Uprooting Tradition: Methodological Naturalism and Personal Identity

Sarah M.E. Harrison

Huron University College, sharri94@uwo.ca

Follow this and additional works at: https://ir.lib.uwo.ca/lajur

Part of the Arts and Humanities Commons, and the Social and Behavioral Sciences Commons

Recommended Citation
Available at: https://ir.lib.uwo.ca/lajur/vol1/iss1/4

This Article is brought to you for free and open access by the Huron University College at Scholarship@Western. It has been accepted for inclusion in Liberated Arts: a journal for undergraduate research by an authorized editor of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlswadmin@uwo.ca.
UPROOTING TRADITION: METHODOLOGICAL NATURALISM AND PERSONAL IDENTITY

S.M.E. Harrison, Huron University College

Abstract: This article seeks to undermine the two fundamental tenets of the traditional view of personal identity originally articulated by John Locke. One is the metaphysical claim that the self consists in a unified, continuous stream of consciousness, and the second is, the epistemological claim that introspection is a reliable method of apprehending the self. The paper considers psychologist Michael Gazzaniga’s research with split-brain patients, which reveals that the brain is highly modularized and that each hemisphere possesses a distinct stream of mental states that is superficially unified by the left-brain interpreter. These findings imply that consciousness is neither unified nor continuous. It then discusses behavioural studies conducted by Richard Nisbett and Timothy Wilson that demonstrate private introspective access to mental processes is far more limited than anticipated. Based on these findings, the paper makes two suggestions: 1) that the traditional view of personal identity is untenable, as is any view that relies on either of its tenets, and 2) that personal identity should be studied by means of naturalistic methods.

Keywords: Personal identity; methodological naturalism; Locke; Nisbett and Wilson; Gazzaniga

The traditional and still predominant view of personal identity posits that identity consists in a unified, continuous stream of consciousness, including mental states, emotions, dreams, desires, memories, and so on. Additionally, the traditional view presupposes that agents possess unconditionally reliable private introspective access to their conscious experience and, as a result, they are the definitive authority on themselves, given that information about our identity amounts to knowledge about our stream of consciousness. The intent of this article is to question this view of personal identity and argue that it is unsatisfactory because its central presuppositions can be falsified by empirical evidence. I will argue, in agreement with methodological naturalism, that theories of personal identity should be studied through empirical methods. Consequently, any legitimate claims about personal identity must cohere with empirical evidence, although they do not necessarily need to originate from empirical research, and any claims that fail to do so are not compelling.

The traditional conception of identity derives from John Locke’s theory of personal identity. Locke argued that the source of personal identity must be something that persists over time that can be accessed by introspection (1959: 447). The soul cannot be the source of personal identity because it is unobservable and it is impossible to tell whether one’s soul was previously possessed by another agent. He posits that the only observable and unchanging feature of a person’s existence is her unique stream of consciousness, and as a result this feature must be the source of personal identity: “For, since consciousness always accompanies thinking, and it is that which makes everyone to be what he calls self, and thereby distinguishes himself from all other thinking things, in this alone consists personal identity” (Ibid. 448). Thus, a self is comprised of an individual’s memories, thoughts, perceptions, and desires which occur in an unbroken stream of conscious mental states that are memorable.

While there have been dramatic alterations of Locke’s theory, there are two significant

---

1 The author wishes to thank Dr. Steve Bland for his mentorship.
tenets that have persisted in its modern versions: 1) the metaphysical claim that the self consists in a unified, continuous stream of consciousness, and 2) the epistemological claim that self-reflection is an unconditionally reliable method of apprehending the self. These two tenets are noteworthy because they remain prevalent in the majority of modern theories of personal identity, which are, for the most part, neo-Lockian: S. Shoemaker (1970: 269), Nozick (1981: 656), Lewis (1983: 66), D. Parfit (1984: 216), L.R. Baker (2000: 138), and G. Yaffe (2011: 387). However, despite the popular acceptance of these tenets, they will be the focus of this article because they are problematic. I will argue that both tenets are significantly, if not entirely, undermined by modern research in neuroscience, psychology, and behavioural studies. As a consequence, I will suggest that we must find an alternative account of personal identity that is consistent with natural science.

