Problems for Introspection as a Basis for Reasoning about the Self

Dillon W. Bak

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Arts
In
Philosophy

Kelly Trogdon
Benjamin Jantzen
James Klagge

25 April 2018
Blacksburg, VA

Keywords: self, phenomenology, introspection, metaphysics

© 2018 Dillon Bak
Problems for Introspection as a Basis for Reasoning about the Self

Dillon W. Bak

ABSTRACT

Through introspection we may gain insight into phenomenology and thereby learn about our own mental lives. One aspect of our phenomenology that we might wish to introspect is our experience of selfhood. In particular, Galen Strawson views phenomenology as particularly useful for reasoning about the self. He expresses this in what he calls the Equivalence Thesis, which states that there are selves if and only if there is something that has properties attributed to the self in every instance of self-experience, where self-experience refers to a phenomenological experience of selfhood. In order to arrive at a phenomenological characterization, any set of properties that characterizes the self via the Equivalence Thesis, one must examine the phenomenology of self-experience through introspection. The Equivalence Thesis can run into difficulties in at least two ways with respect to its reliance on introspection. If introspection is unreliable then the Equivalence Thesis fails as we cannot accurately examine our phenomenology. While some of the consequences of such unreliability will be explored this will not be the main focus. Instead I call into question whether or not introspection provides the information that Strawson says it does. The Equivalence Thesis depends on the ability of introspection to provide us with information about so called mental elements, which give structure to our overall phenomenology. However, this is implausible. When we introspect we can learn directly about the kind of experience we are having, but it will not allow us to form an acceptable phenomenological characterization.
Problems for Introspection as a Basis for Reasoning about the Self
Dillon W. Bak

GENERAL AUDIENCE ABSTRACT

Throughout life we all have various phenomenological experiences. One may feel happy, see a familiar face, or think about solutions to a problem. Each one of these experiences has a first personal element to it. That is we perceive these as belonging to a singular entity, a self. In fact, we often think of ourselves as selves due at least in part to the fact that we are conscious. Some of our conscious experiences even contribute to the way we think about selves. Galen Strawson argues that these experiences, which he calls self-experiences, provide a suitable foundation for our metaphysical reasoning about the self in the Equivalence Thesis. I argue, however, that our phenomenological experiences should not be used to provide such a foundation.
Acknowledgment
I am grateful to the members of my committee, Dr. Kelly Trogdon, Dr. Benjamin Jantzen, and Dr. James Klagge, for all the guidance and feedback they having given me on this project.
Table of Contents

Section 1: Introduction.................................................................................1
Section 2: Self-experience and Strawson’s Project.................................2
Section 3: Unreliable Introspection.........................................................4
Section 4: Problems for Introspection.......................................................5
Section 5: Conclusion.................................................................................8
References .................................................................................................10
Section 1: Introduction

Through introspection we may gain insight into phenomenology and thereby learn about our own mental lives. Under ordinary circumstances the information gained from introspection can be quite useful. You might learn that you are hungry, cold, or happy among other things and then respond accordingly. However you may have had experiences where the opposite is true, where introspection is misleading, such as being unaware that you are anxious about something, and it is having a negative effect on your behavior. So we come to realize that there are limitations to our introspection. Moreover, it is not always immediately clear in what situations we gain useful information from introspection. We must then take caution in using introspection and assess whether its use is justifiable in a given context.

One aspect of our phenomenology that we might wish to introspect is our experience of selfhood. In particular, Galen Strawson views phenomenology as particularly useful for reasoning about the self. He expresses this in what he calls the Equivalence Thesis, which states that there are selves if and only if there is something that has properties attributed to the self in every instance of self-experience, where self-experience refers to a phenomenological experience of selfhood\(^1\). In the following any set of properties that characterizes the self via the Equivalence Thesis is called a phenomenological characterization. In order to arrive at a phenomenological characterization one must examine the phenomenology of self-experience through introspection. After this process of introspection Strawson needs two things to be true. First, there is some phenomenological characterization of the self that we can find by examining our phenomenal experiences. Second, this phenomenological characterization will be common to everyone who has self-experiences. To explain this second point, there must be a common phenomenological characterization since Strawson is suggesting that our phenomenology establishes a valid metaphysical foundation for reasoning about the self. So through introspection everyone should be able to come to the same phenomenological characterization after sufficient reflection so as to work from the same metaphysical basis.

