INTROSPECTION AND SELF-KNOWLEDGE IN KANT

INTROSPECÇÃO E AUTO-CONHECIMENTO EM KANT

PEDRO STEPANENKO

Universidad Nacional Autónoma de México

pedros@unam.mx

Resumo: O objetivo deste artigo é oferecer uma reconstrução da objeção de Kant contra o Cartesianismo em relação ao auto-conhecimento usando uma teoria da introspecção HOT. Depois de expor brevemente a crítica de Kant do auto-conhecimento cartesiano como conhecimento de mim mesmo como um indivíduo, me concentro em interpretar a afirmação kantiana de que Descartes confunde a intuição do ser com a unidade da consciência quando ele supõe que não tem qualquer conhecimento de objetos externos. Argumento que essa unidade da consciência não pode ser a unidade de nossas experiências, porque, de acordo com Kant, esta unidade pressupõe conhecimento de nossas experiências, a qual também pressupõe conhecimento de seus objetos. Esta unidade pode somente equivaler a unidade de nossos pensamentos que acompanham nossa experiência.

Palavras-chave: Kant, Auto-conhecimento Cartesiano, Teoria HOT da consciência, Unidade da consciência, Paralogismos.

Abstract: The aim of this paper is to offer a reconstruction of Kant's objections against Cartesianism concerning self-knowledge using a higher-order thought theory of introspection. After expounding briefly Kant's critique of Cartesian self-knowledge as knowledge of myself as an individual, I concentrate on interpreting the Kantian statement that Descartes confounds the intuition of the self with the unity of consciousness when he supposes he hasn't any knowledge of external objects. I argue that this unity of consciousness cannot be the unity of our experiences, because, according to Kant, this unity presupposes knowledge of our experiences, which also presupposes knowledge of their objects. This unity can only amounts to the unity of our thoughts that accompany our experiences.

Keywords: Kant, Cartesian Self-knowledge, HOT Theory of Consciousness, Unity of Consciousness, Paralogisms

There is a tension in the *Critique of Pure Reason* between Kant's subscription of the dichotomy of inner and outer sense, on the one hand, and his objections against the Cartesian conception of self-knowledge, on the other. One of the passages that help to understand this

tension is at the end of the fourth Paralogism, where Kant is led to accept an "empirical dualism".

> …in the connection of experience matter, as substance in the [field of] appearance, is really given to outer sense, just as the thinking 'I', also as substance in the [field of] appearance, is given to inner sense. (A 379 ; trans. Kemp Smith)[1]

Compare this passage with the Refutation of Idealism, where Kant denies that inner sense can yield an abiding representation to which to apply the category of substance: "…we have nothing permanent on which, as intuition, we can base the concept of a substance, save only *matter*..." (B 277-278).[2]

The standard way to explain this contradiction appeals to the differences between the first and the second edition of the *Critique*. While in the first, both inner and outer sense allow us to know individuals, in the second, only outer sense provides this knowledge. The job of inner sense comes to be only the presentation of a sequence of mental states.

But, even if we accept this view of the second edition, the doctrine of inner sense remains in tension with Kant's critique of what we might call "the epistemic independence of self-knowledge". According to this critique, we cannot claim knowledge of a sequence of experiences without presupposing knowledge of at least some of their objects. The Kantian doctrine of inner sense, on the other hand, suggests that through introspection we acquire second order perceptions about our current experiences.[3] Now, since Kant conceives the relation of inner sense and its objects as immediate awareness,[4] those second order perceptions seem to constitute sufficient evidence to claim knowledge of the experiences they are about, independently of the fact that we take the representational content of those experiences as veridical.[5]

I think this later account of how we become aware of our own experiences is a Cartesian inheritance that precludes Kant's arguments against the epistemic independence of self-

[1] "…in dem Zusammenhange der Erfahrung ist wirklich Materie, als Substanz in der Erscheinung, dem äußeren Sinne, so wie das denkende Ich, gleichfalls als Substanz in der Erscheinung, vor dem inneren Sinne gegeben und nach den Regeln, welche diese Kategorie in den Zusammenhang unserer äußerer sowohl als innerer Wahrnehmungen zu einer Erfahrung hineinbringt, müssen auch beiderseits Erscheinungen unter sich verknüpft werden." (A 379)

[2] "…haben wir so gar nichts Beharrliches, was wir dem Begriffe einer Substanz, als Anschauung, unterlegen könnten, als bloß die Materie..." (B 277-278)

[3] B 68 ("inner perception of the manifold already given in the subject")

[4] A 19/B 33. See Paton (1970), 390.

