**DIPLOMAMUNKA**

**MA THESIS**

Radivojević Mihajlo

Logic and Theory

of Science MA

2015

EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Bölcsészettudományi Kar

**DIPLOMAMUNKA**

**MA THESIS**

*Az öntudat egysége Kantnál*

*Unity of the Self in Kant*

**Témavezető: Készítette:**

 Szalai Judit Radivojević Mihajlo

 Egyetemi adjunktus Logic and Theory of Science MA

2015

A HKR 346. § ad 76. § (4) c) pontja értelmében:

„… A szakdolgozathoz csatolni kell egy nyilatkozatot arról, hogy a munka a hallgató saját szellemi terméke…”

**Szerzőségi Nyilatkozat**

Alulírott Radivojević Mihajlo …………………(Neptun-kód) ezennel kijelentem és aláírásommal megerősítem, hogy az ELTE BTK Logic and Theory of Science angol nyelvű mesterszakján írt jelen diplomamunkám saját szellemi termékem, melyet korábban más szakon még nem nyújtottam be szakdolgozatként és amelybe mások munkáját (könyv, tanulmány, kézirat, internetes forrás, személyes közlés stb.) idézőjel és pontos hivatkozások nélkül nem építettem be.

Budapest, 20

a hallgató aláírása

Contents

[Introduction 4](#_Toc416722005)

[1. Kant's Account of the Unity 8](#_Toc416722006)

[1.1. Background 8](#_Toc416722007)

[1.2. Transcendental Apperception 11](#_Toc416722008)

[1.3. The Second Project 20](#_Toc416722009)

[1.3.1. The Substance Interpretation 25](#_Toc416722010)

[1.3.2. The Personal Identity Interpretation 28](#_Toc416722011)

[1.4. Potential Empirical Aspects of Transcendental Apperception 36](#_Toc416722012)

[2. Transcendental Apperception and Dissociative Identity Disorder 45](#_Toc416722013)

[2.1. Features of DID 45](#_Toc416722014)

[2.2. DID and Transcendental Apperception 50](#_Toc416722015)

[3. Self as a Centre of Narrative Gravity 55](#_Toc416722016)

[3.1. Dennett's Theory 55](#_Toc416722017)

[3.2. Dennett and Kant 59](#_Toc416722018)

[References 64](#_Toc416722019)

# Introduction

The issue of the unity of the self is recognized as an important issue throughout a large part of the history of Western philosophy. Yet, as every significant philosophical problem that survived through several radically different philosophical eras it was subject to substantial changes, not just in the way it was analyzed, but also regarding the very scope of the investigated problem, the motivation behind it, as well as regarding the language used in this investigation. The problem of the unity of self can be recognized as early as in Greek philosophy, most prominently in Plato's discussions about the nature of the soul. Subsequently, for a number of medieval philosophers who confronted the problem, the point of these investigations was to prove the immortality of soul by proving its simplicity and unity. For philosophers belonging to schools of empiricism and rationalism the issue was somewhat similar and it was investigated from different angles and with different results, but the biggest gain of demonstrating this unity was still that it entailed immortality of the soul in this way or another. There are some suggestions among scholars that Kant had the same thing in mind[[1]](#footnote-2), and it is certainly one of the problems that will be dealt with in this paper.

In contemporary philosophy, we recognize that a very similar set of questions to those that are asked when studying the unity of self (or soul) is asked when we consider the problem of personal identity. Admittedly, the question of immortality is not the focus of contemporary discussions any more, and philosophers avoid using the word "soul" in these considerations. Furthermore, a huge development of the sciences of the brain changes the focus of the issue and brings new nuances to it. However, the problem of personal identity and the problem of the unity of self certainly overlap and it can definitely be said that they have the same main aim and - to put it maybe a bit too simply - this aim is to figure out whether there is *something* real that unites all our experiences and that gives rise to the idea of ourselves as being identical at all times, or is the self and its identity just an illusion. Moreover, if it is an illusion, we should be able to explain where it comes from and why it is so widespread and pervasive.

