1-1-2016

Berkeley on the Source of Self-Knowledge: Introspection and Causal Maxim

Scott Harkema
University of Mississippi, scott.harkema@gmail.com

Follow this and additional works at: https://egrove.olemiss.edu/etd

Part of the Philosophy Commons

Recommended Citation
https://egrove.olemiss.edu/etd/1211

This Dissertation is brought to you for free and open access by the Graduate School at eGrove. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.
BERKELEY ON THE SOURCE OF SELF-KNOWLEDGE: INTROSPECTION AND THE
CAUSAL MAXIM

A Thesis
Presented in partial fulfillment of requirements
For the degree of Master of Arts
In the Department of Philosophy
University of Mississippi

by
Scott Harkema
Spring 2016
ABSTRACT:
Most Berkeley commentators agree that Berkeley’s theory of self-awareness depends on some type of direct introspective access to the self. In this paper, I challenge this consensus view, arguing that Berkeley’s theory does not claim that there is direct introspective access to the self until after his first publication of the *Principles of Human Knowledge* in 1710. The first edition of the *Principles*, as well as Berkeley’s *Philosophical Notebooks*, reveal a significantly different, perhaps more “Humean,” perspective concerning self-awareness than his works after 1710. During this period, Berkeley thought that the self cannot be encountered directly through introspection, but is in fact knowable only by means of an inference which integrates a crucial causal maxim. Further, I argue that Berkeley thought the causal maxim which grounds his argument for the existence of the self is itself grounded in experience. Berkeley’s early position on self-knowledge interestingly anticipates Hume’s criticism of the introspective availability of the self while denying any skepticism concerning causation.
DEDICATION:

To the Late George Berkeley, Bishop of Cloyne

"...we have first raised a dust, and then complain, we cannot see." (Introduction, 3)
ACKNOWLEDGMENTS:

Special thanks are in order for Dr. Timothy Yenter, Dr. Donovan Wishon, and Dr. Steven Skultety for helpful comments and assistance throughout the writing process. A version of this paper was presented at the Midsouth Philosophy Conference in February 2016.
TABLE OF CONTENTS

ABSTRACT: ........................................................................................................................................ II

DEDICATION: .................................................................................................................................... III

ACKNOWLEDGMENTS: ......................................................................................................................IV

THESIS: ........................................................................................................................................... 1

BIBLIOGRAPHY ................................................................................................................................. 32

CURRICULUM VITAE ....................................................................................................................... 35
THESIS

Theories of self-knowledge that argue for a substantial self distinct from material bodies often rely upon some type of special, direct introspective access to one’s own self, mind, or soul. The theories diverge with regards to exactly how the self is encountered in introspection, but all share in that they ground self-knowledge in some type of direct introspective awareness of the self. For instance, according to some interpreters, Descartes claims that one can be acquainted with the self as an object of the understanding. Russell also at one point considered that there might be acquaintance with the self, though he ultimately rejected the position. Others like Malebranche endorse a theory where the self is known by an “internal consciousness” of oneself as the subject of thought or action. These types of accounts all share the claim that the self can be encountered directly through introspection; a mere act of turning the mind towards itself yields a direct apprehension of the self, which justifies knowledge of the existence of the self.

But many philosophers, myself included, are skeptical of such claims to introspective access to the self, and therefore remain dubious of such accounts of self-knowledge. For many, Hume’s attack on these types of attempts to find the self through introspection motivates this doubt. Whatever substantial, distinct self that is commonly supposed to exist mysteriously eludes us.

---


2 Russell was considering this possibility through 1912 with his publication of the Problems of Philosophy. However, by 1913, his views had changed; his unpublished manuscript from that time shows him claiming that the self can only be known by description and not by any type of acquaintance. This 1913 position is in many ways similar to the position I argue Berkeley accepts from 1708-1710.
whenever we search for it, even to the very limits of our experience. Where we expect a clear and determinate experience of ourselves, we are instead confronted by the absence of any such experience. This surprising absence often leads empiricists to a skeptical or deflationary stance towards the self.3

However, there is one potentially surprising exception to the conventional empiricist response to the non-introspectability of the self: George Berkeley. Though as his career progresses his position waxes Malebranchean, Berkeley’s early position occupies a unique position which denies any introspective awareness of the self and yet affirms the existence of a substantial self (spirit) distinct from the objects of experience (ideas). In this way, Berkeley actually anticipates Hume’s criticism of the introspective theories of self-knowledge, and develops a response that adheres to this Humean criticism while making a case for the existence of a substantial and distinct self.

In this paper I argue that in his early Philosophical Notebooks (c. 1708) as well as his first edition of the Principles of Human Knowledge (1710), Berkeley maintained that the self cannot be known by some special introspective awareness, but rather can only be inferred based on one’s experience of the objects of perception and thought. This inference involves a type of causal maxim, which Berkeley intriguingly also attempts to ground in experience. If this causal maxim is true, then I think Berkeley’s argument for the self is in fact sound, and thus one could indeed claim to know that she/he exists without having any special direct introspective access to the self. And while the causal maxim and Berkeley’s attempt to ground it in experience is riddled with difficulties, Berkeley’s early theory of self-knowledge may be useful to any empiricist who like

3 There are many relatively non-empirical responses to Hume’s criticism of introspection which accept that self-awareness is not direct and yet affirm that there is indirect awareness of the self. These accounts usually rely on a principle like “every thought has a thinker,” which is not justified empirically. For a robust account of many current accounts of self-awareness, see Brie Gertler, Self-Knowledge, (London: Routledge, 2010), pg. 208-252.
Hume denies that there is any direct experience of the self, yet is not so skeptical concerning the causal maxim.

In §1, I give a sketch of my interpretation of Berkeley’s theory of mind in the *Notebooks* and first edition of the *Principles*. I pay particular attention to his denial that there can be an idea of the self and his affirmation of the existence of a distinct, active self. In §2, I discuss the common contemporary interpretations of Berkeley’s theory of self-knowledge, which allege that Berkeley thinks the self is known through a direct introspective awareness of the self. In §3, I argue that these interpretations are deficient because they do not recognize the important shift in Berkeley’s theory between his 1710 publication of the *Principles* and 1713 publication of the *Dialogues*. In §4, I delineate what I take to be Berkeley’s early argument for the existence of the self, which relies on a causal maxim. I resolve some difficulties with Berkeley’s argument by modifying his notion of causation in a way that is consistent his theory of mind discussed in §1-3. In §5, I discuss Berkeley’s grounding of the causal maxim with regards to Humean worries about causation. I suggest that Berkeley actually believed (at least in his early period) that causation is directly experienced, though not by way of idea.

**Berkeley’s Theory of Mind**

In this section, I consider Berkeley’s theory of mind as it develops in the *Notebooks* to its restatement in the *Principles*. I focus on two key characteristics of Berkeley’s theory of mind; first, anticipating Hume, Berkeley denies that there can be any direct apprehension of the self *via* idea, and second, Berkeley affirms the existence of an active being that is itself distinct from the objects of its experience. This being he calls spirit, soul, or mind, and he identifies the self with it.