[5] For that reason I think that the Kantian doctrine of inner sense fits better with what Gertler (2008) calls "Unmediated Observational Model of Self-knowledge" than with the inner sense model proposed by D. Amstrong and W. Lycan. See Allison (1983), p. 258, where he states that for inner intuition in Kant seems to hold Berkeley's principle that "the *esse* of a mental content is its *percipi*."

knowledge being successful.[6] In order to reconstruct these arguments I think we better have to look for the alternative account of introspection in Kant's philosophy, namely, apperception.[7] The principle of the transcendental unity of apperception, according to which "it must be possible for the 'I think' to accompany all my representations" (B 132) suggests by itself that introspection amounts to having thoughts about our current mental states.[8]

The aim of this paper is to explore a strategy to reconstruct Kant's critique against the epistemic independence of self-knowledge deploying a higher-order thought theory of introspection, instead of the higher-order perception theory of introspection that seems to support the doctrine of inner sense.[9] I will understand the thesis of the epistemic independence of self-knowledge as the claim that *I can acquire knowledge of myself even if I suppose there are no physical objects.* The reconstruction of Kant's arguments will be divided into two parts corresponding to two different senses of self-knowledge, namely as *knowledge of myself as a particular objec*t and as *knowledge of my experiences*. There is also a third sense of self-knowledge that will play an important role in this exploration: knowledge of our own thoughts (in the Kantian sense; not in the Cartesian sense). But it will not be argued against this sense.[10]

---

[6] In any case, the doctrine of inner sense creates such great difficulties that the Kantian epistemology doesn't seem able to solve. For some of these difficulties, see Collins (1999): 114-115.

[7] Even when Kant mostly uses "apperception" to refer to consciousness of the act of thinking, it can also refer to consciousness of our experiences as follows from the principle of the unity of apperception. What distinguishes inner sense from apperception in these case is that through inner sense we become aware of our experiences by means of sensibility (by being affected), whereas through apperception we become aware of them by means of thinking about them. For the difference between inner sense and apperception see: Allison (1983), pp. 275-278; Pippin (1987), pp. 463-466; Thomas (1997), pp. 285-289.

[8] Andrew Brook (personal communication) thinks it is wrong to talk of higher-order thoughts in Kant's philosophy of mind. But Kant clearly asserts in A 109 that our representations can be objects of other representations. And, since thoughts are representations, there isn't any reason to exclude that they can have perceptions or experiences as their objects. In Brook (2001) he contrasts Kant's apperceptive awareness as consciousness of the act of thinking, or more generally of the act of representing, with Rosenthal's higher-order thought theory of consciousness. The objection, as I understand it, is that we don't need to have higher-order thoughts about the act of representing in order to be conscious of the act of representing. That could be right, but it doesn't exclude the idea that in order to be conscious of our experiences we need to have thoughts about them. The Kantian idea that we need to use concepts in order to integrate perceptions to the unity of consciousness seems clearly to require that in order to be conscious of our own perceptions we need to have thoughts about them. Brook's statement that for Kant "A representation itself has the power to make us aware of it" (p. 19) is wrong, because Kant accepts the possibility of intuitive representations of which we weren't conscious (A. A. XI, 51-52).

[9] David Rosenthal (2002, pp. 247-250), Rocco Gennaro (2005, pp. 10-11) and Greg Janzen (2008, p. 87) have seen in Kant a forerunner of the HOT theory of consciousness. I don't know whether all cases of consciousness in Kant's philosophy could be described as cases of HOT. It isn't clearly that for Kant in order to be conscious of a thought we need a higher-order thought about the first-order thought. Nevertheless, I think that Kant would accept a HOT theory concerning consciousness of our own experiences.