The Equivalence Thesis can run into difficulties in at least two ways with respect to its reliance on introspection. First if introspection is unreliable then the Equivalence Thesis fails as we cannot accurately examine our phenomenology. While some of the consequences of such unreliability will be explored this will not be the main focus. Instead I call into question whether or not introspection provides the information that Strawson says it does. The Equivalence Thesis depends on the ability of introspection to provide us with information about so called mental elements, which give structure to our overall phenomenology. However, this is implausible. When we introspect we can learn directly about the kind of experience we are having, but it will not allow us to form an acceptable phenomenological characterization. This will be shown in two ways: by suggesting that introspection does not provide us with a way of choosing between competing phenomenological characterizations and by proposing that attempting to use introspection to learn about the mental elements associated with self-experience is likely affected by beliefs we already hold about the self. In summary, after using introspection to identify mental elements common to all self-experience Strawson wishes to offer a minimal phe-

\(^1\) Strawson p.55
nomenological characterization of the self via the Equivalence Thesis. However, introspection does not support the kinds of conclusions he wishes to draw from it. So the Equivalence Thesis cannot be used for reasoning about the metaphysics of the self as the way it relies on introspection is unrealistic.

**Section 2: Self-experience and Strawson’s Project**

Strawson depends on an analysis of self-experience in order to advance his theory. In order to define self-experience one must understand the method by which beliefs about self-experience are formed in Strawson’s account. He notes that we have, at first vaguely, some experience of being a locus of consciousness or of existing as some sort of mental presence.² It is from these experiences that we derive our self experiences, but Strawson does not provide an account of how this occurs. The following will attempt to bridge the gap between locus of consciousness experiences and self-experience, and then on to how one would form a phenomenological characterization.

As noted we begin by picking out a certain kind of experience, that of being a locus of consciousness. This is an experience which is “something extremely basic,” the likes of which we begin to have in early childhood when we begin to realize that our inner mental life is not present to others as it is for ourselves.³ As a result we begin to think of ourselves as in some way being or possessing a self. Self-experience then is “a term for a certain kind of essentially cognitively informed experience: experience informed by the notion…[of the self mental element]”.⁴ At this point we have identified experiences with a certain phenomenal character, where ‘phenomenal character’ refers to the “subjectively discernible manner in which phenomenally conscious states may differ among themselves”, but we cannot say which kinds of experience have this particular phenomenal character or of what this phenomenal character really consists.⁵ To proceed, an outline of how introspection enables us to form conclusions about the mental elements that produce the phenomenal characters of the experiences in question is needed.

So far what we mean when we say we have self-experience is nothing more than to say we have an experience of being a locus of consciousness, and that this experience leads us to posit the existence of a self. Given this, the following steps outline the role introspection plays in advancing our understanding of these experiences. First we reflect on which experiences are those in which we conceive of ourselves as being a locus of consciousness. According to Strawson we will find that these experiences are hardly rare. We may come to have such an experience alone and deep in thought or even in a crowd of people. What is essential to the experience is that it is one which is a form of cognitive awareness. That is to say it has more in common with a thought than say a sensory mental state. Next we turn our attention to the phenomenal character of our ordinary notion of self-experience. Strawson suggests a list of mental elements that accompany ordinary human self-experience; namely these are subject, thing, mental, single, persisting, agent, distinct, and personality. In other words, ordinary human self-experience is an experience of selfhood involving “a subject of experience, an agent, a single persisting mental thing

---

² Strawson 2009 p. 2
³ ibid p.8
⁴ ibid p. 3
⁵ Siewert 2007 p.5
that is distinct from the human being considered as a whole and possesses a certain personality". To get from the vague description to one characterized by distinct features there must have been some method of reflection that preceded our ability to offer a more detailed description. The first step in this introspection is to become aware, as Strawson has made us, of this certain kind of experience so that we may focus on it. We know that each person has had a multitude of self-experiences throughout life, so it should be easy to begin to generate self-experiences that we may introspect as well as call on past instances to see if by introspection of our memories of these experiences we might better characterize them. For now then we presume that in doing so we notice that our experience has a phenomenal character that can be described in the terms Strawson has given. That is to say we form beliefs of the form ‘This experience depends on the mental elements single, mental, subject, thing, persisting, agent, distinct, and personality’. After a certain point then we would have analyzed enough of these experiences to start to form conclusions about the typical human experience of selfhood. We then have a more exact way of describing self-experience that is far more robust than merely referring to some locus of consciousness.

