2012

Selfhood and the Unity of Consciousness

Kate Hanniball
khanniball@pugetsound.edu

Selfhood and the Unity of Consciousness

Authored by: Kate Hanniball

Collier-Bangs Summer Research Grant 2012

Advisor: Justin Tiehen

I. Introduction

The concept of the self is one which is at once puzzling, as well as vastly important in the most personal sense to the experiences we undergo as human beings as well as the way we receive and process sensory information from the external world. Selfhood, or the idea of an existent self seems inseparable from our existence in both a profoundly abstract and scientifically concrete sense. In the 17th century Descartes posited mind body dualism to explain the existence of the self, giving birth to the infamous concept of something akin to a "ghost in the machine". Descartes, Locke, and other contemporaries created this concept by positing the self as a sort of immaterial soul which resides in or above the body, and by means of some mysterious form of communication, controls the body much in the way a puppet master would control her dolls. This vision of the self as an immaterial substance which is something more than just the physical body is one which has strong intuitive pull. The idea that we cannot simply be boiled down to randomly firing neurotransmitters conforms to the strong intuition held by most that what it *means to be me*, or have a self, is something which is a great deal more profound than the physical capabilities of my body. This intuition is brought out famously by the philosopher John Locke in the scenario known as the "Prince and the Cobbler." In this scenario there exist a prince and a cobbler who undergo a brain swap scenario in which the entire psychologies of the prince and the cobbler are swapped. Suppose that pervious to the swap, the prince committed a heinous crime; also suppose that the prince *knew* that the swap was going to occur, and planned to use the swap to evade prosecution. When the guards arrive to take the prince to jail, they take away the princes body, despite the fact that "the person" who is inside the prince is actually the psychology of the cobbler. When we evaluate our intuitions concerning this scenario it seems as though there has been a gross miscarriage of justice, the "person" who should be punished is the

scheming individual who now resides in the body of the cobbler, as that is the consciousness that knowingly committed the crime. As such, as is illustrated by the manipulation of our sense of morality concerning the "perpetrator" of the crime in the scenario described above, it seems as though there is a great deal more to the self than the mere physical body. However, the idea of the self as something immaterial which communicates directly, (by some unknown sense) with a physical body located squarely in space and time, seems distinctly unscientific, is not (and perhaps never could be) empirically verifiable, nor is it observable in any sense. These shortcomings make mind-body dualism a theory which is easily dismissed as implausible due to a lack of any sort of scientific foundation.

The idea of an existent self however, is not constrained to the realm of mind body dualism, and it is not only in an abstract sense that we form an idea of ourselves. As Dennett points out in his "Origin of Selves" (Dennett, 1989) there is an inherently recognizable distinction between "the self" and "the other" which is a deeply rooted biological principle. This principle is so central to our existence and the existence of the other creatures with which we share this earth, that it is an implicitly foundational principle to evolutionary theory. Specifically, in order to ensure our survival, (leading to the survival of species) there must be some boundary between the self and the other so that we may act according to self preservation, ensuring our survival, and the survival of our genes through reproduction. Put more simply, we must be able to make the distinction between "us and them," the internal and the external, if we are to make it very far in the race for survival. Again, here we have a way of classifying the self, which includes some of our intuitions, (namely that the idea of self ought to have some empirically scientific roots, and answer some of our questions about why exactly the idea of self is so essentially important to human experience) but again falls short in that it seems there is

something profound about the self or "what it means to be me" that goes beyond the skeleton of strictly scientific theory.

As we can see, it seems clear that selfhood plays an integral role our concept of human experience, in both a profoundly abstract as well as foundationally concrete sense. It will be a goal of this paper however, to argue that the age old epistemological pursuit of "what it means" to have a persistent self is ultimately unimportant; and that there may not be (strictly speaking) any fact of the matter about whether selves exist or not. Further, this paper will take for granted that which has been assumed by Derek Parfit concerning the self as a persistent being[1]and work within this framework to address the important questions of what it means to have a self at any one temporal moment; as well as whether it is possible to have a singular self which persists through time. It will be the aim of this paper to argue that these two questions are inextricably linked, and that the answer to one entails the solution to the other, namely that the reality of selfhood (at any given moment in time) ensures that a "persistent self" is an impossibility.