[10] Actually I think Kant didn't have arguments against self-knowledge as knowledge of our own thoughts. But he never refers to knowledge of our own thoughts as self-knowledge. He refers to it as self-consciousness or apperception.

1. Kant's critique against epistemic independence of knowledge of myself as an individual

The main idea in the Refutation of Idealism, also present in the Critique of the First Paralogism, is that the epistemic value of the concept of substance depends upon its application to "something permanent in perception". (B 275) The claim that we can't apply the concept of substance if we don't perceive something permanent can be formulated more precisely as it follows: we can't apply the concept of substance if we don't conceive the objects of our perceptions as something recognizable in different perceptions. So, we can know only those objects that can be objects of different possible experiences.

Now, only physical objects satisfy the prior requirement. This follows from the statement of the Refutation of Idealism I quoted above: "…we have nothing permanent on which, as intuition, we can base the concept of a substance, save only *matter*..." (B 277-278).

With these two points in mind we can formulate the following argument:

1. Premise 1: Only those individuals that are objects of different possible experiences can be known.

2. Premise 2: Only physical objects can be objects of different possible experiences.

3. If I am not a physical object, I can't acquire knowledge of the individual I am.

4. If I suppose there aren't physical objects, I also suppose that I can't acquire knowledge of myself as an individual.

Now, what is interesting in Kant's critique against the epistemic independence of self-knowledge is his explanation of how we fall prey of this "natural illusion", of the illusion that we can acquire knowledge of ourselves, even when we suppose there are no physical objects. The famous passage that synthesizes how this happens is the following: "The unity of consciousness, which underlies the categories, is here mistaken for an intuition of the subject as object, and the category of substance is then applied to it."[11](B 422) If we suppose there are no physical objects we are still aware of a manifold of mental states. These mental states are unified in such a way that it makes us think that we are recognizing the same particular object (myself) in different mental states. The unified manifold makes us think there is a particular object to be known.

[11] trans. N. Kemp Smith

There could be another way to understand Kant's diagnosis, namely, that in such a scenario I take the "simple, and in itself completely empty of content, representation 'I'"[12] as "an intuition of the subject as object". But I will not follow this interpretation for two reasons: 1) This interpretation helps Kant to attack the possibility of acquiring knowledge of ourselves without taking into account experience, as Rational Psychology is supposed to do, whereas the defender of the epistemic independence of self-knowledge, as I formulated this position, doesn't need to claim a priori knowledge of ourselves. 2) I don't understand what could it be a unity "completely empty of content". For, in order to talk about a unity there should be at least a manifold of elements and a set of relations among them.

So, if the illusion of epistemic independence of self-knowledge is due to the confusion between an individual and a manifold of unified conscious mental states, it should be said that when we suppose there are no physical objects we are led to take knowledge of this manifold as knowledge of an individual.

There are different ways of characterizing the unity of consciousness in the *Critique of Pure Reason*. The main contrast, obviously, is between empirical and transcendental unity of consciousness. The latter should be understood as the structure of the former, namely, as the set of conditions that all actual (empirical) unity must satisfy. The transcendental unity can also be described as the unity of all possible experiences. The explanation of the illusion behind Rational Psychology can be formulated as the confusion between the transcendental unity and an immaterial substance. But this explanation makes sense once we have accepted the Kantian account of the mind. It cannot be used to argue against the defender of the epistemic independence of self-knowledge without begging the point. We have to focus on the unity of the conscious mental states we actually have and argue that there is a necessary structure of possible mental states. We have to point out what kind of relations among our actual mental states can make us think that there is a particular object that unifies them.