I believe Kant's answer to these questions, when properly revitalized, can offer a strong background for some new solutions for these issues. The problem, however, to put it in Brooks words, is that *"Kant's epistemology is one of the pillars of his fame - his ethics is another - and his psychology was a mere byproduct of his epistemology, a body of ideas that even he viewed as incomplete and inessential to his main project".[[2]](#footnote-3)* That does not only give us troubles when we try to properly interpret everything Kant claims about the soul, self or personhood, but also obliges us to, when presenting his theory, start from the very beginning and from the very foundations of his philosophy. Explaining the main ideas of Kant's philosophy and showing how his theories of the self and the soul fit into Kant's general project, as well as the extensive analysis of these theories, will thus be the first and the most important aim of this paper. This will be the topic of the first chapter.

In the first section of this chapter I will attempt to situate where the problem of the self belongs in Kant's philosophy and how the general idea of Kant's transcendental project relates to the traditional ideas, most prominently those of empiricism. Then, in the second and third section of this chapter I will try to analyze what precisely Kant's theory of the unity of self consists in, and point out some of the problems with the theory. In dealing with these issues, I will make use of Brook's idea of "two main projects" of Kant's psychology.[[3]](#footnote-4)

Namely, according to Brook, the first main project is proving that the results of natural science are to be held as necessary truths. On the other hand, the goal of the second project is to isolate religion and morality from science, thus diminishing the authority of physics and limiting it to the world of sensory experience, while at the same time introducing morality as an exclusive subject of practical philosophy. The first project is presented in the introductory parts of the *Transcendental Analytic* (with special emphasis on the *Transcendental Deduction*), while the second project is laid out in the chapter about the paralogisms of pure reason. Roughly speaking, the topic of the second section of the first chapter of this paper will be the "first project", namely the introduction of transcendental apperception as a solution to some of the problems empiricist philosophers, most famously Hume, have put before us. The aim of the third section of the first chapter will be to deal with the "second project" - issues presented in the Paralogisms - proving that unity of self can (and must) exist even without postulating the self as an entity, either material or immaterial. This is the point in the Critique of Pure Reason where Kant grapples with rationalist illusions about the *res cogitans,* and also explains where this urge to conflate the "I" of the transcendental apperception and the I as a substance comes from.

The fourth and final section of the first chapter will deal with the various attempts to interpret Kant's transcendental self in a physical way - either by saying that the awareness of this "I" is nothing else than the awareness of one's own body, or by trying to give a neurological explanation of the issue.

The second and third chapter will deal with an important issue of rethinking Kant's theory in a modern context and with one possible application of the Kantian way of thinking in solving current philosophical issues. The second chapter will be devoted to the inquiry on how Kant's theory of the self corresponds to a psychological disorder very much related to this issue - the dissociative identity disorder (DID, formerly known as multiple personality disorder, MPD). When faced with the reality of the phenomenon of multiple personality, there is an apparent contradiction in Kant's demand that one and the *same* "I" always follows all my representations. It will be shown that these two can be in fact reconciled and that Kant's theory is consistent with the existence of multiple personalities in one body.

In the third and final chapter of the paper I will show that Daniel Dennett's approach to the problem of the unity of self - the theory of the self as a center of narrative gravity - is Kantian in several very important ways. At the same time, the vast differences between the two theories will be pointed out, since the theories have very different motivations and ontological assumptions, so drawing analogies between them has to be done cautiously. The point of the chapter will be to show that a modified and readjusted Kantian theory of the self, when stripped off of some of its ontological assumptions and anachronous ideas, can be made good use of when philosophizing about problems we still encounter in contemporary philosophy of mind and, more specifically, theories of the self. This will especially be clear when comparing Dennett's account of the problem of DID to the way Kant's theory could deal with it.

Revitalizing Kant's theory of the unity of self might not be important only for the sake of philosophical discussion, but also when solving some of the practical and legal issues concerning the concept of the self and personhood. To quote Grant Gillet, "*The unity of conscious states attributable to a given human individual is not merely a foundation for theoretical inquiry but is of immense practical importance. In everyday life we base our 'reactive attitudes' on the idea that folk are conscious of their own thoughts and can be seen as acting on the basis of a coherent set of purposes and beliefs.*"[[4]](#footnote-5) Therefore, putting these problems into focus can be of importance for everyday life, too.

