

PhilonoUS

Issue 2, Vol 1 | Spring 2017 | Undergraduate Journal of Philosophy

Pitting Kant's Theory of Self-Consciousness Against Humean Associationism

Daniel Cheong

Cornell University

pp. 61-66



The
University
Of
Sheffield.

Visit our website at:
www.uosphilosophyjournal.wordpress.com

© 2017 The University of Sheffield
All works contained within are the property of their respective owners.

Pitting Kant's Theory of Self-Consciousness Against Humean Associationism

Daniel Cheong

Cornell University

In §16 of the B-Deduction of *The Critique of Pure Reason*, Immanuel Kant presents a theory of self-consciousness that asserts the existence of a subject independent of all perceptions. Kant reaches this conclusion via the use of a transcendental argument. Part I of this essay will briefly explain what Kant defines as a transcendental argument, and how it is used. It will then attempt to elucidate the construction of Kant's theory of self-consciousness iterated via the specific transcendental argument used in §16. Part II will consider David Hume's competing theory of associationism, as laid out in his *Treatise of Human Nature*. In particular, this part will examine Hume's position, inferring from the text several potential rebuttals to Kant's theory of self-consciousness. Finally, part III will evaluate the arguments presented by Kant and Hume in their respective texts, and find that Kant's argument for the perception-independent subject has insufficient rational force against the Humean position.

Section One: Kant's Transcendental Unity of Self-Consciousness

Before we delve into Kant's treatment of self-consciousness in §16, let us clarify the tool he uses in that section – the transcendental argument. In constructing a transcendental argument, Kant first begins with a relatively 'slender' premise about one's thought or experience, or, in other words, a premise that would be widely accepted by most who hear it. From this, Kant takes a step back to consider the condition that allows for this premise. To put this in more formal terms, Kant posits a condition *necessary* for the accepted premise. Broadly speaking, an example of a transcendental argument would be as such: if we accept X, and given that X can only be explained by the truth of Y, then Y must be true. In §16 Kant uses a kind of transcendental argument, which is part of the transcendental deduction of the categories, to demonstrate that the resultant premise applies to true experience. Since a detailed analysis of the transcendental method is not in the scope of this essay, this general framework set down will suffice to understand Kant's treatment of self-consciousness in §16.

In the first stage of §16 Kant begins with a premise about our thought that is compelling and would be widely accepted. He breaks this into two parts, the first of which entails that accompanying every representation¹ is the "I think". He writes that the "*I think* must be *capable* of accompanying all my representations" (B132). That is to say, if one were to observe that "the sky is blue", then one would also be able to

¹ Here, Kant means an observation via the senses. In this essay, "representation" and "perception" (Hume's preferred word) will be used interchangeably.

say, at least that “*I think* that the sky is blue”. Kant asserts that the “*I think*” is crucial, and can accompany *every* single representation. To justify this further, Kant offers the counterfactual to this proposition. If it were the case that the “*I think*” does not accompany a representation X, that that would be to say that X is not thought. For Kant, this in turn means one of two things: (i) that the representation is impossible, or (ii) that the representation “is nothing to me” (B132), i.e. does not exist to whoever the representation supposedly appears to. Kant further asserts that the representation of the “*I think*” is one that does not belong “to sensibility” (B132), or, in other words, is not derived from the senses. This is what Kant terms the “*pure apperception*”, or the “*original apperception*” (B132). The former term distinguishes the apperception as *a priori* as opposed to empirically obtained, whilst the latter term identifies this as independent of any further representation (B132).

The second part of this premise proceeds as such: since this pure or original apperception “*I think*” accompanies all representations and is the same in all of these representations (or as Kant terms, this “consciousness”), there is a *unity* of these apperceptions. Kant states here that the “*I think*” is “one and the same in all consciousness” (B132). Put literally, Kant asserts that the individual “*I*”s in these separate “*I think*”s refer to the same subject. Kant goes further to assert that such a unity is *necessary*. He writes that they “must conform necessarily to the condition under which alone they can stand together in one universal self-consciousness, since otherwise they would not thoroughly belong to me” (B133). Kant asserts that only through the similarity of what is conscious of the representations can there be self-attributions of these representations. Put simply, the ability to self-attribute representations *necessitates* the unity of apperceptions. This thus allows Kant to formulate a necessary unity of mental states that he terms the *transcendental unity of self-consciousness* (B132, B133).