I think that there are two main characterizations of the unity of consciousness which can play this role: 1) as a manifold of experiences in a temporal sequence and 2) as a manifold of thoughts related according to inferential rules. Peter Strawson's reconstruction of the Transcendental Deduction endorses the first, while Wilfrid Sellars' conception of the categories suggests the second.[13]

[12] "die einfache und für sich selbst an Inhalt gänzlich leere Vorstellung: Ich" (A 346-347/B 404)

[13] Strawson's famous Objectivity Thesis ("For a series of diverse experiences to belong to a single consciousness it is necessary that they should be so connected as to constitute a temporally extended experience of a unified objective world") takes unity of consciousness as a unified manifold of experiences (Strawson (1966), p. 97). If

It can be objected that none of these two characterizations can be a good candidate for the unity of consciousness, because each one concerns only some kind of mental states, but not all. Nevertheless, it can be replied that each of these characterizations concerns some feature shared by both kinds of mental states. Thoughts are mental states that happen in temporal sequences. Experiences have propositional contents that are inferencially related to propositional contents of thoughts. But the individuation of thoughts is different from the individuation of experiences. We can individuate a thought merely by taking into account the inferential relations of its propositional content, whereas we must take into account the phenomenal character as well as time relations in order to individuate an experience.

According to the first characterization of the unity of consciousness what unifies a manifold of experiences and thoughts is the fact that they belong to a single temporal sequence. Inner sense is what allows us to be aware of this flow of mental states. Now, since time is the form of inner sense and inner sense is what makes us know whether our thoughts about experiences are true, we acquire knowledge of an experience at the moment we are having it. The attribution of knowledge is then incorrigible, since we can't run time backwards. Memory can only help to preserve the knowledge we have acquired.

According to the second characterization, what unifies a manifold of experiences and thoughts is the fact that they have propositional content related to the propositional content of other possible experiences and thoughts. The understanding is the faculty that allows us to be aware of these relations. Kant denies that the understanding alone can give us knowledge. But there is a sense in which we can say we know something when we merely entertain a thought: we know what we are thinking. We can individuate our thoughts taking into account their inferential relations. I think some passages of the Paralogism can be understood as asserting that the unity of consciousness in this sense is what we take mistakenly as knowledge of the particular we are.[14]

---

we take into account the characterization of the unity of apperception in § 19 of the B-Deduction as a unity that makes possible objective judgments and Sellars' interpretation of the theory of the categories as a classification of the inferential connections of empirical judgments (Sellars (1974), pp 53 and 337), it is possible to conceive the Kantian unity of consciousness as a unity of thoughts related according to inferential rules.

[14] See: A 402: "Gleichwohl ist nichts natürlicher und verführerischer, als der Schein, die Einheit in der Synthesis der Gedanken für eine wahrgenommene Einheit im Subjekte dieser Gedanken zu halten. Man könnte ihn die Subreption des hypostasierten Bewußtseins (apperceptiones substantiatae) nennen." ["Nevertheless there is nothing more natural and more misleading than the illusion which leads us to regard the unity in the synthesis of thoughts as a perceived unity in the subject of these thoughts. We might call it the subreption of the hypostatised consciousness (*apperceptionis substantiatae*)." Trans. Kemp-Smith, p. 365]

If we want to give a full description of the unity of consciousness, we should certainly take into account both characterizations and say: "unity of consciousness is a manifold of experiences and thoughts of which we are aware by means of inner sense and whose propositional content are inferentially connected." But this full description can't help us to understand which connections among experiences and thoughts are responsible of the confusion between knowledge of an individual and knowledge of a unified manifold. In order to answer this question, we have to answer which kind of knowledge could remain when we suppose there are no physical objects. According to the foregoing distinction we have two possible answers: 1) what remains is knowledge of a manifold of experiences in a temporal sequence; or 2) what remains is knowledge of a manifold of thoughts inferentially related. One of these two kinds of knowledge is what leads us to think that we know the individual we are, even when we suppose there are no physical objects.