Kant also emphasizes, in a widely quoted and differently interpreted paragraph, that resolving the issue of the unity of self has a practical value as well. "*Meanwhile, the concept of personality, just like the concept of substance and the simple, can remain (insofar as it is merely transcendental, i.e., a unity of the subject which is otherwise unknown to us, but in whose determinations there is a thoroughgoing connection of apperception), and to this extent this concept is also necessary and sufficient for practical use*".[[5]](#footnote-6) The Kantian, transcendental solution, although taking away the privilege of being a substance from the self, is still important and good enough for Kant to draw the conclusions he draws in his practical philosophy. This will also be a matter of discussion later in the paper. The issue of the unity, how it works and how we can prove it, however, largely belongs to Kant's theoretical philosophy and is most extensively presented in the *Critique of Pure Reason* in the aforementioned chapters (basically, *Transcendental Deduction* and *Paralogisms*). Therefore I will focus on these in the main chapters of the paper, however turning to issues of Kant's practical philosophy whenever that seems to be necessary.

# 1.Kant's Account of the Unity

## 1.1. Background

The basis of Kant's critical project is given as early as in the Introduction of the *Critique of Pure Reason*. The best way to explain briefly what the key points of critical philosophy are is to present it in the context of its relation to empiricism. Kant's theory of knowledge has one thing in common with empiricism - namely, Kant agrees with the empiricistic conclusion that we cannot find any objective connection in the mere perceptions, if we assume that our experience is made merely of the perceptions of outer world. However, as Hume shows this means that all the "regularities" we observe in the world, most importantly causality and, with it, physical laws, are just illusions based on repeated contiguity and succession of impressions and events and have no objective validity. "*For after a frequent repetition, I find, that upon the appearance of one of the objects, the mind is determin'd by custom to consider its usual attendant, and to consider it in a stronger light upon account of its relation to the first object. 'Tis this impression then, or determination, which afford me the idea of necessity."[[6]](#footnote-7)* All the law-like relations in the world are the results of mere habits. Kant agrees that this is the conclusion that follows from the premises, but the undesirable consequence of this position is that it basically denies any legitimacy to the achievements of natural science. Kant tries to find a way around this consequence and thus one of the key tasks of the *Critique* (what Brook calls "the first project") is to "save" physics, to somehow prove the objective validity of physical laws and, also, mathematical truths (since Kant considers mathematical "truths" to be irrefutable facts).

In order to do this, Kant had to made several important distinctions. First of them is the distinction between a phenomenon and a thing-in itself, or a phenomenon and a noumenon. The point where Kant agrees with empiricists is that there is no experience nor knowledge without sensory perceptions, without sensations that affect our senses. However, all we can know about things is how they affect our senses, not what they are or how they are *in themselves*, in an absolute sense. Empiricism takes for granted that these two are the same and that we are just passive receivers of outside sensations, without taking an active role in the process of our own experience. From this Hume rightly concludes to the impossibility of objective connections between things in the world, but Kant thinks that the objects in the world, as we experience them, as phenomena, belong only to our consciousness, while the world *as it is* and things *as they are* not just might be different from that, but we cannot have any theoretical knowledge about them whatsoever. It will, however, turn out that although noumena cannot constitute our theoretical knowledge, they can serve as a regulatory rule for it and that they can also be used extensively in the development of practical philosophy. In any event, Kant reckons that we never experience the world as it is, but instead that a key contribution to experiencing the world is made by our active role in the process of cognition.

The three faculties of the mind - senses, understanding and reason - work together to make this contribution. It turns out that regularities in nature exist not because we find them there, but because we prescribe them to nature. It is in the act of experience that these regularities and laws happen and that is why we cannot say anything about them potentially existing or not existing outside the human experience. All this is still not too different from Hume - after all, Hume also considers the human mind responsible for "supplying" nature with such concepts as physical laws. The main difference is that Hume sees these concepts as motivated by mere repetition of similar associations of perceptions, while Kant sees them as prescribed a priori.