At first glance, it may seem that to claim that there may be *no fact of the matter* about whether selves exist or not, is one which is nothing short of absurd, however it will be the goal of this paper to convince the reader that this is indeed the case. I do not wish to make the claim the vision or idea of the self does not matter, but rather that there is no profound truth about whether selves exist factually. It is instead it is the appearance of selves which is of importance in the debate surrounding selfhood. I will argue that past the surface of appearances, there is no deep fact of the matter concerning the self nor is it particularly metaphysically interesting to engage in the conversation of whether selves exist in factual reality.

In order to illustrate my point to the fullest extent it is helpful to consider a distinction

[1] See Derek Parfit, *Reasons and Persons,* chapters 10-13, p.199-306.

between the immediate world of experience, and the possibility of the existence of a removed metaphysical world containing the realities of things in themselves. Specifically, the immediate world can be supposed to contain all objects of experience, (objects, events, subjective experiences) which we experience through the veil of our subjective sensibilities and intuitions. Contrarily, the metaphysical world can be imagined to contain "things in themselves" or the metaphysical realities of the way things are, sans our human perceptions of the events in question. It is my claim that the objective truth about whether or not there are "selves" falls into the realm of something akin to the metaphysical world described above. Specifically, because of the realities and limitations of our brains there is an insurmountable epistemological gap between the so called "reality of selfhood" and the appearance of selfhood we all have. In short, we will never be able to turn an objective eye to what selfhood entails because we simply cannot get past the subjective veil of our own selfhood to do so. However, as it turns out, rather than some objective truth, it is the appearance of the self we all develop which is of metaphysical importance. This is of the central role our vision of selfhood plays in the way we experience the world, the way we process external information, and the way it affects the formation of our desires, intentions and subsequent actions. Therefore the following sections of this paper will attempt an examination of the central tenets of the universal picture of selfhood, and further show why investigating *the appearance* of the self is a metaphysically worthwhile pursuit.

II. Unity of Consciousness

Although there are many aspects of selfhood that may be considered foundational to our idea of the self, for the purposes of the present paper, a focus will be maintained upon the so-called "unity of consciousness" that many of us consider to be at the center of our vision of the self. The theory of unity of consciousness attempts to explain the sense of unity we have

concerning our conscious mental states, as well as our sense of ownership over our mental states. Take the example of my first college class: during my first college class I enjoyed many different phenomenal states; specifically I entertained the visual experiences of eagerly examining my fellow classmates, registering the professor at the front of the room, and taking in the details of how the classroom appeared to me. During this time I simultaneously experienced auditory sensations such as those of listening to my professor giving her class lecture, listening to the comments of my classmates, and so on. Finally, also during this time, I experienced emotional and cognitive sensations of various kinds, such as thinking critically about what the professor was saying, feeling excitement and anxiety about being in my first college class, as well as a deep awareness of my sense of self and of how I was presenting myself to my peers—I wanted to come off as smart and intelligent without being pretentious or conceited, quick witted without being the class clown, and so on and so forth. Through this description of my first college class, one can begin to understand what unity of consciousness entails: the unity of all of my simultaneous experiences rolled into one subjective experience: that of my first college class.

Timothy Bayne (Bayne, 2008) famously described phenomenal unity of consciousness as the experience of "what it is like" to be me. We can see how unity of consciousness is necessary for understanding our subjective experiences as well as the subjective experiences of others because unity of consciousness consolidates our experiences in such a way that there is something it would "be like" to have them. Another way of thinking about unity of consciousness, concerns the sense of ownership we have over our subjective experiences. The experience of *my* first college class is something I share with no one, it is entirely and completely mine because there is only one "me" or one mind which underwent the actual experience of my first college class, there are not two or three other "selves" which have had parts of the

experience of my first college class but rather just the one "me" who has. Despite the seemingly obvious nature concerning ownership of our mental states, this ownership has been contested several times in debates concerning unity of consciousness.