Berkeley’s apparently Humean remarks appear in his early *Philosophical Notebooks*. Here, Berkeley offers a psychological description of attempts to introspect the self, in a way that is strikingly similar to Hume’s famous remarks in *Treatise* 1.4.6.3:
Consult, ramsack yr Understanding wt find you there besides several perceptions or thoughts. Wt mean you by the word mind you must mean something that you perceive or yt you do not perceive. a thing not perceived is a contradiction. to mean also a thing you do not perceive is a contradiction. (Notebooks 579)

The argument in this passage can be summarized as follows:

1. If $x$ exists, $x$ can be perceived (i.e. there can be an idea of $x$).
2. Therefore, if the mind exists, then the mind can be perceived (i.e. there can be an idea of the mind).
3. There can be no perception of a mind (i.e there can be no idea of a mind).

Therefore, the mind does not exist.

Following this argument Berkeley redefines mind in a way that makes it consistent with (1) and thus (2):

- Mind is a congeries of Perceptions. Take away Perceptions & you take away the Mind put the Perceptions & you put the mind. (Notebooks 580)
- Say you the Mind is not the Perceptions. but that thing perceives. I answer you are abus’d by the words that & thing these are vague empty words without a meaning. (Notebooks 581)

This bundle theory temporarily saves the concept of the mind for Berkeley. But Berkeley’s positivism wanes as the Notebooks progress, and soon he denied premise (1) and (2) of the argument he makes in 579. But (3), which claims that the mind is not perceivable, Berkeley continues to affirm throughout his philosophical career.

---

4 There is some debate concerning whether or not Berkeley is actually endorsing a bundle theory in these passages, with most of the controversy revolving around whether or not Berkeley’s “+” sign in the margins indicated that he later rejected the view. Stephen Daniel has argued that these entries do not endorse a bundle theory and are in fact consistent with Berkeley’s mature theory of spirit. His theory appears most recently in “Berkeley’s Doctrine of Mind and the ‘Black List Hypothesis’: A Dialogue,” Southern Journal of Philosophy, 51, no. 1 (Mar 2013): 24-41. An example of a more traditional interpretation of Berkeley as a bundle theorist appears in Bertil Belfrage, “Berkeley’s Four Concepts of the Soul (1707-1709),” in Reexamining Berkeley’s Philosophy, ed. Stephen Daniel (Toronto, University of Toronto Press, 2007), pg. 172-187.
Berkeley’s retreat from the positivism of premise (1) began when he affirmed the existence of the will, despite its lack of appearance in one’s perceptions.

The grand Cause of perplexity & darkness In treating of the Will, is that we Imagine it to be an object of thought,..., we think we may perceive, contemplate & view it like any of our Ideas whereas in truth ‘tis no idea. Nor is there an Idea of it. tis toto coelo different from the Understanding i.e. from all our ideas. (Notebooks 643)

Berkeley here affirms the existence of something (the will) which is not itself an idea and cannot be found in the understanding. So while he maintained that the will is not introspectively available, he no longer thought that this was a reason to reject the existence of the will. In other words, he denied the first premise of his earlier argument, that “If x exists, x can be perceived (i.e. there can be an idea of x).” This anti-positivist turn remains with Berkeley throughout his philosophical career, though the being that exists beyond perception/experience changes as his views develop.

After turning from positivism by affirming the existence of unperceived will, Berkeley identified “spirit” or active being exclusively with the will:

The Spirit the Active thing that which is Soul & God is the Will alone[.] The Ideas are effects impotent things. (Notebooks 712)

But he later changed his mind, identifying spirit with “all that is active,” and this includes not willing alone but also the acts of perceiving and understanding. This willing, perceiving, thinking being Berkeley calls the mind, soul, spirit, or self:

One argumentative move that may confuse some readers is Berkeley’s move from the claim that the will is not an idea to the claim that there can be no idea of the will. As he puts it, “in truth ‘tis no idea. Nor is there an Idea of it.” To us, this might appear downright fallacious. Why does Berkeley think we cannot have an idea of something that is not itself an idea? The answer, as the seasoned reader of Berkeley might suspect, lies in Berkeley’s likeness principle, which he appeals to in Notebooks 657: “To ask have we an idea of ye Will or volition is nonsense. an idea can resemble nothing but an idea.” (See also Notebooks 684) For more on why Berkeley thinks we should accept the likeness principle, see Kenneth Winkler, Berkeley: an Interpretation (Oxford: Clarendon Press, 1989), pg. 137-149.
S. But the Grand Mistake is that we know not what we mean by we our selves, our mind &c. tis most sure & certain that our Ideas are distinct from the Mind i.e. the Will, the Spirit. (*Notebooks 847*)

S. I must not Mention the Understanding as a faculty or part of the Mind, I must include Understanding & Will &c in the word Spirit by which I mean all that is active. I must not Say that the Understanding differs not from the particular Ideas, or the Will from particular Volitions. (*Notebooks 848*)

S. The Spirit, the Mind, is neither a Volition nor an Idea. (*Notebooks 849*)

Here Berkeley clearly denies any possibility of a bundle theory of the will, understanding, or mind. Instead he posits a sort of dualism; on the one hand, there is an active being which wills and understands, which is called spirit, and on the other hand, there are “objects of experience”, that is, passive ideas. And while he is not advocating a dualism of substance, Berkeley does claim that it is “most sure and certain that our ideas are distinct from the mind.”6 So, although at one point in the *Notebooks* he endorsed a bundle theory of the mind, Berkeley later rejects this and affirms that the mind is a distinct, active being, one that can neither be an idea nor be represented by an idea.

Berkeley continued to affirm both that the self is not introspectable and that it is distinct from ideas in his first edition of the *Principles*. First, regarding the notion that the mind or spirit is a being distinct from its objects, Berkeley writes:

> But besides all that endless variety of ideas or objects of knowledge, there is likewise something which knows or perceives them, and exercises divers operations, as willing, imagining, remembering about them. This perceiving, active being is what I call mind, spirit, soul or myself. By which words I do not denote any one of my ideas, but a thing entirely distinct from them, wherein they exist... (*Principles 2*)7

---

6 Berkeley thinks spirit is the only substance; ideas are not substances because they are wholly dependent upon spirit for their existence.

7 References to all of Berkeley’s works other than the *Philosophical Commentaries* are from: George Berkeley, *The Works of George Berkeley Bishop of Cloyne*, ed. A. A. Luce and T. E. Jessop (London: Thomas Nelson and Sons Ltd, 1949). References to the *Principles* include the section number; references to *De Motu* and the *Dialogues* include page number.
The remarkable resemblance this passage bears to Berkeley’s late remarks in the *Notebooks* indicates that Berkeley likely wrote this section of the *Principles* with these entries of the *Notebooks* in mind. At the very least, it should be clear that Berkeley’s notion of mind did not change between the late remarks in the *Notebooks* and the first edition of the *Principles*.

It is also clear that Berkeley continued to maintain that there can be no idea of the mind, and therefore that it cannot be found through introspection. Again, he writes:

...there can be no idea formed of a soul or spirit... so far as I can see, the words will, soul, spirit, do not stand for different ideas, or in truth, for any idea at all, but for something which is very different from ideas, and which being an agent cannot be like unto, or represented by, any idea whatsoever. (*Principles* 27)

Again, here we see Berkeley claiming that to have an idea of spirit is impossible. Thus, Berkeley accepted his Humean position from the *Notebooks* concerning the non-introspectability of the self at least as late as his first publication of the *Principles*.