The key to understanding this is making another important distinction - a priori vs. a posteriori, or pure vs. empirical. Our inclination to connect our perceptions into meaningful law-like networks is not just a matter of habit but is actually a necessary feature of the workings of our mental faculties. The rules that our minds use and apply to the mere sensations affecting us, in order to make sense of them, are necessary and universal - we *all* apply the same rules and we *always* apply the same rules, and that is why we are able to communicate meaningful ideas at any level. One of these rules is the concept of causation - we never find causation in the objective world, as Hume correctly remarks, and this is exactly why it is necessary to think of it as an a priori concept of the human understanding that is ascribed to the world by us. "*Thus we ourselves bring into the appearances that order and regularity in them that we call nature, and moreover we would not be able to find it there if we, or the nature of our mind, had not originally put it there."* [[7]](#footnote-8) Causation, among other concepts, is the basis of the physical laws we all agree on, because without having causation as a proper concept, it would be absurd to talk about physical laws. Explaining causation away by stating that it is nothing more than a mere habit was the too-easy way out for Kant.

This is the essence of the appeal that synthetic a priori judgments have for Kant - they are stating something about the world that are not mere analytical truths, they contribute to our knowledge, yet they are universal, necessary and not dependent on our experience, because, as Hume already pointed out, and Kant repeats it numerous times in the *Critique*, these judgments cannot be derived from mere sensory data we passively receive. From the fact that they are "omnipresent", universal and necessary, yet not derivable from experience, Kant draws the conclusion that the only reasonable explanation is that they are the very basis of our experience. At this point, the most important thing for Kant to show is how physics and mathematics are *possible* and that there is nothing paradoxical about them. Synthetic a priori judgments prove to be essential for this task because they serve as a basis for all synthetic a posteriori judgments, for all the statements that claim something about the objective world and thus broaden our knowledge. To put it in Cassirer's words: "*For it is not the case that the world of cognitions and truths as an imprint and copy of the objects exists because the world of objects exists, but it exists because certain judgments unconditionally exist - judgments whose validity does not depend on the individual empirical subject that asserts them, nor on the particular empirical and temporal conditions under which they are generated - because for us there is an order that should not be designated only as the order of impressions and representations but as the order of objects*."[[8]](#footnote-9)

The inner workings of pure intuitions of space and time and pure concepts of understanding are therefore the first concern of critical philosophy, because it is through them that any kind of objective knowledge is established. The objective of the first project is to make a sensible and plausible sketch of these inner workings. The objective of the second project is to show what the limits of this kind of knowledge are, knowledge that is based on intuitions and to prevent it from entering the realm of morality and religion. We start with the first project and its crucial device that is also of major concern for Kant's psychology and thus our topic - the device of transcendental apperception.

## 1.2. Transcendental Apperception

In the first part of the *Critique*, *Transcendental Aesthetics*, Kant explains how our senses work and claims to have discovered time and space as *pure intuitions*, as a kind of "filters" our senses use when encountering brute sensations, frames that are used to encompass them all and make them apt for human experience. Kant defines an *intuition* as "*an objective perception... [that is] singular and immediately related to object*"[[9]](#footnote-10). In other words, intuitions are sensations that have been processed by our senses, that went through "filters" of space and time, but of which we still have no concept; intuitions represent mere sensory experience that we have become conscious, but have no discursive knowledge of. Space and time are pure intuitions, a priori intuitions, that are the necessary, but still not sufficient conditions of experience as we know it. Everything we sense, we sense in space and time, which are not absolutely objective properties of things-in-themselves or relations between them, but are pure forms of intuition that serve as a foundation stone of our sensory experience. Space is defined as the form of the outer sense, while time is defined as the form of inner sense, which is a distinction we will later touch upon again, but is not of relevance at this point of the paper.

What else is important is the faculty of *understanding*, which is analyzed in the second part of the book, *Transcendental Analytic.* Without understanding, senses would simply be passive recipients of sensations that might somehow fit into pure forms of space and time, but would be completely disconnected and unintelligible without the faculty of understanding that reads another "portion" of a priori rules into what is received by the senses. Therefore Kant calls senses the faculty of receptivity, while understanding is called the faculty of spontaneity, because thanks to understanding we spontaneously engage into the *act* of prescribing the a priori rules of our mind to the material we *passively* receive by the senses, through pure intuitions of space and time. How does this enterprise work and what are these rules of understanding?

