)

Suppose the three concepts hypothesis is correct. If so, it could explain both why most people judge that they survive in all four conditions: for in all such conditions they take there to be simple survival. It also explains why (setting aside the bodily pain condition) fewer people judge that they survive in the failure conditions than the success conditions. For we would expect some interaction between these concepts. We would expect people to be more confident that they survive when all three concepts are satisfied than when only one (or two) or them is. That is just what we find. In success conditions two such concepts are satisfied, whereas in failure conditions only one is.

Is there, then, any other reason to suppose that our judgements in these cases are subject to some kind of cognitive bias? We remain unsure why the presence of pain tended to undermine the effect of failure in the bodily condition. *Perhaps* this local effect is the product of cognitive bias. But even if it is, it is hardly reason to conclude that in general people’s judgements in *all* these cases are the product of cognitive bias.

What of our broader results then? Well, our results suggest that none of the further hypotheses we made, explains this pattern of results. We tested a number of possible features to which people could be sensitive in making the judgements they do. We did not find that people’s judgements are in fact sensitive to *any* of these factors except the success/failure of psychological continuity. This includes factors that are relevant (and hence not evidence of cognitive bias) such as whether there is some minimal level of psychological continuity present, and what sort of view of personal identity the relevant experts appear to endorse. It also includes factors that are not strictly relevant to question of personal identity, such as whether people have the conceptual resources available that would allow them to distinguish the person pre- and post- procedure, and hence which might generate performance errors (without being evidence of cognitive bias per se). Finally, it includes factors that, if explanatory, would seem to be evidence of cognitive bias, such as whether the scenario is described in first or third-personal terms, or whether there is the presence of pain or not.

Notably, our results (in concert with those of Nichols and Bruno when it comes to the first/third personal factor) did not find that *any* of these factors plays a role in people’s judgements.

Hence, we found no evidence that *relevant* factors explain people’s judgements. On the other hand, we also found no evidence that *irrelevant* factors explain people’s judgements either. As we see it, then, there is no reason, arising from these results, to think that these judgements are the product of cognitive bias. No such source of bias has been located. Further, the pattern of judgements is not itself evidence of some kind of bias, since that pattern is not inconsistent. What it suggests is simply that people are not either psychological or bodily continuity theorists, at least about minimal survival.

What implications does this have for theorising about PI using thought experiments? Well, we think it undermines the idea that there is some *special* problem for PI thought experiments that outstrips more general worries one might have about their role in theorising. We see no reason, based on our results, to be especially sceptical about the role of thought experiments when it comes to theorising about PI, for we see no evidence that our judgments in this arena are the product of special cognitive biases. Of course, this leaves very much open what role thought experiments *should* play. It is consistent with what we have said here that there are other more general reasons to be sceptical of thought experiments, and hence *a fortiori* of PI thought experiments. So we do not take ourselves to have vindicated the use of thought experiments here, or elsewhere: rather, we take ourselves to have gone some way towards undermining the idea that there is something especially problematic about the use of thought experiments in this domain of philosophy.

References

Ahler, D., Roush, C., & Sood, G. The micro-task market for lemons: data quality on Amazon’s Mechanical Turk. Unpublished manuscript, 22 January 2020. URL: <http://gsood.com/research/papers/turk.pdf>A

Alexander, J., & Weinberg, J. M. 2007. Analytic epistemology and experimental philosophy. *Philosophy compass*, *2*(1), 56-80.

Baz, Avner. 2018. *The Crisis of Method in Contemporary Analytic Philosophy.* Oxford: Oxford University Press.

Bengson, John. 2013. “Experimental Attacks on Intuitions and Answers.” *Philosophy and phenomenological research* 86, no. 3: 495–532.

Braddon Mitchell, David and Miller, Kristie (2020a). “Surviving, to a degree”. *Philosophical Studies* <https://doi.org/10.1007/s11098-019-01410-0>

Cullen, Simon. 2010. Survey-driven romanticism. *Review of philosophy and psychology,* 1(2), 275-296.