Berkeley’s notion of spirit in the *Notebooks* and *Principles* is thus in one respect Humean; his denial of any direct experience of ourselves or of spirit in general anticipates Hume’s critique of the introspectability of the self. But nonetheless he affirms that the self is an active being distinct from the objects of perception, and this brings up the question, “how in fact can we know that there is an active, distinct being that is the self?” Hume and most others who assent to his critique of the introspectability of the self deny the existence of a distinct being that is the perceiving, thinking, and willing self. Is Berkeley being inconsistent? Or does he have an argument for the existence of the self and spirit that does not depend on the introspectability of the self? Most commentators agree that Berkeley’s response to this question is not consistent with non-

---

8 Many of the passages from the *Principles* mirror entries from the *Notebooks*, from which many commentators have inferred that Berkeley used these *Notebooks* as a reference when writing the *Principles*, and maybe even the *Dialogues* and *De Motu*. At the very least, it seems Berkeley did not write the *Notebooks* and then leave them alone.

9 See also *Principles* 89, 135, 138, 139, and 142.
introspectability of the self; they claim that Berkeley upholds some form of direct awareness of the self, upon which our knowledge of the existence of the self is founded.

Berkeley on Self-Knowledge: Traditional Interpretations

The opinion that Berkeley’s theory of self-knowledge throughout his mature period (1710-1734) relies on some special first-personal access to the self exists even among his best commentators. While there are various types of positions attributed to Berkeley ranging from acquaintance theories to inner-sense theories, all share in that the self is perceived or felt “directly,” and on this basis we can be said to know many things about the self, including most importantly that it exists. For the sake of brevity, here I consider only one such interpretation, that of Kenneth Winkler. I have chosen Winkler’s interpretation only because I think it is the best and most textually sensitive interpretation, and not because I have some particular quarrel with his view which I do not have with others. Winkler’s interpretation is by and large correct, except for his small error of omission, which every other commentator also appears to make. My criticism of Winkler’s interpretation should apply equally to any other theory which supposes that Berkeley holds an introspectivist position regarding self-knowledge.

\[10\]

This includes Winkler, Dicker, Adams, Cummins, and as far as I know, anyone else who has written about Berkeley’s theory of self-knowledge. Bennett is actually the only one who appears to challenge this perspective. He claims, “there are also indications that [Berkeley] thinks that his own existence is inferred from ideas which are immediately given, through the principle that there cannot be unowned ideas.” See Jonathan Bennett, *Learning from Six Philosophers*, vol. 2 (Oxford: Clarendon Press, 2001), pg. 156.

\[11\]

Winkler argues that Berkeley held that the self is known through “subject-consciousness” rather than “object-consciousness.” The self is experienced as the subject of actions, rather than as an object of one’s understanding. Knowledge of the self derives from this first-person experience of oneself as the subject or performer of these acts. Thus the self is presented or known in a way that is immediate and non-inferential; to put it in popular contemporary terminology, it is “epistemically direct.” We are not simply aware of certain things from which we infer the existence of ourselves; rather, we are directly aware of the self itself, with no inferential apparatus necessary.

The evidence for this position, though somewhat sparse given Berkeley’s rare treatment of the topic of self-knowledge, is very compelling. In the Principles, Berkeley claims that “We comprehend our own existence by inward feeling or reflection, and that of other spirits by reason” (Principles 89, second edition). Philonous in the Dialogues claims,

I who am a spirit or thinking substance, exist as certainly, as I know my ideas exist. Farther, I know what I mean by the terms I and myself; and I know this immediately, or intuitively, though I do not perceive it as I perceive [an idea]. (Dialogues 231)

...the being of my self, that is, my own soul, mind or thinking principle, I evidently know by reflexion. (Dialogues 233)

Finally, in De Motu, Berkeley claims that “the sentient, percipient, thinking thing we know by conscientia quadam interna [a certain internal consciousness]” (De Motu 21 [latin], 36

---

12 The scarcity of Berkeley’s remarks about the nature of the mind is likely due to Berkeley’s loss of a manuscript in which he was planning to disseminate his theory of mind. He claims in a letter to his friend Samuel Johnson in 1729 that he “had made a considerable progress in it, but the manuscript was lost [in 1715] during [his] travels in Italy,” and he never rewrote and published what was lost (Philosophical Correspondence with Samuel Johnson, pg. 282). He does remark on numerous occasions in the Notebooks his plans for a second and even third manuscript (which were to be part 2 and 3 of the Principles). The second’s subject matter was to contain his theory of mind and ethics, while the third was to contain his natural philosophy. The third was likely reworked into his 1721 De Motu, after Berkeley abandoned his plans for the tripartite Principles. See Notebooks 508, 583, 738, 792, 807 and 878 for references to his intentions for the tripartite Principles.
All these seem to imply that there is some special, direct awareness of the self that is unlike our apprehension of the objects of thought or perception. Winkler proposes that in all these passages, Berkeley is suggesting that this direct awareness of the self is subject self-awareness. This means:

> The mind itself [is not] an object that we contemplate and view, but a thing known through acting, operating, or performing... to know spirit is to experience oneself as acting, willing, and operating - to experience oneself as the subject of acts, rather than as an object.\(^{14}\)

Winkler is suggesting that Berkeley thought that experience or perception does not only include the objects of the understanding, but also the self.

If the self is known through an epistemically direct subject self-consciousness, does this mean that the self is in fact introspectable? If yes, then my claim that Berkeley did not accept the introspectability of the self in his early career conflicts with Winkler’s interpretation.\(^{15}\) But to see if subject self-consciousness does in fact imply the self is introspectable, the definition of introspection must first be clarified. Ordinary definitions of introspection can be grouped into roughly two different definitions:

- **Introspection\(^1\):** The act of observing all present ideas (representations, objects of the understanding).

- **Introspection\(^2\):** The act of observing all present experience (representational and non-representational).

The former version is about only representational ideas, or those ideas of purely sensory content. The latter version includes these ideas, but also takes note of other objects of experience, like passions or feelings, volitions, the passing of time, etc. These phenomena are still a part of the

---

\(^{13}\) This section is loosely adapted from Winkler 2011, pg. 231.
\(^{14}\) Winkler 2011, pg. 230.
\(^{15}\) Of course, Winkler’s interpretation doesn’t have to be entirely correct to put pressure on my view, in fact, probably any other Berkeley commentator’s version will do. Most if not all commentators affirm that Berkeley thinks our awareness of the self is at least epistemically direct, if not metaphysically direct, and for this reason, they put pressure on my claim that Berkeley thinks the self is non-introspectable.
general character of experience, but they are non-representational; one cannot frame a representation of them in the mind as one does with simple and complex ideas. So the latter definition is broader than the former. The former definition includes only epistemically direct representations, whereas the latter definition includes all epistemically direct experience, representational or otherwise.

If we adopt the former definition of introspection, then it follows that Berkeley clearly denies the introspectability of the self. For he continues to insist throughout his mature career, as he did in the Notebooks, that “all ideas... cannot represent unto us, by way of image or likeness, that which acts” (Principles 27). But if this is all we mean by introspection, then Berkeley may still affirm that there are non-representational ways of encountering the self directly while upholding the non-introspectability of the self. Winkler’s interpretation fits this description: subject self-awareness is a non-representational but still epistemically direct means of self-knowledge. Thus, if by introspection we mean introspection in this former sense, Winkler’s Berkeley can claim that the self is not introspectable and yet maintain that the self is known directly.