These rules are a priori concepts, or *categories,* that lie beneath any experience and make it possible. How do we deduce these rules? Kant's answer is, on the one side, suspicious and widely disputed, and on the other side, original and very fruitful. We should rely on the basic rules of our formal logic. Kant's method is to first extract the 12 basic types of judgments.[[10]](#footnote-11) The fact that, by investigating the rules of formal logic, we arrive at these 12 basic types of judgments and cannot define any more or reduce them to any less is indicative. For Kant, that simply means that these 12 types represent the necessary and universal ways in which we think, in which we relate the objects of experience one to another in our representations. We cannot make any statements that would not be definable as one of these. We cannot connect our experiences, cognize them or assert anything about the objects in the world in any other way but these 12. Based on these 12 types of judgments we can define 12 basic categories of understanding, 12 a priori concepts we inevitably use every time we process what was passively received by the senses to make sense of it.[[11]](#footnote-12)

Of course, in several important ways, Kant's theory is anachronous. Insofar as the science of logic went through a huge development in the 19th and 20th century, conceiving of these 12 types of judgments as necessary, universal, unvarying and not dependent on the initial choice of axioms and definitions is very out of date. Also, as Strawson points out[[12]](#footnote-13), it is not just the case that we have no reason whatsoever to believe that it is these 12 types to be considered as an absolute basis of all human thought, but the further development of logic actually proves that some of these are interdefinable and thus reducible to more primitive concepts. However, rejecting Kant's theory because of these facts would mean neglecting the historical context of Kant's philosophy, as well as farsightedness of the idea that lies behind the concept of transcendental categories.

And the idea is, as was pointed out earlier, that it is our minds that play an active role in our experience, thus rejecting to assume that our role is one of a mere passive receiver of the outside sensations that are somehow mechanically imprinted in our consciousness. Categories are "concepts of objects in general", that "*lie at the ground of all experiential cognition as a priori conditions, consequently the objective validity of the categories, as a priori concepts, rests on the fact that through them alone is experience possible (as far as the form of thinking is concerned). For they then are related necessarily and a priori to objects of experience, since only by means of them can any object of experience be thought at all."* [[13]](#footnote-14) Still, in order to work, Kant's conception of human experience needs one additional element that ties all the categories together - transcendental apperception.

Kant considers that Hume's problem of disparity of perceptions in an ordinary mind would still not be solved just by mere application of categories to what is perceived. Best we could have (and it is not too probable that we could even have that) are some kind of disconnected contingent sequences and associations, without any possibility of establishing stable objective connections between different representations. For instance, in order to apply the category of causality, we would need something prior to the category to make an initial connection between disparate intuitions. For the fact that an event A is close in time and space to an event B cannot be established unless there is a unified mind that would be able to place event A and event B in time and space which seem initially completely unrelated. If *my* consciousness of one event has nothing to do with *my* consciousness of another, then, for me, there is no basis or reason to ever connect these events. If it is me who is conscious of event A and then at next moment someone else who is conscious of event B, no matter how close they are in time and space, contact between the two does not exist. Therefore, the essential demand of the theory in order for categories to be able to work is to have a *unified consciousness* behind all intuitions and judgments that would be the fundamental point of contact between diverse representations. It is presupposed if the possibility of applying categories to given sensory data is to hold.

Establishing these connections is the task of one of the central chapters of the *Critique* - *Transcendental Deduction of Categories*, for which Kant says that "*It is the exhibition of the pure concepts of the understanding (and with them of all theoretical cognition a priori) as principles of the possibility of experience, but of the latter as the determination of appearances in space and time in general - and the latter, finally, from the principle of original synthetic unity of apperception, as the form of the understanding in relation to space and time, as original forms of sensibility*."[[14]](#footnote-15) The aim of the *Transcendental Deduction*  is therefore to show how categories "drive" our experience in a fashion that is universal and necessary, as well as to prove that it is the one united consciousness behind all intuitions and judgments, the one and united "I" that synthesizes them all and keeps them together. This primordial unity is the first condition of any sensible knowledge a human mind can obtain - how am I supposed to make connections between objects and events distant in space and time if there is no one and the same consciousness that is aware of all of them? It is perhaps trivial to say that I cannot make connections between things that do not belong to my consciousness, but this is an assumption that has very important consequences in Kant's philosophy.