**Continuity of consciousness**

First, I will consider the metaphysical claim that the self consists in a unified, continuous stream of consciousness, which here will be taken to refer to a single series of coherent mental states. Findings in neuroscience and psychology question this apparently self-evident claim. For instance, Michael Gazzaniga has conducted extensive research with split-brain patients who have had their corpus callosum severed, which effectively disconnects the left and right hemispheres of the brain, making communication between them impossible. Because of the drastic nature of this procedure, it would seem reasonable to expect that it would cause disruption to a patient’s stream of consciousness. However, patients claim that “...they do not feel any different after the surgery than they did before” and “...[t]hey do not have any sense of the dual consciousness implied by the notion of having two brains” (2000: 1319). Thus, despite the two hemispheres being disconnected, a split-brain patient still reports having the same experience of a unified, continuous stream of consciousness as an individual with a fully intact brain. This result raises the question of what causes the continuity of consciousness if it is not the product of the physical unity of the brain. It is important to highlight that in both normal brain and split-brain patients the functions and capacities of both the left and right brain have not altered; they are simply incapable of influencing each other. As a result, split-brain patients allow us to learn about normal brain patients because they expose the processes that are constantly taking place but are undetectable.

Gazzaniga’s research reveals that rather than having a fundamentally unified brain that is lateralized into the right and left hemispheres, the brain is highly modularized, i.e., it is composed of numerous autonomous information processing systems. Each system primarily functions independently, although it can still be influenced by other systems (Ibid. 1314). Consequently, it seems odd that there is any continuity of consciousness at all if the brain functions more like a computer, with multiple independent systems, than a clock, with a series of systems all designed to complete the same task. Gazzaniga argues that the seeming harmony of the brain is the result of the left brain “interpreter” (Ibid. 1319). Language, both spoken and written, is primarily processed in the left hemisphere2, and so too is the ability to “...interpret

---

2 As a result, split-brain patients are unable to verbally respond to a stimulus that is only shown to the right hemisphere, although they are able to tactically respond by drawing a picture or pointing (Gazzaniga 2000: 1299). Significantly, the right hemisphere’s inability to verbally communicate creates the appearance of mental consistency, since the two hemispheres cannot attempt to speak over each other, even though they are still in conflict (Ibid. 1299).
behaviour and unconsciously driven emotional states” (Ibid. 1315). Because both of these abilities reside in the left hemisphere, it is able to act as the “interpreter” of the brain as a whole (Ibid. 1319). As the interpreter, the left hemisphere acts as the spokesperson of the entire brain by creating “...a running narrative of our actions, emotions, thoughts, and dreams” (Ibid. 1320).

Essentially, each of the autonomous systems in the brain sends information to the left hemisphere, which compiles this information into a narrative. Significantly, even without this information the left brain will still construct a narrative. This phenomenon is revealed in studies with split-brain patients where the left hemisphere cannot receive information from the right hemisphere: “The interpreter is driven to generate explanations and hypotheses regardless of circumstances. The left hemisphere of split-brain patients does not hesitate to offer explanations for behaviours which are generated by the right hemisphere” (Ibid. 1319). Gazzaniga argues that because of the left hemisphere’s generation of a personal narrative “...the left hemisphere interpreter may generate a feeling in all of us that we are integrated and unified,” which explains why split-brain patients still believe their consciousness is unified and continuous (Ibid.). These results demonstrate that the unity and continuity of consciousness is simply an illusion produced by the narrative of the left-brain interpreter.

Additionally, because the narrative created by the left hemisphere is consistent and uninterrupted, it produces the feeling that there is one system within the brain that is in control of all other functions, even though there is no single system that oversees or moderates all the functions of the brain. Numerous experiments with split-brain patients have illustrated that, when divided, the left and right brain will make decisions separately, meaning that they can act autonomously and are not under the moderation of a single dominant system.