However, it must be admitted that if introspection is defined in this former sense, this definition would not be expansive enough to satisfy Hume. For, when denying the introspectability of the self, Hume does not simply call to attention his representational ideas, he instead claims he cannot find the self among all his impressions, including “Pain and pleasure, grief and joy, passions and sensations” (Treatise 1.4.6.2) and again, “heat or cold, light or shade, love or hatred, pain or pleasure” (Treatise 1.4.6.3).16 These are no mere mental pictures (to use the Wittgensteinian term) or representational ideas; these are acts, or at the very least they

contain some active content. Putatively, they tell us something about what the mind is doing, not just what is happening to it. And even among these things, which are known not just through sensory perception, but rather through some other means of awareness, Hume claims he can not find himself. Therefore, Hume affirms the non-introspectability of the self not just on the former, narrower definition of introspection, but even on the latter, broader definition. He does not just think that the self is not representationally available, it is not experientially available. On this definition of introspection, Winkler’s Berkeley would clearly be violating the non-introspectability of the self. For Winkler’s Berkeley believes that the self is experienced directly, although not representationally.

Winkler’s Berkeley can be seen as upholding the non-introspectability of the self only in the former sense of introspection, and not the latter. But I nonetheless contend that Berkeley, like Hume, believed the self was not introspectable in both senses of introspection, though only during the early part of his career (1708-1710). I defend my interpretation in the next section.

A Shift Regarding Self-Knowledge

In this section, I aim to show that in the Notebooks and Principles, Berkeley thought that the self could be known by its effects alone, not directly by introspection. It is my contention that the view that is commonly attributed to Berkeley, one which like Winkler’s entails that the self is at least introspectable in some sense, only applies to Berkeley beginning in 1713, when he published the Dialogues. So, on my interpretation, Berkeley shifts his position regarding knowledge of the self from his early period (Notebooks [c. 1708] and Principles [1710]) to his middle period (beginning with Dialogues [1713]). The early view represents Berkeley in a somewhat more Humean light, while the later view is more Lockean or Malebranchian.

The leading piece of evidence for my interpretation comes from Principles §27. Here Berkeley claims:
A spirit is one simple, undivided, active being... hence there can be no idea formed of a soul or spirit: for all ideas whatever, being passive and inert, vide Sect. 25, they cannot represent unto us, by way of image or likeness, that which acts. A little attention will make it plain to anyone, that to have an idea which shall be like that active principle of motion and change of ideas, is absolutely impossible. Such is the nature of spirit or that which acts, that it cannot be of itself perceived, but only by the effects which it produces. (Principles 27)

Particularly the last sentence of this passage supports my interpretation. Berkeley’s claim that spirit “cannot be of itself perceived, but only by the effects which it produces” seems to imply that Berkeley does not think there is any epistemically direct experience of any spirit, including the spirit that is oneself. Rather, it is “perceived by the effects which it produces,” meaning that the self is inferred from its effects, or that which it causes.

Obviously such an interpretation contradicts Winkler’s interpretation, or any other which supposes some direct apprehension of the self. Winkler recognizes this, claiming:

If to say that spirit cannot itself be “perceived” is to say that it cannot itself be experienced, my proposal must be on the wrong track. But Principles 27, so understood, is flatly incompatible with Berkeley’s repeated declarations to the contrary.17

Winkler then attempts to reinterpret 27 in such a way that makes it compatible with his interpretation, claiming that “the main burden of Principles 27 is to argue that there can be no idea of soul or spirit. This is not to deny that it can be experienced or known in a way ideas are not.”18

Winkler’s motivation for reinterpreting Principles 27 is to make it consistent with what he calls “Berkeley’s repeated declarations to the contrary.” By this Winkler appears to be referring to the evidence I presented in §2, compiled from the Principles, Dialogues, and De Motu. All these passages claim that the self or spirit is known by reflexion or conscientia, which I am convinced imply that Berkeley, at the time of writing these passages, believed the self is experienced directly,

17 Winkler 2011, pg. 234.
18 Ibid.
though not by way of representation. But almost every Berkeley commentator fails to note a fundamental aspect about the dating of all these endorsements of reflexion or conscientia. None of these passages about reflexion or conscientia appear in the first edition of the Principles. The only passage in the Principles that endorses such a view is in Principles 89, but was added into the second edition, published 24 years later in 1734. The first edition of the Principles contains absolutely no mention of the self being known by reflection, inward sense, or conscientia; they only began to appear in the Dialogues, in 1713. So even if Berkeley did advocate for some type of direct experience of the self in 1710, he never mentions this in the Principles.

This alone should be enough to make us question the fairness of using Berkeley’s later remarks (1713-1734) to reinterpret his earlier remarks. But there is even more evidence that suggests Berkeley did not yet accept the reflection/conscientia theory until later:

S. De Vrie will have it that we know the Mind as we do Hunger[,] not by Idea but sense or Conscientia[,] So will Malbranch. This is a Vain distinction. (Notebooks 888)¹⁹

Here we have a passage that was very likely written after August 28th, 1708, where Berkeley explicitly denounces the inner sense theory of Malebranche and Gerard De Vries.²⁰ This is a direct contradiction with his 1721 claim in De Motu that spirit is known by “a certain internal conscientia.”

To make the progression more clear, consider the following timeline:

¹⁹ See also Notebooks 724, also a reference to (and criticism of) Malebranche’s theory. Also, Principles 136 is the section of the Principles indebted to this entry.

²⁰ On the very same page where Berkeley writes his final philosophical entries at the bottom (including entry 888, the last entry of the Notebooks), above these Berkeley wrote “August 28th. 1708 the Adventure of the Shirt.” (See Thomas, “Editor’s Introduction” to Philosophical Commentaries, page x) The philosophical entries at the bottom are written “small and crowded” into the space below the comment regarding the “adventure of the shirt,” indicating that Berkeley had written the “adventure of the shirt” when there was plenty of space on the page, only for it to become crowded when he later filled it with the philosophical entries. Thus, it is likely, though not certain, that these entries were written after August 28th, 1708.
1708: Berkeley denounces inner-sense/conscientia as a means for self-knowledge in the Notebooks.

1710: Berkeley never mentions inner-sense/conscientia/reflectio as a means for self-knowledge in the Principles. He does say things that are incompatible with inner-sense/conscientia/reflectio in 25, 27, and 136.

1713: Berkeley first endorses “reflection” and “reflex-act” as the means of knowing spirit or self in the Dialogues.

1721: Berkeley endorses that spirit or self is known by “internal conscientia” in De Motu.

1734: Berkeley endorses “reflection” or “inward feeling” as the means of knowing the self in the second edition of the Principles and Dialogues.

It seems to me that the most reasonable explanation of Berkeley’s theory of self-knowledge is not to assume, as Winkler and others have previously done, that Berkeley affirmed the reflexion/conscientia theory throughout his published works. Rather, a better explanation portrays Berkeley as undergoing a shift between 1710 and 1713. For there is no indication of it in the 1710 edition of the Principles (and even some contrary evidence), as well as evidence that less than two years prior, he explicitly criticizes the conscientia theory. Therefore, I conclude that Berkeley did not think that spirit or self was to be known by reflexion/conscientia in the first edition of the Principles, and that if in fact there is to be knowledge of the self or spirit, he thought it must come by way of inference from spirit’s effects.

Berkeley’s Causal Arguments from the “Effects”

The task at hand, then, is to discover from the first edition of the Principles just how Berkeley thought spirit or self is known by its effects. The theory must not rely upon any first-personal special access to the self through introspection; for to appeal to this kind of evidence would be to violate the claim that spirit is not known by anything other than its effects.