Kant then moves on to the second stage of his argument in §16, where he examines what is necessary for the premise of the transcendental unity of self-consciousness. Kant argues that there is no way one comes to realize the unity of apperception without consciousness of the “synthesis of representations” (B133, B134). In other words, what Kant means is that there must be an active gathering of the “*I think*”s within each representation for one to be aware of the unity of consciousness. To clarify this assertion further, let us consider the empirical alternative that Kant offers. He argues that the “empirical consciousness that accompanies different representations is essentially scattered and without any reference to the subject’s identity” (B133). In other words, if the instances of apperception were merely empirical, i.e. associated to the representation *independent* of one another, then there would be no *necessary unity of self-consciousness* as accepted in the first stage of the argument. For Kant, such an alternative would prevent one from concluding that there is any *necessary unity of self-consciousness*, as representations would be attributed to discrete subjects as such and fail to refer to a single “subject’s identity”

(B133). This leads Kant to conclude that only an addition of "one representation to another and being conscious of this synthesis" (B133) allows the premise of the *transcendental unity of self-consciousness* to hold.

Section Two: Humean Response to Kant's Transcendental Unity of Self-Consciousness

Before proceeding to infer the appropriate Humean response to Kant's theory of self-consciousness, let us first briefly sketch out Hume's view of personal identity and self-consciousness. Broadly speaking, Hume rejects the idea that there is consciousness of the self, independent of perceptions. Hume accepts that many see the self as a node unto which "all impressions and ideas are supposed to be related" (132). Yet, he questions the method in which they belong or are connected with this idea of the self. Since Hume confronts particular perceptions upon looking inwards, and not any specific idea of the self, he asserts that one is simply a "bundle of perceptions that follow each other enormously quickly and are in a perpetual flux and movement" (133). In other words, what Hume means is that there exists only a grouping of perceptions with weak associations to each other, which in turn prevents any conclusion that asserts a strong idea of the self.

We term Hume's view of the self as "associationism", which conflicts with Kant's unity of self-consciousness. Hume would therefore probably not endorse this Kantian premise. Following along with Kant's construction of the transcendental argument, it seems at first plausible that Hume would accept the presence of an "I think", or apperception, that accompanies every representation or perception. We can infer this much as Hume writes that he is able to be *aware* of "some particular assumption of heat or cold, light or shade..." (132), which suggests he would agree that there is some degree of consciousness accompanying each perception. Such consciousness of one's perceptions is a tacit acceptance of the presence of the "I think" that accompanies each of these perceptions.

However, Hume would probably rebuff the subsequent moves Kant makes in three parts. First, Hume would not go so far as to say that there is a unity of apperceptions. This, for Hume, would not at all constitute a "slender" premise about our experience that is compelling. Instead, Hume would argue that the unity of the individual "I think"s that Kant asserts is the product of confusion between (i) "thinking about a sequence of related objects" and (ii) "thinking about one uninterrupted and invariable object" (134). Hume would argue that the bundle of perceptions are only a sequence of related objects, and are neither uninterrupted nor invariable. They may be related to each other by resemblance, contiguity, and cause and effect, but such relation is not *necessary*, as Kant so suggests. This supposed *necessary unity of apperception* that Kant offers is a product of substituting the notion of identity in (ii) with mere relation in (i). That is to say, Kant confounds the individual and independent "I think"s that may accompany each separate perception into a non-existent unity (134). In fact,

Hume would argue that Kant's "unity of apperception" is nothing but a "new and unintelligible thing" (134) that hardly constitutes a compelling premise as required by a transcendental argument.

Second, even if Hume assents to the presence or accompaniment of apperception to all perceptions, Hume may not necessarily agree with the conclusion that Kant reaches on the subject that is independent of perceptions. Hume also argues that he "never catch[es] [him]self *without* a perception, and never observe anything *but* the perception" (132). Therefore, even if Hume accepts the "I think" that accompanies each perception, it is something that is secondary, for what he observes primarily is the perception itself. Furthermore, this "I think" only ever exists *along with* the perception; there is no notion of independent existence in absence of perception, as Kant suggests.