Several years ago we observed how the left, dominant-speaking hemisphere dealt with behaviours we had elicited from the disconnected right hemisphere...In one example of this kind of test, a picture of a chicken claw was flashed to the left hemisphere and a picture of a snow scene to the right hemisphere. Of the array of pictures placed in front of the subject, the obviously correct association is a chicken for the chicken claw and a shovel for the snow scene. Patient P.S. responded by choosing the shovel with the left hand and the chicken with the right. When asked why he chose these items, his left hemisphere replied ‘Oh, that’s simple. The chicken claw goes with the chicken, and you need a shovel to clean out the chicken shed’. Here the left brain, observing the left hand’s response, interprets that response in a context consistent with its sphere of knowledge—one that does not include information about the left hemifield snow scene. We called this left hemisphere process ‘the interpreter.’ (Ibid. 1316)

These results suggest that each hemisphere of the brain possesses a distinct series of mental states that can conflict, although this is obscured by communication between the hemispheres in normal brain patients. If both hemispheres can function autonomously and conflict, then it does not seem accurate to describe the brain, or consciousness, as unified, when it is composed of independent parts that can conflict and are only superficially strung together by the left-brain interpreter.

Moreover, the previous experiment also illuminates how split-brain patients are equally confident in their responses to stimuli as normal brain patients. This self-assurance is a result of the interpreter immediately creating the most logical or coherent narrative possible and
presenting it as absolutely true: “The left hemisphere without batting an eye, would incorporate the right hemisphere’s response into the framework… [but] the left hemisphere could merely guess. Yet, the left did not offer its suggestion in a guessing vein but rather as a statement of fact” (Gazzaniga and LeDoux 1978: 149). Thus, these trials also reveal how unaware subjects are that this narrative is not a result of their own cognitive initiative, but a consequence of the left brain interpreter making an educated guess. The narrative makes it appear as though all the brain’s mental states are consistent or continuous over time, even though the mental states of the left and right brain can conflict, which is revealed in split-brain patient studies.

These findings present a drastic blow to the traditional view of personal identity. If the brain is actually a highly modularized system, which is only unified superficially by the narrative of the left hemisphere interpreter, then there is no unified, continuous stream of consciousness, and it cannot be the source of personal identity. This conclusion poses serious problems for any theory of personal identity that claims or presupposes that the self is the unity and continuity of consciousness, and it calls for novel theories of personal identity. Furthermore, because this first tenet of the traditional view has been struck down, it puts the epistemological claim that self-reflection is a reliable method of perceiving the self on precarious ground since our allegedly trustworthy introspective access was not able to prevent us from making this metaphysical blunder.

**Introspection**

An extensive study conducted by Richard Nisbett and Timothy Wilson provides evidence that one’s introspective access to the self is significantly more limited than it is assumed to be, and that we can often feel confident in our supposed introspective judgements even when they are incorrect. In particular, Nisbett and Wilson are concerned with test subjects’ introspective awareness of their mental processes, rather than the contents of their mental states (1977: 232). Mental states are the products of mental processes. For instance, a thought is a mental state created by the mental process of thinking. Nisbett and Wilson argue that while subjects can almost always accurately describe their mental states, such as their thoughts, beliefs, or preferences, their accuracy in describing first order mental processes, such as thinking or decision making, is mostly dependent on implicit a priori causal theories, rather than direct introspective access.

Nisbett and Wilson refer to numerous experiments testing the ability of subjects to report on their mental processes (Ibid. 239). If subjects did have complete and direct introspective access to all of their cognitive processes, they would presumably be able to report accurately on external influences that affect their mental processes, such as decision making. In the following study, however, subjects were unable to identify the stimuli influencing their decisions.

In order to test subjects’ ability to report influences on their associative behavior, we had 81 male introductory psychology students memorize a list of word pairs. Some of these word pairs were intended to generate associative processes that would elicit certain target words in a word association task to be performed at a later point in the experiment…The average effect of the semantic cuing was to double the frequency of target responses, from 10% to 20%. (Ibid. 243)
For instance, the given word-pair ‘ocean-moon’ was intended to prime subjects to answer ‘Tide’ when asked to name a brand of detergent. Despite the ten-percent increase in targeted responses based on the associated word-pairs, all of the subjects gave an alternative explanation for their response. Almost no students associated their choice with the word-pair ‘ocean-moon’; rather they would focus on other aspects, such as a feature of the product, e.g. “Tide is the best known detergent” (Ibid.). Subjects do not give these responses as guesses or uncertain answers; they provide these responses as absolute facts regarding their decision making. Although these are plausible reasons for why one might have chosen to answer ‘Tide’ rather than any other brand of laundry detergent, these were clearly not the actual causes in the experiment where subjects were primed to give particular answers with a ten-percent increase from the control group.