Berkeley’s argument for the existence of spirit appears in Principles §25-33. In Principles 25, Berkeley begins by inviting his reader to participate in a thought experiment concerning our ideas:
All our ideas... are visibly inactive, there is nothing of power or agency included in them... To be satisfied of the truth of this, there is nothing else requisite but a bare observation of our ideas. For since they and every part of them exist only in the mind, it follows that there is nothing in them but what is perceived. But whoever shall attend to his ideas, whether of sense or reflection, will not perceive in them any power or activity; there is therefore no such thing contained in them. (*Principles* 25)

Jonathan Bennett dismisses this thought experiment, because he does “not know what thought-experiment [he is] being invited to perform.”\(^{21}\) He asks, “How do I go about looking for activity in my ideas?... [and] why should I infer that this holds for all ideas always?”\(^{22}\) The latter question is easily dealt with\(^^{23}\), but the former question is formidable since Berkeley does not explicitly define activity in *Principles* 1-25. But shortly after, in *Principles* 28, he does offer a characterization of activity based on experience:

> It is no more than willing, and straightway this or that idea arises in my fancy: and by the same power it is obliterated, and makes way for another. This making and unmaking of ideas does very properly denominate the mind active. Thus much is certain, and grounded on experience... (*Principles* 28)

And in *Principles* 53, Berkeley equates activity with efficient causation:

> [There are some who] will have God alone to be the immediate efficient cause of all things.\(^{24}\) These men saw that amongst all the objects of sense, there was none which had any power or activity included in it... (*Principles* 53)

These passages seem to suggest that a thing is active only insofar as it causes ideas.\(^{25}\) But while in *Principles* 53 Berkeley calls this “efficient causation”, his own notion of causation never really

\(^{21}\) Bennett 2001, pg. 159. Bennett, rejecting the thought experiment because of its apparent lack of set-up, proposes that Berkeley thinks the passivity of ideas can be known *a priori*. I, however, see no reason to deny that Berkeley thought his thought experiment was decisive, and thus think there is no reason to find an *a priori* argument.

\(^{22}\) Ibid.

\(^{23}\) By inviting us to examine our ideas, Berkeley uses “idea” in a broad sense. This would include ideas of memory. Thus, if we consider a sample of ideas from our current perceptions or past memories, we will find that every single one is passive in exactly the same way. To inductively infer that all ideas will always be passive would seem a very plausible move to Berkeley.

\(^{24}\) Certainly Berkeley considered Malebranche to be a member of this group. He may well have, after publication, considered Newton to be a member of this group as well, but Newton’s “General Scholium” to his *Principia* was added in the 1713 edition.

\(^{25}\) See also *Notebooks* 830-831.
distinguishes between efficient and final causation. For Berkeley, the causal status of a being with regards to any idea depends on whether the idea follows from the being’s will. We can define causation for Berkeley like this:

**BERKELEY’S CAUSE:** $S$ causes $x$ iff (1) $S$ is a subject that wills that $x$, and (2) $x$ occurs as a direct result of $S$ willing that $x$.

This notion of a cause is primarily one of efficient causation, for it requires that $S$ is the principle of change in the realization of $x$. It also, however, sneaks in some type of final causation, in that for something to be a cause it must will that $x$ occur. This willing, presumably, cannot be purely random; the very nature of willing requires that there is a reason for $x$, an end for which it is done (at least Berkeley appeared to think this was the case). So Berkeley’s theory of causation, at least here, appears to blend together notions of efficient and final causation.

I do not mean to defend Berkeley’s strict notion of causation; in fact, I find it to be one of the weaker areas of his philosophy. But it should be clear that given Berkeley’s identification of activity with causation, Berkeley is right that one will not find any such thing among his/her ideas. Berkeley expects, and I think rightly so, that we do not perceive causes as we perceive our ideas; causes elude us in our sensory phenomenology. So, since we do not have ideas of causal powers, we may properly denominate our ideas “inactive,” or rather “passive.” Further, since for Berkeley an idea can only represent another idea, it follows that we cannot have an idea of (i.e. an idea that represents) an active thing. So all of our phenomenology, *qua* idea, is inactive, nor can it represent to us an active thing.

---

\textsuperscript{26} The claim that “an idea can only represent another idea” is an instantiation of Berkeley’s likeness principle. I am of the opinion that if we accept a strictly sensory definition of an idea, then Berkeley’s likeness principle is true. See footnote 6 for more on Berkeley’s controversial likeness principle.
Having made the case that ideas are wholly passive and thus can neither be nor represent causes, Berkeley makes his argument for the existence of spirit:

We perceive a continual succession of ideas, some are anew excited, others are changed or totally disappear. There is therefore some cause of these ideas whereon they depend, and which produces and changes them. That this cause cannot be any quality or idea or combination of ideas, is clear from the preceding section. It must therefore be a substance... the cause of ideas is an incorporeal active substance or spirit. (*Principles* 26)

There are actually two similar yet separable arguments being offered in this passage, one from change of ideas and one from the very existence of ideas:

**Argument from the Change of Ideas:**

1. There is a change (i.e. succession) of ideas.
2. If there is a change of ideas, there must be a cause of this change in ideas.
3. Ideas cannot be causes.
4. Therefore, the cause of this perceived change in ideas cannot be something that is itself an idea.

Therefore, there must be something non-ideal that causes the perceived change in ideas (and this we call spirit).

**Argument from the Existence of Ideas:**

1. There are ideas.
2. If there are ideas, there must be a cause of these ideas.
3. Ideas cannot be causes.
4. Therefore, the cause of the existence of these ideas cannot be something that is itself an idea.

Therefore, there must be something non-ideal that causes the ideas to exist (and this we call spirit).
Both arguments share the same general form: they argue from (1) an observed feature about ideas along with (2) a causal principle and (3) the causal inefficacy of ideas to the final conclusion that there is an active thing that is not itself an idea. Premise (3) was already discussed, in Berkeley’s claim that we do not have ideas of causes, and thus ideas cannot be causes. Premise (2) in each is obviously the most controversial and problematic, and is the premise with which Hume would surely disagree. But before discussing the importance of (2), I want to flag another problem with these arguments. Both arrive at the conclusion that spirit exists in a causal relation to ideas, but this conclusion contains no other information about spirit. There is a significant gap between the notion of spirit argued for here, and Berkeley’s definitions of spirit elsewhere in the *Principles*:

For by the word spirit we mean only that which thinks, wills, and perceives; this, and this alone, constitutes the signification of the term. (*Principles* 138)

This perceiving, active being is what I call mind, spirit, soul, or myself. (*Principles* 2)

A spirit is one simple, undivided, active being: as it perceives ideas, it is called the understanding, and as it produces or otherwise operates about them, it is called the will. (*Principles* 27)

In all these passages, Berkeley defines spirit in the same way he does at the end of the *Notebooks*; recall his claim that “I must include Understanding & Will [etc.] in the word Spirit by [which] I mean all that is active” (*Notebooks* 848). But the conclusion of the two arguments includes only activity in the narrower sense of willing alone; it disregards these other important aspects of active being that are fundamental to Berkeley’s (and most other’s) concept of what the self is. Therefore, even if the above arguments are sound, they do not offer knowledge of spirit as a perceiving and thinking being, but rather only as a willing being.