Third, given the attack on the first "slender" premise that Kant offers, Hume would assert that the subsequent premise of "synthesis" so derived by the transcendental argument would be in turn invalid. Since there is no agreement that there is any necessary unity of apperceptions in the first place, then there is by definition no necessary condition that such "synthesis" even exists. Hence, Kant's conclusion of "synthesis of representations" would also not hold.

Section Three: Pitting Kant against Hume

From the standpoint of a neutral third party, Kant's premise of the unity of self-consciousness does not have sufficient rational force against the Humean position. As illustrated by Hume's inferred response, the central vulnerability of Kant's argument is in the first stage, where he leaps from the presence of apperceptions to its necessary unity.

First, Kant's argument for the unity of these apperceptions seems at best obscure, and at worst, circumlocutory in nature. Kant writes that there must be "one universal self-consciousness", for otherwise instances of perception "would not thoroughly belong to me" (B133). However, this presupposes that it is the same individual that is in receipt of these perceptions. To conclude from this that there is self-consciousness would be in effect saying:

- (1) There are perceptions. (premise)
- (2) Accompanying these perceptions are the "I think", or *apperception*. (premise)
- (3) I observe all these perceptions. (1, premise)
- (4) Since these perceptions to belong to me, all the accompanying *apperceptions* must be unified, or, in other words, there must be some necessary unity in my self-consciousness. (1, 2, 3, 4)

Saying that one observes all the perceptions in question is equivalent to saying that "I think" of all these perceptions. Hence, the conclusion in (4) that Kant reaches seems to be a mere restatement of the premise in (3), for which he does not offer any explicit justification, but merely an analytical truth. Assuming that we accept the existence of apperception in (2), claiming that we observe all perceptions in question by (3) necessitates the unity of individual accompanying apperceptions in (4) by definition. However, given Hume's associationist view, we are not certain that the perceptions we have are truly 'ours' in the first place. According to Hume, there is only a "bundle of perceptions" (133), and hence we may mistake these perceptions as being attributed to an uninterrupted subject instead of different "I"s which are merely *related*. This following example may clarify the difficulty faced. Say that in the first instance I observe a tree. In a separate instance, I observe a boat. One can say "*I think* I observe a tree" and "*I think* I observe a boat". However, this does not automatically lead to the conclusion that these separate instances of "I think", or apperception, are one in the same. Hume argues that even a "small or inconspicuous part" added to a mass makes it different than before this part was added (135). In a similar manner, there is no *guarantee* as such that the "I"s in each instance of apperception are indeed unified. Kant takes for granted that separate perceptions attribute to the same entity, when in fact they may be merely related in a non-necessary way, and are in fact different. Such an argument hence cannot be used to justify in a non-circumlocutory sense that there is a *necessary unity of self-consciousness*.

Second, and as a result of the previous point, Kant's argument of a unity of apperception seems to require an additional premise in order to defeat the Humean notion of associationism. In other words, there needs to be some evidence that individual instances of the "I think" accompanying perceptions are *truly* one and the same, that is, a *necessary unity of apperception*. For instance, Kant would need an additional argument that asserts the possibility of one being contemporaneously conscious of multiple instances of perception (Van Cleve 1999). This would affirm that the accompanying "I think" in these simultaneous perceptions refer to the same, unified self-consciousness. However, it seems questionable if such an argument made can even be tenable.

Section Four: Conclusion

In conclusion, it appears that Kant's theory of self-consciousness faces a central difficulty in illustrating a *necessary unity of apperceptions*. The argument he offers seems to be circumlocutory in nature, and probably needs an additional premise that might not be at all defensible. Overall, Kant's initial premise of the *unity of apperceptions* is subject to Hume's criticism of "unintelligibility" (133), which in effect undermines the second stage of his transcendental argument. It is neither evident

nor defended with a convincing premise, and hence the argument as a whole fails to possess sufficient rational force against the Humean position.

References

Kant, Immanuel. "Critique of Pure Reason." in *Modern Philosophy*. Eds. Roger Ariew and Eric Watkins. (Indianapolis: Hackett Publishing, 2009), 663-4.

Hume, David. "Treatise of Human Nature." in *Early Modern Texts*. Trans. Jonathan Bennett. (2004), 132-41. Retrieved April 18, 2015 from http://www.earlymoderntexts.com/pdfs/descartes1641_2.pdf

van Cleve, James. *Problems From Kant*. (Oxford: Oxford University Press, 1999).