If there were no exceptions to the rule of singly unified ownership over our subjective experiences then we could happily conclude that what matters in the debate about selfhood concerns whether the subject is able to entertain phenomenally unified personal experiences and leave it at that. However there has been great debate since the early 1970's about whether or not our vision of the self as a unified consciousness is actually coherent. This debate often centers around neuropsychological discoveries concerning bi-hemispheric communication between the right and left hemispheres in the brain, and the strange psychological phenomena that occurs when the corpus callosum is severed. During the 1960's brain bisection operations were sometimes performed as a last ditch effort to attempt to control the seizures of severely epileptic patients. In brain bisection operations the corpus callosum, (a strand of neurons which connects the right and left hemispheres allowing for communication within the brain) is cut, the goal being to stop epileptic seizures from spreading from the left to the right (or vice versa) lobes of the brain. For a long time it was assumed that these operations had no effect on the patient other than effects desired due to the fact that patients would continue their normal lives exhibiting no signs of impaired functioning. However sometime later, experiments done using brain bisected patients told a somewhat different story, leading psychologists and philosophers to speak of a "dual consciousness" in brain bisected patients. The facts concerning brain functioning are as follows: the left hemisphere is associated with the right side of the body, while the right hemisphere is associated with the left side, as such tactile stimuli and sensations are transmitted contralaterally. Visual stimuli are also transmitted contralaterally in that the left half of each

retina (which scans the right half of each visual field) sends information to the left hemisphere, while the right half of each retina, (scanning the left half of the visual field) sends information to the right hemisphere. Further, the left hemisphere is associated with the control and production of speech. Laboratory experiments making use of such knowledge concerning brain functioning have been conducted on split brain patients with surprising results. These results are summarized by Nagel (Nagel, 1971) in his paper "Brain Bisection and the Unity of Consciousness" and are as follows: in brain bisected patients,

"What is flashed to the right half of the visual field, or felt unseen by the right hand can be reported verbally. What is flashed to the left half of the visual field, or felt unseen by the left hand cannot be reported verbally. However, if the word "hat" is flashed on the left the left hand will retrieve a hat from a group of concealed objects if the person is told to pick out what he has seen. At the same time he will insist verbally that he saw nothing" (Nagel, 1971, p. 400).

This strange behavior seems to indicate a dual consciousness, or the development of 2 separate consciousnesses residing inside a single body, immediately raising concerns about the extent to which we can confidently attribute a single countable mind to such patients. Indeed, such experiments have lead some philosophers to question whether unity of consciousness is something we can confidently attribute even to normal individuals due to the fact that it seems highly unlikely that experimental conditions can elicit (from an otherwise normally functioning individual) a new consciousness to spring into existence and then equally quickly disappear once the individual goes back to daily life. Indeed, the results from such experiments make it seem rather more likely that we have made an error in assuming that "the self" we all have is a single unified global consciousness which subsumes all of our mental states creating a larger cohesive unit.

However, there is something distinctly disconcerting about the idea of sharing our body with multiple "selves" or unities of consciousness, and there is a strong intuitive pull towards any viewpoint which posits only one "self" operating at a time, certainly this is what it *feels like* is going on. Indeed to refer to the point made by Timothy Bayne, it does not seem as though there is anything it would "be like" to share one's body with multiple competing selves which operate simultaneously. Therefore, how do we reconcile the appearance of selfhood with the logical inconsistencies raised by such problems as brain bisection or other strange psychological phenomena? I will defend make the claim that something akin to the successive selves theory described by Parfit in his book "Reasons and Persons" (Parfit, 1984) is the closest we can hope to come to explaining the appearance of the self. Not only does the successive selves theory provide a coherent response to the logical problem raised by brain bisection experiments, it also provides a coherent answer to the puzzling question of how we manage to stay the same person over time even though quite literally *everything* about us changes.