The main consequence is that some form of self-consciousness, of apperception, is necessary in order to have any kind of experience. But this transcendental type of apperception, which follows all our representations and lies in the basis of our experience, is not itself an experience, is not a kind of knowledge in the ordinary sense[[15]](#footnote-16). It is the spontaneousact of self-consciousness - the expression "*performance*"[[16]](#footnote-17) that Susan Mendus uses might be of help when trying to understand it. We are implicitly conscious of ourselves in any of our representations; it must be possible for the transcendental "I" to follow any representation that I have, otherwise we have no reason to classify it as mine nor I could possibly obtain any knowledge from such a representation. It is important to point out that the act of representing makes us aware of three things: firstly, the unified and always-present "I", secondly, our particular representation, and finally, the object of representation - we necessarily have all these in a single act of consciousness.[[17]](#footnote-18) Indeed, we cannot even have the consciousness of the transcendental "I" outside of representations of some kind of objects - "*this awareness of myself as myself is not via some independent representation of myself. The only representations involved are representations of the objects of which I am aware.*"[[18]](#footnote-19) The awareness of myself as united throughout all my representations and as the unifying factor of representations can come *only* through these very same representations of objects. These two cannot be disentangled - in classical terminology, we could almost say that the unified "I" is the *ratio essendi* of our representations, but our representations are *ratio cognoscendi* of the unified "I".

Now we arrive to an important issue of objectivity. It was mentioned in the introduction of the paper that the main objective of this chapter is to "save" physics, to save objective validity of law-like connections in the world. The aspect of Kant's theory that is in charge of doing this is the fact that transcendental apperception, categories, as well as time and space are universal and necessary conditions of experience. This means that they apply to the entire human kind or any human-like sentient being. "*We must necessarily ascribe to things a priori all the properties that constitute the conditions under which alone we think them. Now I cannot have the least representation of a thinking being through an external experience, but only through self-consciousness. Thus such objects are nothing further than the transference of this consciousness of mine to other things, which can be represented as thinking beings only in this way.*"[[19]](#footnote-20) So these conditions of experience do not lead into some kind of solipsism in which only one subject is supposed to perceive things in a certain way (or in any way), because these are universal rules of cognition in the human kind, and this is what gives our knowledge, which is based on these rules, objectivity. How can we show this? By making use of the fact that physical and mathematical truths exist and we all recognize them. But was not that supposed to be proved in the first place? It seems at a first glance like Kant is running in circles - proving physics by appealing to categories, yet proving the reality and objectivity of categories by appealing to the existence of physics.

The solution seems to be that all Kant wants to show is that there is a cognitive mechanism that can support the existence and validity of physics and mathematics, to save it from the conclusions that can be drawn from the philosophy of empiricism. One of the main tasks of critical philosophy is to show that physics is *possible,* and it seems to be reasonable to believe its results as long as there is a mechanism that can prove it is not paradoxal. If there is then a choice whether to believe we apparently all use the same a priori rules and get the same results in physical experiments or to believe that this is still all a matter of habit or some kind of preestablished harmony, then Kant's choice is obvious, although, I reckon, not necessary.

In any case, validity of statements about objects comes from its universality among humans. Thus we could talk about two senses of the term "objectivity" - the first one more colloquial and everyday, in which what is "objective" about the world is what we could say about the thing-in-itself, which is the way of speaking Kant rejects, since nothing can be possibly said about "objects" in that sense. In fact, from Kant's perspective, any kind of communication would be impossible if this was the true meaning of "objectivity". However, there is a different, Kantian sense of objectivity, and in this sense, phenomena as perceived by us, through the "filters" we apply, are objective, since these "filters" are the same for all of us. Thus Kant writes about the "empirical reality" of space and time and puts much effort to distinguish a *phenomenon* from an *illusion*. An illusion in this sense is something that cannot be verified by the others and is something essentially private, while statements about phenomena are subject to corrigibility. Thus I quote a very important passage from the *Aesthetics*: "*If I say: in space and time intuition represents both outer objects as well as the self-intuition of the mind as each affects our senses, i.e., as it appears, that is not to say that these objects would be a mere illusion. For in the appearance of the objects, indeed even properties that we attribute to them, are always regarded as something really given, only insofar as this property depends only on the kind of intuition of the subject in relation of the given object as appearance is to be distinguished from itself as object in itself."[[20]](#footnote-21)* Therefore, objects *as appearance* are still valid objects of our judgments, in fact the only valid objects of our judgments, since the object in itself is not an intelligible concept. It seems perfectly simple at the first glance - we have the same a priori concepts and intuitions and we use them to communicate empirical concepts that are formed after our senses are affected by the actually-existing objects.[[21]](#footnote-22)