I do not think that this problem devastates Berkeley’s proposed theory of how the self is known using the causal maxim; rather, I only think it indicates that Berkeley’s strict notion of

---

27 See also 854 and 871.
causation/activity which he appeals to in *Principles* 26 needs to be broadened in such a way that
is consistent with some of his other remarks about activity. As noted above, in the *Notebooks*
Berkeley conceived of spirit not just as a thing that wills, but rather as “all that is active.” Also,
throughout the *Notebooks* and the first edition of the Principles, Berkeley seems to think of
perception, willing, imagining, remembering, thinking, feeling, etc. as acts, presumably all
included in spirit, or “all that is active.” Thus, willing is just one way that spirit acts/causes. It
might be helpful to conceive of spirit as an active being from which the acts of perceiving, willing,
thinking, etc., “emanate.” Spirit is a simple, active principle, and its various acts as we conceive of
them are its emanations or fulgurations.

Is activity conceived of in this way still properly termed causal? Yes, if we qualify Berkeley’s
notion of causation. Recall that earlier I defined Berkeley’s cause in this way:

**Berkeley’s Cause:** $S$ causes $x$ iff (1) $S$ is a subject that wills that $x$, and (2) $x$ occurs as a
direct result of $S$ willing that $x$.

This definition of causation, it was noted, builds in notions of efficient and final causation. But we
might distinguish a second type of cause, one that is not mutually exclusive with the above
definition, but is broader and less restrictive:

**Berkeley’s Active Cause:** $S$ causes $x$ iff $S$ is a subject that is necessary for the realization
of $x$.

An active cause is not necessarily a final cause, an efficient cause, or a total cause, though if certain
conditions are met, an active cause could be any one of these things. But in its most minimal sense,
an active cause is none of these. It is not necessarily a final cause (in Berkeley’s sense), because
the causes’ effects need not follow from the will of the being that is the active cause. Nor is it
necessarily an efficient cause, for it may be that some other thing is the primary or foremost
principle of motion or change of the effects. Nor is it necessarily a total cause, for it may exist and
yet not be sufficient for the realization of the effects. For $x$ to be an active cause of $y$, all that is
required is that the full realization of $y$ would not happen if not for $x$. All that is implied by claiming
$x$ is an active cause of $y$ is that $y$ would not be fully actualized if not for $x$. An active cause is just necessary for making something actual, rather than less than actual, or potential.

To get a better sense of how this theory of active causation might work, let us apply this notion of active causation to one of the various effects that Berkeley thinks spirit in some sense enacts: a perception. An actual perception as opposed to a potential perception will be a perception actually perceived by a mind. A potential perception will vary depending on one’s metaphysical intuitions about the status of the potentially perceivable world. For Berkeley (depending on idealist/phenomenalist interpretation) all potential perceptions are either ideas in the mind of God (idealist interpretation) or powers to make a human being have an idea (phenomenalist interpretation). But for a materialist or dualist, potential perceptions would just be unperceived matter. For a phenomenalist of a non-Berkeleyan bent, potential perceptions would just be the “permanent possibilities” for perception (Mill). Regardless of one’s metaphysics of substance, there is usually at least some sense in which that which is actually perceived is different from that which is potentially but not actually perceived. And whatever this potentiality of perception may be, the actuality of a perception is, on this theory, accounted for by the mind’s activity qua active causation. It is the activity of the mind that makes a potential perception into an actual one.

Perception is just the paradigm case used here, but we might apply the notion of an active cause to whatever else might be thought to exist in the bundle of experience. Depending on what one finds in their bundle of experience, this might include thoughts, ideas, volitions, passions, feelings, etc. All of these could plausibly have an active cause.

\[28\] Again, this will depend on what one takes to be evident in one’s phenomenology, even in the broader sense of introspection. Only the most extreme reductive empiricist, or perhaps the most hard-core Parmenidean rationalist, will deny that there is any first-person experience.
Based on this notion of an active cause, Berkeley’s argument can be reworked in a way that gives a more satisfactory definition of spirit:

**ARGUMENT FROM THE EXISTENCE OF EXPERIENCE:**

1. There exists an (actual) bundle of experience.
2. If there exists an (actual) bundle of experience, then there exists a distinct entity which is the active cause of it (i.e. which makes the potential experience into actual experience).

Therefore, there exists a distinct entity which is the active cause of this bundle of experience.

We can then call an entity that is the active cause of the bundle of experience **spirit**:

\[
\text{Spirit} = \text{df} \text{ an entity that produces a bundle of experience.}
\]

Therefore, if the above argument is indeed sound, we do in fact know that spirit exists, insofar as by “spirit” we mean the above definition. It is then a definitional matter to move to knowledge of the self:

\[
\text{Self} = \text{df} \text{ the spirit that is the active cause of the bundle of experience that is “specially present.”}
\]

Here, “specially present” is an attempt to define one’s phenomenology without begging the question of there existing a distinct experiencer of the phenomenology. It is meant to denote one’s present experience, regardless of whether it is experienced or produced by some distinct being. Such a premise even the most skeptical philosopher regarding the substantiality of the self can agree to; Hume, Wittgenstein, Lichtenberg, Mill, Mach, and (early) Quine would all seemingly
admit of its truth. Unless one is to claim that there is no extant phenomenology which is in some sense specially present, this part of the definition of the self ought not to be objectionable.29

From this it should be clear that if the argument is sound, we can be said to have knowledge not just of spirit in general but also of the self, for few if any will object to the definition’s key claim that there is some bundle of experience that is “specially present.” But is the argument in fact sound? Premise (1) is rather uncontroversial; in fact, if one admits that there is a bundle of experience that is “specially present,” it will follow that there exists at least one bundle of experience. The argument is also clearly valid, it just being an instance of modus ponens. But premise (2) is clearly contestable, particularly for those who seek to ground all knowledge in experience. This key premise is the subject of the following section.

Causal Maxims

In each of the three arguments for the self offered above, premise (2) is a variation of the same type of maxim: a causal maxim. A causal maxim simply claims that if some observable thing or event exists/happens, then there must be a cause of that thing/event. This maxim plays a key role in the arguments above; for the arguments to indeed work, this premise must be known. But Hume, who Berkeley mirrors with regards to the introspectability of the self, criticizes this maxim, claiming that it cannot be known certainly, neither by intuition nor demonstration. And the notion

29 I have, in this paragraph, continued to use language in a way that might imply there is a distinct being, separate from the “specially present” phenomenology. I do this not because I want to imply this, but because our language community does not have an expression for the concept I am trying to develop; in ordinary speech we always refer to “my perceptions” and “my phenomenology.” The point of this paragraph is to remove the “my” without losing the connotation that there is some phenomenology that exists which bears all the marks of an acquaintance relationship.