III. Successive Selves

The version of the successive self view I wish to defend plays out as follows: with every change in mental state, sensory input, or cognitive processing, the "self" we identify with as a person changes, in this way each cohesive self (all possessing an instantaneous unity of consciousness) passes out of existence as quickly as each "new" self comes into being to replace the previous self. In short, every moment (conscious or unconscious) long enough for us to formulate an awareness of our existence is presided over by a selfhood, one which is replaced by a new selfhood the next instant. Each of these successive selves has as much claim to "who we are" as the previous, but is the most relevant to our idea of selfhood in the moment in which each self is in immediate existence. However, it is not this instantaneous "unity of consciousness"

which matters in the idea of selfhood, but rather the connection between each successive self. The connection between selves has to do with a kind of continuity of mental states rather than simply the continuation of the physical body. However, selves that have passed out of immediate existence many years ago have a different kind of connection to the currently presiding self than do those who are passing out of immediate existence as I speak. Here Parfit's distinction between psychological continuity and psychological connectedness is helpful. Specifically Parfit states that psychological connectedness refers to "the holding of particular direct psychological connections" (Parfit, 1984, p.206) as such psychological connectedness is a transitive relation and can hold to a matter of degree. This differs from psychological continuity which Parfit defines as the "holding of overlapping chains of *strong* connectedness" (Parfit, 1984, p. 206). This distinction is extremely helpful when discussing the relationships between successive selves because it helps to illustrate their relevance to the self at hand. Specifically, my current self is psychologically continuous with the immediately previous selves because there is a strong and overlapping connection between the present and immediately previous selves which orients me to the continuous and uninterrupted flow of daily life that I am accustomed to. However this strong relation does not hold between my present self and the self I was at some moment 13 years prior. However, there is *some* relation there, which allows me to refer to the person I was at that moment 13 years prior as me. This relation holds because the self I was at some moment 13 years prior is psychologically connected with the present self to some degree by virtue of the transitivity principle of psychological connection.

Due to the fleeting nature of the "unity of consciousness" that is in charge in each moment, what actually matters when we discuss selfhood are the various psychological relations we have to each of our successive selves. When this view is adopted we can explain the

appearance of selfhood solving the puzzling issue of our persistence across time as well as dealing with brain bi-section. Successive selves provides a solution to the brain bi-section problem because it shifts the focus away from considering one global and time expansive unity of consciousness as "the real me" and places emphasis instead on the connections that hold between successive instantaneous unities, therefore allowing us to attribute the strange discrepancies that occur in the behavior of brain bisected patients to malfunctioning in the physical brain rather than a strange "dualility of unified consciousness" or simultaneously operating selves. Indeed, even further support for this is provided by information gleaned from some of the laboratory experiments conducted upon split brain patients when the right hemisphere is faced with a communication problem. Dennett brings our attention to the resourcefulness of some brain bisected patients in his paper "The Self as a Center of Narrative Gravity" (Dennett, 1992). Experiments have been conducted where split brain subjects must reach into a closed bag with the left hand to feel an object which they are then asked to identify verbally, because of the removal of the communication link between hemispheres, the right hemisphere gets information about the object from the left hand, but the left hemisphere, (which controls speech) is kept in the dark, therefore the patient is unable to publicly announce what has been felt in the bag. However, there is a solution to this communication problem which some split brain patients have been observed to discover. As quoted by Dennett

"Whereas ordinary tactile sensations are represented contralaterally, pain signals are represented ipsilaterally (meaning pain sensations go to both hemispheres) suppose the object in the bag is a pencil. The right hemisphere will sometimes hit upon a very clever tactic: hold the pencil in your left hand so its point is pressed hard into the palm; this creates pain letting the left hemisphere know there's something sharp in the bag, which is enough of a hint that it can begin guessing; the right hemisphere will then

signal "getting warmer" and "got it" signals by smiling or other controllable signs and in a very short time the subject will be able to announce the correct answer" (Dennett, 1992).

As such it seems as though these experiments reveal a talent for overcoming accessibility and communication problems in patients who have impaired brain functioning. When applied to the theory of successive selves it would seem as though the unified self "in charge" at a particular moment is experiencing difficulty (due to physical damage) in creating a cohesive story for the present experiences of the patient at hand. However, due to the resourcefulness of the brain and our determination to force a unity into our experiences it is possible for the self at hand (by means of clues and hints given) to combine fragments of a phenomenal experience together in such a way that allows us to maintain the appearance of unity—much in the same way we have the appearance of a persistent expansive unified consciousness despite the passing in and out of existence of successive selves. This is because it is a psychological tendency to create unity where perhaps no one appears, or to organize our life experiences into a unified story.