The above gives an account of our ordinary self-experience. Strawson recognizes though that this ordinary account may exclude some instances of self-experience as conceivably two experiences may count as self-experiences while having different phenomenal characters. To refer back to the vague notion, this is to say that this experience of being a locus of consciousness may include some experiences that are out of the ordinary. We might suspect, for example, that an experience that had many of the features of the above list but say lacked a sense of personality could still count as a self-experience. To account for this Strawson wishes to determine what self-experience is in general, and to do so he will prune the list of ordinary features until only those features which allow us to describe any self-experience remain.

In what he calls the whittling argument Strawson proposes to reduce the list of mental elements that shape our self-experience to as few members as possible. Doing so requires a focus once again on introspection but also draws upon the testimony of others and thought experiment throughout in an attempt to bridge gaps that may exist in one’s own experiences. It is not necessary for the purpose of this paper to recount the reasons for which each mental element is either dismissed or retained by Strawson. What is important though is that he arrives at a list of four core elements, subject, thing, mental, and single, which correlates to an experience of selfhood as a subject of experience that is a single mental thing. This result is supposed to describe the phenomenology of any given instance of self-experience. Furthermore it can be assumed that this result would be a continuation of the process outlined above. Presumably you would come to hold beliefs such as ‘This experience depends on the mental elements single, mental, subject, and thing’ or failing to do so for ourselves such beliefs could be arrived at from the testimony of others. The same would hold true for the recognition of the absence of some mental elements. In introspecting we come to hold beliefs such as ‘This experience does not depend on the mental element agent’. After sufficient reflection, we would then come to make generalized assertions about the content of all self-experiences and thus what is

---

6 Strawson 2009 p.161
common to all of them. Additionally, it must be the case that through suitably diligent introspection anyone with ordinary mental capacities would arrive at this conclusion independently, as this method aims to find commonality between all self-experience.

Over the course of his work Strawson appeals to self-experience as an avenue through which we can establish a foundation for our metaphysical reasoning about the self. This is articulated in the Equivalence Thesis, which states that “selves exist if and only if there is something that has properties denoted by those [cognitive mental elements] that feature in every genuine form of self-experience”.\(^7\) Strawson’s goal then is to deconstruct self-experience to understand in what way the self mental element structures our experience of being a self. Strawson hopes to find not a simple, inscrutable self mental element but rather further mental elements that together all contribute to the formation of the self mental element. The self is figured as having certain other properties that shape our self-experience and without which we could not be said to have genuine self-experience. So identifying these as outlined previously allows for the phenomenological characterization of the phenomenal character of self-experience. Strawson hopes to use this phenomenological characterization to make a metaphysical claim: the self’s attributes, should it exist, will be identical to those common to every instance of self-experience. Strawson’s plan for his argument is to identify the thought-elements that contribute to the core content of genuine self-experience in order to determine the nature of the self. It is important to note that, in abiding by the Equivalence Thesis, he does not presuppose the existence of the self, as it could be that once the phenomenal character of self-experience is established nothing corresponds to this characterization. If however once his characterization is complete there is something that has all the properties implied thereby, Strawson argues that entity is a self and any putative self must likewise have the properties proposed by his phenomenological characterization. The Equivalence Thesis’s reliance on phenomenology however is suspect. In the following two sections issues with relying on phenomenology for our reasoning about the self will be explored.