There are still some philosophers who may deny that there is a “specially present” phenomenology. Particularly, some hard-core reductionists may deny that they have any phenomenology. Also, a hard core rationalist, like Parmenides or maybe even Plato, might also deny this.
of causation Hume argues one gains from experience is very “thin”; a far departure from what the Berkeleyan arguments for the self require.\textsuperscript{30} I want to suggest two possible directions that Berkeley might take to avoid the Humean criticism of the causal maxim. The first is to claim (in rather rationalist fashion) that the causal maxim is in fact known \textit{a priori}, contrary to Hume’s various criticisms. The second is to claim (in a more empiricist fashion) that the causal maxim is known by experience. Both answers would not appeal to Hume, but for very different reasons.\textsuperscript{31}

Hume’s criticism of the \textit{a prioricity} and certainty of the causal maxim appears (among other places) in \textit{Treatise} 1.3.3:

\begin{quote}
‘Tis a general maxim in philosophy, that \textit{whatever begins to exist, must have a cause of existence}... [If] we examine this maxim by the idea of knowledge [explained in 1.3.1], we shall discover in it no mark of any such intuitive certainty; but on the contrary shall find, that ‘tis of a nature quite foreign to that species of conviction. (\textit{Treatise} 1.3.3.1)
\end{quote}

Hume’s criticism of this maxim is based on his definitions of knowledge and certainty; he claims “all certainty arises from the comparison of ideas, and from the discovery of such relations as are unalterable, so long as the ideas continue the same” (\textit{Treatise} 1.3.3.2). Not finding causation to be among this class of relations, Hume denies that there can in fact be any certainty about it. He concludes that “it is not from knowledge or any scientific reasoning, that we derive the opinion of

\textsuperscript{30} Again, I am likely being anachronistic here. Just as Hume’s bundle theory is more complicated than I have suggested, his notion of causation is more complicated that simple constant conjunction. But among the plethora of suggestions for how to make sense of Hume’s remarks on causation, no interpretations would sufficiently play the role that the Berkeleyan arguments for self-knowledge require.

\textsuperscript{31} It should be noted that the motivating question of this section (why should one accept the causal maxim in light of Humean critiques of causation) is obviously not a question Berkeley would have considered, since he published mostly before Hume’s criticisms ever appeared (1739), and certainly before they became popular or important. I ask this question in this section not because I expect Berkeley to have a well-formulated answer, but rather because the Humean critiques of causation are difficult and important problems for anyone who accepts a causal maxim to ground a metaphysical truth. If Berkeley’s argument for the existence of the self is to be viable for a contemporary audience, the theory ought to be able to respond to Hume’s criticisms given their apparent power.
the necessity of a cause to every new production, [therefore] that opinion must necessarily arise
from observation and experience” (Treatise 1.3.3.9).

How exactly Hume thinks the causal maxim arises from observation and experience, as well
as Hume’s appraisal of its efficacy in philosophy, are widely disputed topics. However, it is
generally agreed that the notion of causation developed by Hume involves two components:

Hume’s Cause: \( x \) is the cause of \( y \) iff

1) \( x \) is constantly conjoined with \( y \), and
2) \( y \) “necessarily” follows from \( x \).

The key point of contention among interpreters of Hume regards what Hume means by
“necessarily,” and whether or not Hume’s version of (2) is strong or “thick” enough to support the
use of the causal maxim in metaphysics and natural theology. Traditionally, Hume has been
interpreted as endorsing a very weak or “thin” notion of necessity, though this interpretation has
come under fire by some interpreters more recently. It is the Hume of lore that I consider here
rather than these more recent interpretations. This is not because I think the traditional
interpretation is correct, but rather because it portrays Hume in a way that poses a significant and
plausible challenge to the use of the causal maxim in metaphysics, especially for an empiricist.
Viewed this way, Hume’s critique of causation can be seen as a direct criticism of Berkeley’s
version of the causal maxim and its use in justifying knowledge of the self.

Hume seems to define necessity as a matter of custom or expectation of the mind, rather than
any metaphysical fact about the world:

This connexion, therefore, which we feel in the mind, this customary transition of the
imagination from one object to its usual attendant, is the sentiment or impression, from
which we form the idea of power or necessary connexion. Nothing farther is in the case.
(Enquiry 7.2.28)\(^{32}\)

References to Hume’s Enquiry are from: David Hume, An Enquiry Concerning Human
Understanding (London: 1773).

\(^{32}\)
This weak notion of necessity effectively undermines causal maxim’s usefulness in metaphysical speculation; instead making it a habit born of the mind or brain’s expectation. When Hume says $x$ is the cause of $y$, Hume is not claiming that $x$ possesses some actual power to bring about $y$, but is instead claiming that the mind, by habit or custom, expects $y$ to follow from the presence of $x$.

It is clear that this notion of causation is incapable of playing the role Berkeley requires in premise (2), precisely because Berkeley uses premise (2) to ground a *metaphysical* claim about the existence of spirit, and therefore the self. Since Hume’s notion of causation is only a fact about the way humans think rather than a fact about the way the world works, it would fail to establish that there is a metaphysical substance which is an active cause of the bundle of experience. At best, it could infer that there is a customary or habitual expectation of the existence of the self, a conclusion far too weak for Berkeley’s purposes. Therefore, it appears that if Berkeley’s argument for the self is to succeed, Berkeley must reject the Humean definition of causation in favor of a stronger, metaphysical definition.\footnote{As noted above, I am a more traditional, anachronistic interpretation of Hume. There is plenty of evidence that Hume did not accept the conclusions that I am attributing to him.}

As I suggested above, there are at least two distinct approaches that Berkeley might adopt to subvert the Humean rejection of a metaphysical definition of causation. The first approach is the *a priori* approach; on this theory, Berkeley would deny Hume’s claim that the “opinion [of the necessity of a cause] must necessarily arise from observation and experience” (\textit{Treatise} 1.3.3.9). Recall that Hume arrived at this conclusion because the causal maxim did not fit the criteria for intuitive or demonstrative certainty, based on his theory of knowledge and certainty developed in 1.3.1. Thus, if Berkeley is to claim that the causal maxim is known *a priori*, either by intuition or demonstration, he will need to in some way subvert the Humean arguments to the contrary.
While this doesn’t appear to be the route Berkeley takes, it should be noted that his empiricist predecessors, Hobbes and Locke, both attempt to demonstrate a causal maxim; Hume even singles them out (along with Samuel Clarke) in Treatise 1.3.3. While Berkeley does not do so, it may be a possibility for the empiricist to show that Hume’s criticism of the demonstrability of the causal maxim is incorrect. However, I think Berkeley is right in finding this solution to be unpromising.

Of course, a Kantian transcendental solution is also a possibility; on this solution, the causal maxim would be a “synthetic a priori” truth, rather than the more traditionally rationalist analyticity of a priori truths.

Translation from the latin:
Out of nothing, nothing comes. this (saith Spinoza op:posth:p 464) & ye like are called eternal truths because they have no credence outside of the mind.
endorsing a version of Spinoza’s maxim, there is no conclusive evidence from this passage to show he endorses Spinoza’s characterization of it as an eternal truth known by reason alone.\textsuperscript{37}

Further, if we examine the context in which \textit{Notebooks} 831 appears, we find that in the previous entry, Berkeley appeals to experience as evidence for the causal maxim:

\begin{quote}
G Why may we not conceive it possible for God to create things out of Nothing, certainly we our selves create in some wise whenever we imagine. (\textit{Notebooks} 830)
\end{quote}

Here, Berkeley appears to endorse that we do in fact experience our own causal power as we imagine. This reappears in \textit{Principles} 28, where he claims,

\begin{quote}
I find I can excite ideas in my mind at pleasure, and vary and shift the scene as oft as I think fit. It is no more than willing, and straightway this or that idea arises in my fancy; and by the same power it is obliterated, and makes way for another... Thus much is certain, and grounded on experience[.] (\textit{Principles} 28)
\end{quote}

This restatement of \textit{Notebooks} 830 in the \textit{Principles} makes very clear that Berkeley thinks that our notion of causal power is grounded on experience. \textit{Notebooks} 831, on the other hand, does not appear in a restated form in the \textit{Principles}.