Another study similarly concerned with the effects of position on appraisal and choice found the same results—that subjects would attribute their altered behaviour to a rational or salient cause that was simply incorrect in the particular circumstance.

In both studies, conducted in commercial establishments under the guise of a consumer survey, passersby were invited to evaluate articles of clothing [such as]…four identical pairs of nylon stockings…Subjects were asked to say which article of clothing was the best quality and, when they announced a choice, were asked why they had chosen the article they had. There was a pronounced left-to-right position effect, such that the right-most object in the array was heavily over-chosen. For the stockings, the effect was quite large, with the right-most stockings being preferred over the left-most by a factor of almost four to one. When asked about the reasons for their choices, no subject ever mentioned spontaneously the position of the article. (Ibid.)

As in the previously mentioned experiment, subjects would provide alternative explanations for their choice. For instance, they would state that they based their decision on the quality of the material, which would be a reasonable explanation if all the items were not identical (Ibid. 244). Furthermore, when directly asked whether the position of the stockings might have impacted their decision, all test subjects denied the possibility of relative position having any impact at all (Ibid.). Thus, it is clear that in both cases subjects were making dramatically incorrect introspective claims about their mental processes.

Based on these experiments, Nisbett and Wilson argue that subjects do not have direct introspective access to mental processes as previously assumed. They claim that subjects could not have introspective access to these processes because “…the explanations that subjects offer for their behavior…are so removed from the processes that investigators presume to have occurred as to give grounds for considerable doubt that there is direct access to these processes” (Ibid. 238). For instance, in last example, none of the subjects considered the position of the stockings relevant to their decision, but they did provide a prima facie plausible reason for their selection, such as the quality of the material. Thus, because subjects are seemingly unaware of the effects of controlled stimuli on their mental processes in numerous experiments, it is reasonable to conclude that they simply do not have direct cognitive access to these mental processes.

Instead, Nisbett and Wilson propose that subjects rely on a priori causal theories in order to make rational inferences based on the evidence they are given. When the circumstances

---

3 They are a priori because they are inherent, rational generalizations that predate our exposure to the circumstances that require us to use them (Ibid. 238).
are normal, agents usually apply a relevant causal theory that results in an accurate verbal report (Ibid. 248). However, when the circumstances are clouded by experimental controls, agents give incorrect verbal reports because they have applied an incorrect causal theory based on unclear or misleading evidence (Ibid. 252). In the nylon stocking case, for instance, it is not a salient causal theory to suppose the agent chose a pair of stockings because they were on the right hand side, because we normally pick clothing based on considerations of quality or use, and do not generally choose between identical items of clothing in a controlled experiment. Thus, subjects’ incorrect verbal reports are not unrealistic because they are relying on a priori causal theories, which are generally reliable, but are fallible (Ibid. 247).

Moreover, because agents apply a priori causal theories in order to explain mental processes, there is a strong correlation between verbal reports from different sets of test subjects. For instance, one experiment tested five stimulus factors on the perception of a female job candidate: 1) appearance, 2) academic credentials, 3) clumsiness, 4) history of an automobile accident, and 5) whether they would meet the candidate (Ibid. 250). After receiving a profile about the applicant, subjects were asked to judge, based on a scale from zero to five, how much each of the previous stimulus factors had influenced each of the following judgements of the candidate: “(a) how much they liked her, (b) how sympathetic they thought she would be toward clients’ problems, (c) how intelligent they thought she was, and (d) how flexible they thought she would be in dealing with clients’ problems” (Ibid.). The same judgement task was given to another group of observers, who did not get a specific profile of a job candidate, but had to speculate on how the same stimulus factors influence their judgement. The result of the experiment was a strong correlation between the two sets of reports: “Mean subject and observer reports of factor utilization were correlated .89 for the liking judgement, .84 for the sympathy judgement, .99 for the intelligence judgement, and .77 for the flexibility judgement” (Ibid.). The high correlation suggests that any rational agent with sufficient information would produce the same report because all agents are using the same means of evaluation: not private introspective knowledge, but a priori causal theories.