DeGrazia, D. (2005). *Human Identity and Bioethics.* Cambridge University Press

Gigerenzer, G. (2010). “Moral Satisficing: Rethinking Moral Behavior as Bounded Rationality.” *Topics in Cognitive Science* 2 (3):528-554

Hashimoto, Takaaki et al. 2018. “Gettier Was Framed.” In *Epistemology for the Rest of the World*. New York: Oxford University Press.

Herdova, Marcela. (2016). Nothing to Fear: Swap Cases and Personal Identity. *Analytic Philosophy*, *57*(4), 315–337.

Johnston, Mark. 1987. Human Beings. *Journal of Philosophy* 84 (2): 59-83.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow.* Macmillan.

Latham, A. J., Miller, K. Tarsney, C. and Tierney, H. (2021). “Belief in Robust Temporal Passage (Probably) Does Not Explain Future-Bias.” *Philosophical Studies.* 10.1007/s11098-021-01748-4

Latham, A. J. Miller, K., Tarsney, C and Tierney, H. (2021). “Robust Passage Phenomenology Probably Does Not Explain Future-Bias” *Synthese* 10.1007/s11229-022-03514-3

Ludwig, Kirk. (2010). Intuitions and relativity. Philosophical Psychology, 23(4), 427–445.

Ludwig, Kirk (2016). Thought Experiments in Experimental Philosophy. In Mike Stuart, James Robert Brown & Yiftach J. H. Fehige (eds.), *The Routledge Companion to Thought Experiments*. New York: Routledge. pp. 385-405.

Locke, John. 1689/1694. *An Essay Concerning Human Understanding*, (The Clarendon Edition of the Works of John Locke ), Peter H. Nidditch (ed.), Oxford: Oxford University Press, 1975.

Machery, Edouard. 2017. *Philosophy within its proper bounds.* Oxford: Oxford University Press. Martin, Raymond (1997). *Self-Concern: An Experiential Approach to What Matters in Survival*. Cambridge University Press.

Meier, Lukas J. (2022). Can Thought Experiments Solve Problems of Personal Identity? *Synthese* 200 (3):1-23.

Nozick, Robert (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

Nichols, S. & Bruno, M. 2010. Intuitions about personal identity: An empirical study, *Philosophical Psychology*, 23:3, 293-312.

Olson, Eric. (1997). *The Human Animal*. OUP

﻿Parfit, D. (1984). *Reasons and Persons.* Oxford: Clarendon Press.

Petrinovich, Lewis, and Patricia O'Neill. 1996. Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology* 17, no. 3: 145-171.

Reid, Thomas. 1785 [1851]. *Essays on the Intellectual Powers of Man*. Abridged and reprinted, James Walker (ed.), Cambridge, MA: John Bartlett, 1850; second edition, 1851.

Schechtman, M. 1996. The constitution of selves. Ithaca, N.Y.: Cornell University Press.

Tobia, Kevin, Wesley Buckwalter, and Stephen Stich. 2013. Moral intuitions: Are philosophers experts?. *Philosophical Psychology* 26, no. 5: 629-638

Tobia, Kevin P. 2015. Personal identity and the Phineas Gage effect. *Analysis* 75 (3):396-405.

Tversky, A., & Kahneman, D. 1989. Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers* (pp. 81-126). Springer, Berlin, Heidelberg.

Tversky & Kahneman (1981). The framing of decisions and the psychology of choice. *Science* 211: 453-458.

Unger, Peter. (1990). *Identity, consciousness, and value*. New York: Oxford University Press.

Williams, B. (1970). The self and the future. The Philosophical Review, 79, 161–180.

Wilkes, Kathleen V. (1988). *Real People: Personal Identity Without Thought Experiments*. Oxford University Press.