2. Kant's critique against epistemic independence of knowledge of myself as knowledge of my experiences.

Before trying to connect the previous outcome and Kant's attempts to argue against epistemic independence of knowledge of myself as knowledge of my experiences, let me recall briefly the main features of the latter. I think everybody would agree that such attempts are to be found in the Transcendental Deduction and in the Refutation of Idealism. In the first, Kant argues that the unity of a manifold of experiences is made possible by concepts that allow us to think about objects, so that unified experiences are experiences of objects. In Strawson's words: "for a series of diverse experiences to belong to a single consciousness it is necessary that they should be so connected as to constitute a temporally extended experience of a unified objective world" (Strawson 1966, 97). In the Refutation of Idealism, on the other hand, Kant argues that knowledge of a manifold of experiences in time requires knowledge of something that persists in time and that only knowledge of physical objects counts as such.

I will not assess these arguments. I just want to point out that if they were successful, they would exclude the first explanation of the confusion between knowledge of ourselves as individuals and knowledge of a unified manifold of conscious mental states. If Kant succeeds in showing that we can't have knowledge of our experiences without taking into account knowledge of their objects, then, when we suppose there aren't physical objects, we can't say we know our experiences. What could remain is only knowledge of our thoughts. So, if we

want to make compatible Kant's critique against the first sense of epistemic independence of self-knowledge and the second, we should say that for Kant what remains after supposing there are no physical objects is knowledge of a manifold of thoughts inferentially related. But then we have to explain what kind of cognitive relation we have with our own experiences when we suppose there are no physical objects. For we can't deny we are having experiences even if we suppose there are no physical objects. And we have to explain why this cognitive relation doesn't amount to knowledge.

Let me suggest two Kantian conditions of knowledge of an experience: 1) we have knowledge of an experience when we know its propositional content, and 2) we have knowledge of an experience when we know some inferential relations of its propositional content with the propositional content of other (past or possible) experiences. So, if it makes sense to say that we can have conscious experiences without being committed to say we know them, we have to accept some kind of presence of the experiences that don't satisfy the condition of knowing its propositional content. Some would find this as a hopeless situation. I personally think there is some hope if we substitute the Kantian conception of introspection as inner sense for a conception of introspection as higher-order thought. For, if what makes me conscious of an experience is the fact that I have a thought about it when it is present, then it is possible that this thought attributes to the experience a propositional content that the experience doesn't really have and, hence, the right way to describe the situation would be the following: I am conscious of this experience even when I am not allowed to say that I know its propositional content. The thought that accompanies the experience should fix it in some way and makes me possible to check later its propositional content taking into account the propositional content of other experiences. I think sometimes something like this happens: for instance, when we realize that what we really saw wasn't what we thought we were seeing. In that case we don't doubt that we thought such and such; we simply realize that this thought was false. Does that entails that the propositional content of the accompanied experience was also false? I don't think so. It could be perfectly possible that when we realize our thought is false, we also remember the experience and realize its propositional content wasn't the one we thought. The fact that the thought about a current experience is false doesn't entail that the propositional content of the experience is also false.

So, introspection as higher-order thought about our experiences allows Kant to sustain that the Cartesian confounds knowledge of an immaterial substance with knowledge of a manifold of thoughts without denying we are conscious of our experiences when we suppose

there are no physical objects. It makes sense after all to say we are conscious of experiences, even when we are not allowed to say we know them. But this conception of introspection can also help to explain the indubitability the Cartesian attributes to our experiences, that is, the impossibility to doubt that we are having the experiences we think we are having. If this indubitability is understood as the impossibility to doubt that I am having an experience with certain propositional content, the higher-order thought theory of introspection can explain this illusion as the confusion between thoughts that accompany our experiences and those experiences. For, concerning thoughts in contrast to experiences, it makes no sense to say we can be aware of them without knowing its propositional content. Thoughts don't have phenomenal content; all their content is propositional. What are indubitable are the thoughts about a current experience. But, of course, we can doubt that the experience we are actually having is the experience with the propositional content we think it has.