IV. Psychological Tendencies and the "Narrative Self"

The question which now becomes relevant is: why are we so invested in this "unified story"? As indicated above, it is more likely that what really matters in the appearance of the self is the relationships between present selves rather than the appearance of a single, expansive unified consciousness. So why do we find ourselves preoccupied with such a self? I will argue that there is nothing more to this intuition than a psychological tendency to view ourselves as a unified, consistent, and pervasive unit which persists through time. However this tendency has no metaphysical depth other than it is useful for us to think of ourselves in this manner, it helps us make sense of our lives, develop an idea of who "we really are" and aids in our ability to form

relationships and communicate with others around us. However the idea of a persistent self is nothing short of an abstraction which has a psychologically intuitive pull. As seen above, logically what seems to matter about the appearance of a self is a relationship that connects "selves" or present mental states; however we don't necessarily like to think of ourselves in this way. The reason for our belief in the self as a persistent unity of consciousness has been artfully illustrated in a useful analogy coined by Dennett (Dennett, 1992). Dennett draws a helpful parallel between an object's "center of gravity" and our "narrative selves" which I believe, allows us to see the true function of the appearance of a persistent self. Specifically, a center of gravity is a concept of Newtonian physics that has an established place both in the theory and language of science. It figures into explanations that appear to be causal in nature—"why doesn't the lamp fall over?" "Because its center of gravity is so low"—however, a center of gravity cannot be *found* anywhere, it occupies no spatio-temporal location, has no physical properties to speak of, and does not resemble any other physical object in space. Indeed, to try and pinpoint where *exactly* an object's center of gravity is located would be to make a category error, one would not look at a lamp on a table and ask "but where exactly is its center of gravity?" to do so would be absurd. In this sense, a center of gravity is like the idea of a conscious and persistent unified self, it is an abstraction, a fiction or "causal story" used in an offhand way to offer a coherent explanation for the events of our daily lives, but in reality is a somewhat empty notion.

I propose that this narrative "center of gravity" or self, is constantly changing an updated version of which is available with every new successive self or change in mental state. In our daily lives this change is gradual, we do not notice a shift in our "narrative figurehead" from one moment to the next, however if we look back on who we were five or ten years ago, we find that there is often a great discrepancy between who we identify with now and who we felt we were

then. This discrepancy is tangible and can be attributed to the different relationships our presiding narrative self shares with previous selves. Specifically, as mentioned above, the relationship between my current "self" and those immediately previous is much stronger than that which holds between my present self and that of some moment 5 years ago. This is due to the level of psychological connectedness present between each of the selves in question. The level of overlap between my current self and those immediately previous is very high, a direct connection of psychological continuity, however the relationship between myself now and myself 5 years ago is direct only because such a relationship is transitive, and the level of overlap is low or nonexistent. Therefore, it is natural that the constant change in my narrative self that occurs daily seems somewhat seamless, and offers no interruption to daily life.

The claim that our narrative self changes literally from one mental state to the next is not a weak claim. I am proposing that in a very real sense our vision of what constitutes a self ceases to exist and is replaced by a new and updated narrative, or to be put more simply, each narrative self dies and is replaced by a brand new updated self constantly throughout the course of our lives. As such, a common intuitive reaction to such a proposal is to object to it on the grounds that we certainly would notice if, rather than being a persistent being, or having a single unified self which stretches across time, we were instead merely a collection of instantaneous and every changing narrative figureheads. Certainly, it will be objected, there is a distinct enough difference between the death of a self, and its continued existence, that each option is describing profoundly different outcomes. However I will argue that the two options are merely semantically different ways of describing one and the same outcome, that there is no deep metaphysical fact of the matter which makes the death of a self or its continued existence, vastly different scenarios.