1. See for instance Locke’s “the prince and the cobbler” (1689/1894: 2.27.15) and Reid’s “the brave officer” (1785 [1851: 248–249]). More recent instances include Williams (1970); Parfit (1984: Part Three); Martin (1997). [↑](#footnote-ref-2)
2. Alexander and Weinberg 2007; Baz 2018; Bengson 2013; Cullen 2010; Machery 2017. [↑](#footnote-ref-3)
3. Although Williams does not himself draw this conclusion. Instead, he thinks that one set of intuitions is stronger than the other, and hence favours the bodily continuity theory over the psychological continuity theory. [↑](#footnote-ref-4)
4. Johnston (1987); Unger (1990: 88); Herdova (2016: 334-5); Wilkes (1988: ch.1); Meier (2022). [↑](#footnote-ref-5)
5. For instance, one might take the fact that people seem to disagree about personal persistence, as demonstrated by different judgments about teletransportation cases, as evidence either that thought experiments are poor ways to gather people’s judgements, or perhaps even that there are no facts of the matter regarding personal identity that we are tracking with our judgments (where this latter is really scepticism not about thought experiments, but about personal identity itself). Meier (2022) discusses various other reasons one might be sceptical of PI thought experiments, most of which could generalise into reasons to be sceptical of thought experiments in other domains. [↑](#footnote-ref-6)
6. Nichols and Bruno (2010) call this the Lockean Frame. [↑](#footnote-ref-7)
7. Nichols and Bruno (2010) call this the Pain Frame. [↑](#footnote-ref-8)
8. Tversky and Kahneman 1989; Petrinovich and O'Neill 1996; Tobia et al. 2013; Hashimoto et al. 2018. [↑](#footnote-ref-9)
9. See Ahler, Roush & Soud (2019) for a discussion of some of the problems associated with collecting data using MTurk and the prevalence thereof. [↑](#footnote-ref-10)
10. We offered forced choice responses in all our experiments. In Nichols and Bruno’s (2010) some of their experiments (those that aim to replicate Blok, Newman and Rips) were run on a Likert scale, and some were forced choice. For consistency we used forced choice throughout. [↑](#footnote-ref-11)
11. Using a Bonferroni correction for multiple comparisons. [↑](#footnote-ref-12)
12. As in experiment 2, all Psychological conditions are variants of the original Psychological No-pain conditions, and all Bodily conditions are variants of the original Bodily Pain conditions, we shall henceforth omit “No-pain”/“Pain”. [↑](#footnote-ref-13)
13. Here by “a stream of consciousness” we do not mean any well-defined technical sense; instead, it’s intended to be understood by the participants as denoting *some* sort of phenomenological continuity that we consider to be involving less than what’s required of full-blown psychological continuity. We think it is reasonable to assume that the participants are familiar with the term of art and would have little difficulty associating it with the intended meaning. If we instead specified the notion of phenomenological continuity in terms of well-defined philosophical jargon, we would probably have taken up controversial assumptions about consciousness that we could not defend, and the participants would have great difficulty understanding the vignettes. (Thanks to an anonymous reviewer for pressing us to clarify this issue.) [↑](#footnote-ref-14)
14. The Psychological conditions are all No-pain to replicate the original Psychological condition from experiment 1. [↑](#footnote-ref-15)
15. The Bodily conditions were both Pain condition to replicate the original Bodily conditions from experiment 1. [↑](#footnote-ref-16)
16. Using a Bonferroni correction for multiple comparisons. [↑](#footnote-ref-17)
17. Using a Bonferroni correction for multiple comparisons. [↑](#footnote-ref-18)
18. Using a Bonferroni correction for multiple comparisons. [↑](#footnote-ref-19)
19. Johnston (1987: 82-3) suggests that this might explain such judgments. [↑](#footnote-ref-20)
20. There is a related worry which questions not our general ability of concept possession but successful application of concepts in experimental setups. That is, unlike in real life, the participants of experiments would have nothing practical at stake if they make judgements haphazardly, and the monetary reward is insensitive to the answers to probe questions, so one might worry that the responses were immediate and unreflective rather than proper application of concepts based on the recognition and processing of relevant information. This is indeed a long-standing objection against the philosophical relevance of experimental philosophy (see, for example, Ludwig 2010) which we cannot fully address here. But we believe that the attentional check/comprehension questions should have succeeded in filtering out those who only wanted to get the experiment over and done with and receive the reward, and that the vignettes were interesting and detailed enough to keep the participants engaged and prompt them to think carefully. (Thanks to an anonymous reviewer for raising these two worries). [↑](#footnote-ref-21)