If we analyse the inferential consequences of a thought, we are specifying its propositional content. Since thinking is a spontaneous faculty, in reflexion we produce thoughts that are connected with the thought we are specifying. In contrast, analysing an experience can certainly help us to specify its propositional content, but it cannot exclude the possibility of being wrong. Only connections among the propositional content of different experiences could reduce the probability to be wrong. Now, since experiences depend on our receptivity, we cannot confirm (or refute) what we think about the propositional content of an experience merely by analysing its consequences. We have to take into account other experiences that have been or can be given to us. That means that other experiences must be able to give us information about the propositional content of the current experience. That is a consequence of the second condition I proposed above for knowledge of an experience. If I need knowing some inferential relations of the propositional content of the experience I am having in order to know that experience, I must then be able to recognize other experiences that confirm that I am having or that I had that experience. But that would be impossible if I suppose the propositional contents of all my experiences are false. How could it be possible that the propositional content of an experience could give me information of the propositional content of another experience if I suppose it is false?

I think that the foregoing outcome can help us to argue successfully against the epistemic independence of self-knowledge in the second sense. For the defender of the latter position, as I characterized it at the beginning of this paper, supposes that it is possible to acquire knowledge of our experiences about external objects even if all their propositional

contents were false. But, if we need to take into account the propositional content of other experiences in order to acquire knowledge of our current experience, then the possibility this defender is advocating is actually excluded.

REFERENCES

ALLISON, H. (1983), **Kant´s Transcendental Idealism. An Interpretation and Defense.** New Haven: Yale University Press.

ALSTON, W. (1971). **Varieties of Privileged Access.** *American Philosophical Quarterly*, vol. 8, No. 3.

BROOK, A. (2001). Kant, self-awareness and self-reference in: Brook, A. and DeVidi, R.C. (eds.), **Self-reference and Self-awareness.** Amsterdam/Philadelphia: John Benjamins, pp. 9-30.

BURGE, T. (1988). Individualism and Self-Knowledge. **The Journal of Philosophy** Vo. 85, No. 11, pp. 649-663.

BYRNE, A. (1997). Some Like it HOT: Consciousness and Higher-Order Thoughts". **Philosophical Studies,** 86, pp. 103–129.

COLLINS, A. (1999). **Possible Experiences**. Berkeley and Los Angeles: University of California Press.

GENNARO, R. (2005). The HOT Theory of Consciousness. **Journal of Consciousness Studies**, 12, No. 2, pp. 3-21.

GERTLER, B. (2008). Self-Knowledge. in: **Stanford Encyclopedia of Philosophy**, online.

JANZEN, G. (2008). **The Reflexive Nature of Consciousness.** Amsterdam/Philadelphia: John Benjamins.

KANT, I. (1982). **Kritik der reinen Vernunft**, in Werkausgabe, vol. III and IV, ed. by W.

Weischedel. Frankfurt: Suhrkamp.

_____. (1929). **Critique of Pure Reason.** trans. Norman Kemp-Smith. London: Macmillan.

PATON, H. J. (1970). **Kant's Metaphysic of Experience.** vol. II. New York: Humanities Press Inc.

PIPPIN, R. B. (1987). Kant on the Spontaneity of Mind. **Canadian Journal of Philosophy**, 17, 2, pp. 449-476.

ROSENTHAL, David (2002). **Consciousness and Mind**. Oxford: Clarendon Press.

_____. Consciousness and the Mind Iyyun. **The Jerusalem Philosophical Quarterly** 51, pp. 227-251.

SELLARS, W. (1967), **Some Remarks on Kant´s Theory of Experience.** in: SELLARS (1974), pp. 44-61.

_____. (1970), **Towards a Theory of the Categories.** in: SELLARS (1974), pp. 318-339.

_____. (1974), **Essays in Philosophy and its History.** Dordrecht: D. Reidel.

SCHWYZER, H. (1997). Subjectivity in Descartes and Kant. **The Philosophical Quarterly**, vo. 47, no. 188, pp. 342-357.

STRAWSON, P. F. (1966). **The Bounds of Sense: An Essay on Kant's Critique of Pure Reason.** London: Metheun.

THOMAS, A. (1997). Kant, McDowell and the Theory of Consciousness in: **European Journal of Philosophy** 5 (3): 283-289.