Berkeley’s statements in \textit{Principles} 28 and its \textit{Notebooks} counterpart show that it is likely that Berkeley thought the causal maxim was not known \textit{a priori} but rather through experience, at least in his early period (1708-1710). For while Berkeley does often treat the causal maxim as if it is intuitive, nowhere does he explicitly claim the causal maxim can be known without experience. Berkeley is generally disdainful of \textit{a priori} knowledge, and may in fact have thought it impossible. Berkeley did not even think that mathematical truths are known \textit{a priori}. It simply seems more

\begin{footnote*}{37}{It is also worth noting that Berkeley denounces the \textit{veritates aeternae} earlier in \textit{Notebooks} 735: “Qu: wt becomes of the \textit{aeternae veritates}? Ans: they vanish.” We should treat this passage with caution, though, since it develops out of a theory of demonstration and certainty that Berkeley was developing at this time which it is not clear he accepted later. See \textit{Notebooks} 727-740.} \end{footnote*}
likely given his general empiricist leanings that Berkeley did not affirm the causal maxim on *a priori* grounds, but rather on experiential grounds, as he explicitly declares in *Principles* 28.\(^{38}\)

If Berkeley’s theory of the mental experience of causation is to ground knowledge of the causal maxim, then for Berkeley there are two stipulations that it must meet. First, the experience must be of the *causal power itself* (i.e. “causal transparency”), not just the constant conjunction of cause and effect. For if the experience is just of a constant conjunction of cause and effect and not of the causal power itself, then Berkeley cannot properly claim that there is direct experience of causation, and thus will not have any experiential evidence for the causal maxim that Hume did not consider and refute. Second, the experience of causation must not be *via* idea, since ideas are passive, causes active. Thus, the transparency of the causal power of the self must derive from some other introspective acquaintance relationship.

Whether or not our causal power is in fact transparent must be settled on phenomenological grounds. For my part, I must own that just as I cannot introspect myself, so too I cannot introspect the causal power that brings about ideas in my imagination. But I may be looking in the wrong place, as it is not clear from Berkeley’s work exactly what type of awareness this transparency comes from.

Nonetheless, there is still another problem with Berkeley’s attempt to ground the causal maxim in experience. One’s own direct experience of causal power in the exercise of her imagination will justify the following proposition:

\[^{38}\text{ Other more rationalist interpretations of Berkeley like McCracken’s, Ott and Hight’s, Bracken’s, and Loeb’s will likely disagree with me on this point. But most who suppose Berkeley to be hiding rationalist principles are often guilty of inferring that his somewhat rationalist-seeming metaphysics is rooted in a rationalist epistemology. As far as I can tell, Berkeley thought of himself as an unqualified empiricist, despite his occasional use of principles that appear rationalist. For a helpful defence of Berkeley as an empiricist, see Michael Ayers, “Was Berkeley an Empiricist or a Rationalist” in *Cambridge Companion to Berkeley*, ed. Kenneth Winkler (Cambridge: Cambridge University Press, 2005), pg. 34-62.}\]
A: The present imaginative idea has a cause.

Further, suppose we grant that in every instance of one’s exercise of her imagination, she directly experiences the causal power of her imaginative ideas. This would justify the following proposition:

B: Every past or present imaginative idea has a cause.

But it is not clear how one may infer from A or B to the following causal maxim:

C: Every imaginative idea must have a cause.

Nor is it clear how one may infer from A, B, or C to the following more general causal maxim:

D: Every idea must have a cause.

These are general inductive problems that call into question whether one is justified in generalizing to a universal claim from a particular sample. While Berkeley generally appears to be unperturbed by such inductive puzzles39, it is nevertheless a problem which he does not explicitly address.

So, there are at least two potential problems for Berkeley’s attempt to ground the causal maxim in experience. First, there is the phenomenological concern summarized by the following question: do we actually experience the causal power involved in imagining our own ideas? I am inclined to think the answer is no. Second, there is the inductive concern summarized by this question: can experience of the causal power involved in imagining ideas justify a more general causal principle? If the problem of induction is truly a problem, one must also answer no to this question.

I do not intend for these two objections to Berkeley’s theory of the grounding of the causal maxim to be taken as reasons for dismissing Berkeley’s theory altogether. I simply state them as

39 Cf. footnote 24.
challenges which require a response; whether there is in fact a viable response that has not been addressed in this paper can be taken up in future research.

Conclusion

To conclude: I have argued that Berkeley in his early philosophy (1708-1710) held that the self could not be directly apprehended through introspection, even under the broadest definition of introspection. Instead, during this period, Berkeley thought that the self must be known by means of inference, which involves an all-important causal maxim. And while all other components of the argument appear to be at least plausible or even certain, it is not clear, at least by what has been discussed, whether or not the causal maxim should be accepted. We’ve seen that Berkeley’s proposed grounding of the causal maxim in experience faces significant phenomenological and inductive problems, which potentially undermine his proposal.

It is worth noting, however, that if Berkeley’s causal maxim (or one similar) is not granted, then seemingly the only alternative is a bundle theory where at least in some sense the “specially present” bundle of experience is metaphysically brute. Many philosophers, particularly empiricists, may have no trouble accepting this conclusion; for them, there is no implausibility in supposing that there is no active being which contributes to the realization of the bundle of experience. But I suspect there are some empiricists who will find this answer unsatisfactory, perhaps equally or even more so than Berkeley’s proposed answer. These philosophers, myself included, may find that pursuing Berkeley’s answer is worthwhile, however fruitful or unfruitful the result may be.


CURRICULUM VITAE

RESEARCH INTERESTS

Primary:  
Early Modern Philosophy, George Berkeley

Secondary:  
History of Philosophy (general), Philosophy of Mind, Philosophy of Language, Philosophical Methodology

EDUCATION

2014 - B.A in Philosophy, Religion, minor in Greek from Calvin College

APPOINTMENTS

Current:  
2014 - Present: Teaching Assistant, University of Mississippi

AWARDS

Current:  
2014 - Present: Graduate Assistantship, University of Mississippi
2014 - Present: Graduate Fellowship, University of Mississippi

Past:  
2015: Graduate Summer Research Assistantship, University of Mississippi
2013 - 2014: Calvin Honors Scholarship, Calvin College
2010 - 2013: Calvin Knollerest Scholarship, Calvin College

PRESENTATIONS

March: “Berkeley on the Source of Self-Knowledge: Introspection and the Causal Maxim” at Mid-South Philosophy Conference


2014 - March: “The Metaphysics of Moral Luck” at Calvin College Student Philosophy Symposium

COMMENTARIES
2016 - Comments on Clinton Packman’s “Epistemic Consequentialism, Reliabilism, and the Problem of Epistemic Trade-Offs” at Mid-South Philosophy Conference

2015 - Response to Thomas Holden’s “Hobbes’s First Cause” at University of Mississippi Colloquium Series
Response to Michael Della Rocca’s “The Parmenidean Ascent: Substance” at University of Mississippi Colloquium Series

IN PROGRESS WORKS

Berkeley on the Source of Self-Knowledge: Introspection and the Causal Maxim
Berkeley’s Theory of Spirit and the Philosophical Notebooks.
Empirical Theism: Berkeley’s God

LANGUAGES

English: Fluent
Greek (attic): Fair
German: In Progress