**Section 3: Unreliable Introspection**

One way the Equivalence Thesis might be thought to fail is if introspection is unreliable. If it is then there would be no reason to think that we could understand our phenomenology well enough to allow it to dictate our characterization of the self. Here I will briefly comment on the reliability of introspection, some reasons its reliability should be questioned with respect to the phenomenology of self-experience, and ultimately why this route of critique should not be followed. Eric Schwitzgebel has argued that our introspection about experiences of emotion and thought are much less reliable than they are generally taken to be. He notes that if we introspect about our emotions we will find it hard to answer questions such as whether or not we always have the same phenomenal experience accompanying some emotion or whether or not emotions are even always experienced phenomenally. He also points to debate about whether or not thought has its own distinctive phenomenology or if it is rather exhausted by mental imagery. If introspection were reliable, there would be no debate, as introspecting the phenomenology of thought would quickly resolve the matter.\(^8\) Introspection about the phenomenology of selfhood is similar. There is a great amount of disagreement over what selves are—should there be

---

\(^7\) ibid p. 55  
\(^8\) Schwitzgebel 2008
any—and how we experience selfhood. Furthermore, we can question whether or not experiences of self are always accompanied by the same phenomenology or if as Strawson believes the experience can be quite diverse. By parity of reasoning then we have cause to view introspection about self-experience as unreliable. Strawson notes that our self-experience has a phenomenal character which by the Equivalence Thesis posits that the self must be a single, mental thing which is a subject of experience. Given our doubt about the reliability of introspection though, one might suppose that we cannot attribute any one of these properties to the self. Our phenomenology could be consistently misleading. So for instance, when our self-experience is characterized in part as experience of a thing, there may be nothing of the sort that gives rise to this experience despite there nevertheless being something else worthy of being called a self. For example the self, among other possibilities, might be attributable to some mental process, but the Equivalence Thesis would have us conclude that there is not a self as there would be no entity which aligns with our phenomenology. The other characteristics may be similarly misleading.

Strawson may try to circumvent this difficulty by noting that the self does not have to exist in his view. The Equivalence Thesis merely states that what it is to be a self is to have the properties attributed to it in our self-experience. So it might seem that our introspection cannot be incorrect, if we take phenomenology to be fundamental. Whatever it claims the features of the self are in fact the features the self must have if it is to exist. This is still problematic, however. If our introspection is not reliable then one should not specify what it is to be a self on the basis of phenomenology, for the ability or inability to find something with those features is inconsequential. There could be something worthy of being called the self and this entity could be entirely different from what our introspection reports in self-experience. Furthermore, we might suppose that our self-experience does not report all the features of the self. The self may always have some feature which is opaque to us, so that we are never aware of it. Accepting the Equivalence Thesis risks offering a misinformed description of the self.

If the above is true then there are obvious problems for Strawson’s project; unreliable introspection would undermine any project founded on being able to examine phenomenology. So much so that Strawson would have undoubtedly been aware of these dangers but decided to proceed. Moreover, it is worth noting that the impetus for this examination was inspired by cases of naive introspection. In such cases we are relying just on what our introspection tells us without much in the ways of deep reflection. Conclusions gained by naive introspection lack the benefit of being produced by a specific, systematic approach to analyzing phenomenology, instead relying on what occurs to the introspector. However, the process that was outlined above for introspecting self-experience was not a naive one. It relied on a methodical inductive examination of phenomenology revolving around experiences of being a locus of consciousness. It might even be said that the locus of conscious experience is a product of naive introspection, but then Strawson aims to move beyond this to something with a stronger foundation. So if the Equivalence Thesis is wrong, it should be wrong for other reasons.

Section 4: Problems for Introspection

---

9 Strawson argues that in some sense every thing is a process, though he does not seem to suggest that all processes are things. So the suggestion here is that some mental process that does not count as a thing could possibly be a candidate for the self.
Suppose instead that introspection is reliable. Our introspection accurately represents our phenomenology and any features of our phenomenology thereby discovered correspond to something real. Even granting this there is still a problem with Strawson’s reliance on phenomenology in the Equivalence Thesis. There could be competing minimal phenomenological characterizations of the self, and it is unclear whether or not introspection provides information about mental elements in the way that has been supposed. If introspection does not provide the proper information about phenomenology, then Strawson is not justified in asserting his minimal phenomenological characterization.

Recall that in order to form a minimal phenomenological characterization introspection on multiple locus of consciousness experiences is needed in order to become more certain as to what these experiences entailed. Doing so leads us to hold beliefs about particular experiences. Specifically we will come to hold beliefs about the mental elements that structure each experience’s phenomenal character. These will include beliefs such as ‘This experience depends on the mental elements single, mental, subject, and thing’ and ‘This experience lacks the mental element agent’. After making these judgements about many instances of self-experience we can then hope to make general judgements about the phenomenal character of any would-be self-experiences. These would be judgements such as ‘Any self-experience depends on the mental elements single, mental, subject, and thing’ and ‘Some self-experiences lack the mental element agent’. From these a minimal phenomenological characterization of the self is obtained via the Equivalence Thesis.