As a result of the implicit use of a priori causal theories to report on mental processes, Nisbett and Wilson make a distinction between what can be called broad and narrow introspection. Broad introspection encompasses the typical meaning of the term, namely a “correct verbal report” on one’s own cognition (Ibid. 251). In contrast, Nisbett and Wilson propose an alternative narrow definition that limits instances of introspection to “verbal reports which exceed in accuracy [those] obtained from observers provided with a general description of the stimulus and response in question” (Ibid.). This narrow definition eliminates cases in which verbal reports rely on a priori causal theories (Ibid. 255). Thus, Nisbett and Wilson argue that if we want to consider introspection as the necessarily private inward observation of mental states or cognitive processes, then the criteria of introspection need to be redefined in such a way that they only include specific instances in which this kind of access occurs.

Nisbett and Wilson’s research makes it clear that subjects have limited direct introspective access to mental processes, such as evaluation, judgment, problem solving, and the initiation of behavior, which is why subjects rely on a priori causal theories to explain these processes (Ibid. 232). This conclusion poses two problems for theorists of personal identity, especially for those who argue that personal identity can only be known by means of direct and private introspective access. First, there is a wealth of experimental evidence demonstrating the unreliability of introspective verbal reports, which other psychologists have taken as sufficient justification for anti-introspective positions with respect to self-knowledge (Ibid. 250). Recall
Gazzaniga’s research and the apparently self-evident belief in the continuity of consciousness. This belief was seemingly justified by means of introspection, but has been shown to be empirically false.

Secondly, according to Nesbitt and Wilson’s narrow definition of introspection, subjects’ private knowledge is significantly more limited than we assume. Thus, not only do we need a test to judge the accuracy of our introspective reports, but our private knowledge is drastically limited. It is certainly much less persuasive to argue that a theory of personal identity is justified on the basis of private knowledge of mental states, such as thoughts, beliefs or intentions, without reliable knowledge of how those states arose. Even intuitively, this position does not seem to have sufficient evidence and justification to fully comprehend the self in a meaningful way. If we take the self to be something that can make decisions, form habits, arrive at conclusions, etc., then it is clear that the self is not something that can be known in accurate detail through narrow introspection alone. Hence, in order to understand the self as something that includes mental processes, it is necessary to study the self through a naturalized methodology, rather than exclusively from the first-person perspective, because the most accurate information regarding personal identity has come from empirical evidence, not introspection. In particular, the studies considered have 1) falsified the two fundamental tenets of the traditional view, and 2) provided a more complex understanding of the brain and cognitive behaviour, which lends itself to a more well-informed understanding of the self.

Conclusion

It has been demonstrated that the two fundamental tenets of the traditional empiricist view of personal identity are extremely problematic. The metaphysical claim that the self consists in a unified, continuous stream of consciousness has been undermined, since there is empirical evidence that strongly suggests that the perceived unity and continuity of consciousness is an illusion caused by the left-brain interpreter and its narrative. In reality, the brain is highly modularized. In addition, the epistemological claim that self-reflection is an unconditionally reliable method of apprehending the self has been challenged, because actual private knowledge of mental processes seems to be extremely limited. Narrow introspection, in Nisbett and Wilson’s sense, is limited to mental states, without access to the mental processes which give rise to them. Thus, narrow introspection seems to be an insufficient source to understand an entity as complex as the self. Therefore, it is clear that we need an alternative metaphysical account of personal identity, as well as a supplementary epistemological method of discovering and justifying our theory. I have suggested that any investigation into personal identity that does not rely, in part, on naturalistic methods will be severely limited. The suggestion of this paper, then, is that the traditional view of personal identity, as well as any other view reliant on the same tenets, should be uprooted and abandoned in favor of a theory of personal identity that is consistent with empirical evidence.

Bibliography


S.M.E. HARRISON is a fourth year student at Huron University College completing an Honours Double Major in English Language and Literature and Philosophy as well as the Scholar's Elective Program. In September 2015, she will begin attending McMaster University for her MA in English and Cultural Studies, where she has been awarded a CGS-M SSHRC.