In order to make this claim, it is important to understand that my proposal is reductionist in that it is my belief that in some scenarios concerning personal identity there may be *no objective fact of the matter* about whether a self is existent at a particular moment, or whether it has ceased to exist. To put it more simply, there may be scenarios involving personal identity where it makes the most sense to claim that the identity of the thing or "self" in question is distinctly indeterminate in a most unpuzzling way. It may be hard to understand this type of abstract claim without the help of a somewhat more concrete example, as such it will be helpful to consider Parfit's club example to make the claim more apparent, the scenario is as follows:

"Suppose that a certain club exists for several years, holding regular meetings. The meetings then cease. Some years later, some of the members of this club form a club with the same name, and the same rules." (Parfit, 1984, p. 213).

The question we may ask ourselves concerning the nature of this club is: have these people reconvened the very same club? Or have they instead formed a new club which is exactly similar? It is possible that there is some definitive answer to this question, it may be that the club has some rule built into its structure concerning how long it may be disbanded and may be still called the same club upon its reinstatement, however it is equally likely that the club has no such rule and that, when asked, the members of the club wouldn't give it a definitive answer. It would seem then that there would be no real answer to the question "is this the same club." And that either claim concerning the status of the alleged club would be *neither true nor false*. Put more simply, there is no fact of the matter concerning the correct status of the club, because both the claim "this is the same club" and "this is a different club" are simply semantically different ways of describing one and the same outcome. In much the same way, I believe there is something indeterminate about the nature of our identity. Specifically, a self passing out of existence and

being replaced by another, and the continued existence of a persistent and cohesive self are simply semantically different ways of describing the same outcome. Because of our inability to ever get outside of our own brains to examine selfhood objectively, we cannot know whether there would be a perceptually significant difference in our lives if we existed as a bundle of successive selves all connected by some kind of direct relation, or if we existed as a continually coherent overarching and persistent self. However, as shown above, it logically makes more sense to accept the former view. Whether or not there is a fact of the matter about the entity of the self may not have a coherent answer, however the idea we have of some kind of selfhood is a psychologically useful tool for helping us navigate our lives and our relationships with others. Finally, the question of whether or not we would notice the death of one self and the "birth" of another seems to be a somewhat indeterminate question, where there is no metaphysically profound difference between the "life" or "death" of a self because both options are simply two different ways of describing the same outcome.

V. Conclusions

As such, the concept of selfhood is one which is vitally important to us from a psychological perspective because it is through the lens of the self that we view our world, navigate our lives, and create our narrative stories. Despite the importance of the concept of "the self," because of the logical constraints we are faced with (namely our inability to obtain an objective standpoint when regarding the concept of ourselves) we will never be able to conclusively discover whether "selves" actually exist as things in themselves, I think it doubtful that they do and find it more persuasive that the concept of selfhood is nothing more than a psychological tendency. Despite this, *the appearance* of the self is something worth investigating. As I hope to have shown in this paper, the intuitive idea we have of the self as a

persistent and ever changing entity is actually not what matters when we are talking about our vision of selfhood. Instead, it is logically more likely that we are a bundle of successive selves connected by some direct relation, and it is actually the relation, rather than the unified consciousness, which is important when we are discussing selves. Further, the difference between the passing out of and passing into existence of a self, does not describe some deep metaphysical difference but rather simply is two different ways of describing one and the same outcome. Although at first disturbing, I believe this should give us a somewhat more optimistic outlook upon the time when we pass out of existence entirely, because perhaps it will not be as final as once thought, but rather a continuation of the cycle of selfhood, continued without a physical vessel.

References

Bayne, T. (2008). The unity of consciousness and the split-brain syndrome. *The Journal of Philosophy, 105*(6), 227-300.

Dennett, D. (1989). The origins of selves. *Cogito*, 3, 163-173.

Dennett, D. (1992). The self as a center of narrative gravity. In F. Kessel, P. Cole, & D. Johnson (Eds.), *Self and Consciousness: Multiple Perspectives* (275-288). Hillsdale, NJ: Erlbaum.

Nagel, T. (1971). Brain bisection and the unity of consciousness. *Synthese*, 22, 396-413.

Parfit, D. (1984). *Reasons and persons*. New York: Oxford University Press.

Brook, A., & Raymont, P. (2010). "The Unity of Consciousness", *The Stanford Encyclopedia of Philosophy* Zalta, E. N. (ed.), URL = <http://plato.stanford.edu/archives/fall2010/entries/consciousness-unity/>.