One may at this point note that the general judgements about self-experience depend on ampliative reasoning and so may easily be over-turned by a counter example. It could be that self-experiences that are more minimal than those espoused by Strawson’s theory or even a radically different form of self-experience has not presented itself, but may be possible. Any problems ampliative reasoning presents in general are relevant here but are not of concern. Much reasoning depends on similar inferences, though here it may seem particularly obvious. Nevertheless, these judgements should not be doubted based solely on the problem of induction.

That is not to say that such characterizations are guaranteed to be justified. Introspecting our phenomenology about self-experience undoubtedly gives knowledge about something. At the very least we can know that an experience was had, and since it is assumed that our introspection is reliable it is also quite certain that the experience has the phenomenological features that one takes it to have. Yet this alone is not enough to be confident in the resulting phenomenological characterization. It is not clear that using the methods underpinning Strawson’s Equivalence Thesis will lead everyone to the same conclusions about self-experience or to the kinds of conclusions to which Strawson arrives. Starting with the former the following will examine the possibility of disputes arising over introspection about the phenomenology of self-experience. As stated, to come to the generalized judgements concerning the content of any self-experience one must first form beliefs about individual cases. The specific beliefs can be both positive and negative in nature. In cases of positive beliefs a list of mental elements is attributed to an instance of self-experience. In the negative case mental elements absent from the current experience are identified. On Strawson’s account everyone who properly introspects should form very similar, if not identical, positive and negative beliefs about the features of their
self-experiential phenomenologies. However, suppose there are two individuals who noticing that at times they have experiences of being a locus of consciousness begin to introspect in order to gain insight about these experiences in particular. One may come to characterize the self as Strawson does, but it also seems plausible that the other might disagree. The second individual may never have had self-experiences lacking some mental element which the first individual has excluded from the resulting phenomenological characterization or has never had an experience that includes some mental element. Suppose, for example, that the latter never has any locus of consciousness experience that lacks a sense of agency, the experience of being a locus of consciousness is of an active rather than passive entity, and so never forms the belief ‘This experience lacks the mental element agent’ when having such experiences. Such a characterization would view agency as essential, while the former does not. Suppose when they have finished their examinations that the two end up with equally minimal phenomenological characterizations that disagree. In such cases where neither phenomenological characterization is more minimal than the other it is not clear which to choose on Strawson’s account. Each individual in the example has carefully examined self-experience and come to an informed conclusion about the involved phenomenology. Yet, we suppose that due to some circumstance of life events or the like both have come to different conclusions. Favoring parsimony cannot help here as each characterization is equally parsimonious. So it is plausible that even careful introspection could fail to specify what sort of thing the self is. Given the large amount of disagreement on the self, it even seems certain that introspection will end in these sorts of disagreements.

Strawson would likely reply that what he aims to do is pick only that which is common to both characterizations. He wants to know what is common to all self-experience, and so we stand to learn from the disagreement. Both might disagree on some elements, but it is very likely that they will be able to recognize shared elements between their characterizations. But this does not seem entirely right. Strawson’s approach favors sparse phenomenological characterizations without warrant. If the nature of the self is to be accessible through phenomenology, then there must be something we learn about the self when we introspect. What we learn would not be limited to commonalities between our phenomenologies, rather we should think that all the cases are telling us something about the nature of the self, at least to some degree. So considering the self to only be that which has the properties at the intersection of competing characterizations will result in a loss of information about what it is that is being analyzed. So Strawson’s Equivalence Thesis goes too far in saying that the self is just those properties. Instead it would be better to say that the self at least has the properties in our minimal phenomenological characterization. However, this then admits that phenomenology can only be so helpful in determining what the self is.

We may also question whether the two individuals are, while following the retrospective method outlined, truly examining at all times experiences which are self-experiences. Experiences of being a locus of consciousness are the starting point for introspection about self-experience, but this does not justify viewing all such experiences as self-experiences. We could have multiple visual-experiences which while all visual in nature were not all experiences of the same thing. So we might experience being a locus of consciousness with different phenomenal characters, but we should not assume that all of
these are self-experiences. If this assumption is made, one might not find what mental elements all self-experiences have in common but what all locus of consciousness experiences have in common. Introspection is able to tell us that we are having an experience and perhaps even what mental elements it has, but it cannot always tell us what kind of experience it is we are having if we do not already have an understanding external to the experience. For example, we recognize a difference between hunger pain and other abdominal pain after gaining a rudimentary understanding of nutrition at an early age. In the case of selfhood we are not already clear on what counts as a self so our phenomenology can only tell us so much. In other words it seems as though in order to call the experience of being a locus of consciousness self-experience we must already have some idea as to what we are looking for. The Equivalence Thesis aims to use phenomenology to sort out the features of the self, but this cannot be done without first having background information about what sort of phenomenal experiences should be considered.

This raises other questions about Strawson’s phenomenological project. It is not clear that the method of introspection that has been discussed thus far actually corresponds to an accurate depiction of the way we learn from introspection. Start by considering the positive beliefs formed from introspection. Thus far we have supposed that when we introspect a purported self-experience we may identify the mental elements that make up its phenomenal character. It is not clear though that this commonly occurs if at all. When we reflect on experiences of being a locus of consciousness it seems unlikely that any given mental element will be laid bare through introspection. It is more likely that we will merely recognize that an experience of being a locus of consciousness is being had. Moreover this seems true of the negative judgements we might have about some experiences. For example suppose we wanted to know whether the self is always figured in self-experience as an agent. It seems we cannot know whether a past experience’s phenomenal character lacked the mental element ‘agent’ as this neither follows analytically from the concept of being a locus of consciousness nor does it seem that we would be likely to remember past experiences in such detail as to be sure of its contents. There even seems to be a problem for our current introspection. If at this moment one tries to have a self-experience—whatever that entails—it seems challenging to name all of the characteristics it does and does not have. More than likely one would need to go through Strawson’s list and carefully reflect on whether the experience is shaped by a particular mental element. At which point this may lead to a feeling of confusion rather than enlightenment. One might suspect that the experience will be shaped by the desire to confirm or deny a certain assertion about how our phenomenology represents selfhood. It would be no easy task to ascertain whether one thinks that self-experience figures the self as an agent because that is the way such experiences always are or because a certain belief is held about the matter at this moment. So we cannot justify our judgements about the phenomenology of self-experience by reflecting on our past or current experiences.

Section 5: Conclusion

In the preceding the viability of introspection was analyzed as a method for reasoning about the nature of the self. This was done in response to Galen Strawson who argues for the existence of the self as characterized by our phenomenology. This argument rested on the Equivalence Thesis, which posits that something is a self if and only if that something has the properties attributed to it by self-experience. Strawson motivates self-
experience through appeal to the experience of being a locus of consciousness. Consequently to understand self-experience one must introspect about these locus of consciousness experiences in order to learn in what they consist. By reflecting on these experiences we form beliefs about what mental elements do and do not contribute to their phenomenal characters. At the end we have a list of mental elements that give structure to our self-experiences. Taking note of the mental elements common to all self-experiences and applying the Equivalence Thesis results in a minimal phenomenological characterization of the self.

However, it seems the conclusions arrived at through this method are not justified. The investigation began by examining locus of consciousness experiences as they appear to underlie our early understanding of selfhood. By introspecting these experiences there is a risk over generalizing our phenomenological characterization of the self, as it is possible that every self-experience is a locus of consciousness experience without every locus of consciousness experience being a self-experience. Furthermore, it may be doubted that introspection reveals information about mental elements in the way Strawson’s work suggests it must. The Equivalence Thesis assumes that after examining phenomenology the mental elements structuring our self-experience can be determined. However, actual experience of selfhood does not seem to provide this information. We may become aware of being a locus of consciousness, but this is not accompanied by the belief that the experience is one of being a single mental thing which is a subject of experience nor that the experience lacks the mental element of personality for example. If one does introspect with the intent to determine the presence or absence of some mental element this may further obfuscate the introspection as one’s personal beliefs about the self may have an effect on introspection. As such, the Equivalence Thesis should not be used in characterizing the self as the introspective methods that it must rely on cannot operate in the manner suggested by Strawson’s thesis.
References