 Nevertheless, Kant manages to complicate the matter further by introducing the distinction between reproductive and productive imagination[[22]](#footnote-23), or association and affinity of the manifold. This distinction is more explicitly and confidently defined in the *Prolegomena[[23]](#footnote-24)*, which was written between the two editions of the *Critique*, as a distinction between judgments of experience and judgments of perception.All these distinctions come down to the same thing - one of them is the basis of "objectively" established connections among different representations, while the other has only the "*subjective validity*" and is driven by mere "*laws of association*"[[24]](#footnote-25), that are seemingly meant to resemble Hume's idea of association of perceptions. So after all, now it seems that there can be *some* kind of connection between different representations that are *not*  based on a priori rules and that lack objective validity. Let's see in more detail what Kant has to say about this.

A good start to this exposition would be to quote Bennett who calls Kant's use of the word "object" as "*viciously ambiguous*"[[25]](#footnote-26), and it lives up to this reputation. The best way to look into the matter is by quoting one of the classical passages of the *Deduction* regarding this problem: "*In accordance with the latter* [subjective validity] *I could only say 'If I carry a body, I feel a pressure of weight', but not 'It, the body is heavy', which would be to say that these two representations are combined in the object, i.e., regardless of any difference in the condition of the subject, and are not merely found together in the perception"[[26]](#footnote-27)*. The analogous distinction made in the *Prolegomena* comes down to the similar thing: what we claim only about our representations has subjective validity, but if we go beyond that and claim something about objects, not just our representations, it is something we claim to be necessarily true about objects of experience.

However, there are a few problems with this distinction. As Bennett points out[[27]](#footnote-28) first of the Kant's claims - '*If I carry a body, I feel a pressure of weight'* is still a claim about objects, since it involves the concept of a body, and a similar mistake is made in the examples in *Prolegomena[[28]](#footnote-29)*. Bennett also points out that the only difference we could possibly make between the two is that the latter claim involves greater causal commitment, since it asserts something about a body in general, while the former simply asserts that at the moment of carrying a body I have a certain sensation. It seems that the point is that judgments of perception cannot be put to test by a third party, while the judgments of experience could. Nevertheless, it still does not change the fact that both of these statements are about objects, not mere representations, and that both use the cognitive machinery based on transcendental apperception and the application of categories.

Furthermore, the way Kant treats the concept of object and our (in)ability to assert anything without actually claiming something about objects make the problem even more obscure. Namely, when Kant tries to explain where the possibility of this mere association of intuitions - that does not claim anything about objects - comes from, it turns out that this possibility, which is usually due to some function of the imagination, is inevitably based on the synthetic unity of transcendental apperception. This faculty of merely associating intuitions or representations in general is not a separate and independent faculty of mind but depends on the more basic association, a transcendental one, that is used to claim something about objects.[[29]](#footnote-30) Most notoriously, Kant claims that "*all representations, as representations, have their object, and can themselves be objects of other representations in turn*"[[30]](#footnote-31), thus making this at least a two-fold problem. Firstly, if all representations necessarily have objects, then claims about representations do claim something about objects as well. And secondly, if representations themselves can be objects of our experience, then claims about representations are claims about objects. But then again, how is my claim about my representation to be "objective" if there is no third party that can confirm or refute it, since these representations are private? The confusion here exists at many levels.

It is clear that Kant's use of the word "object" gives us a hard time in determining what exactly he is thinking when making a difference between claims that can have only subjective validity and judgments which have an objective validity (for the latter Kant implies in B142 that they are the only proper type of judgments). It is not clear what kind of significance (if any) would the "subjective" claims have for Kant's theory. On the contrary, if Kant's main interest in this part of the *Critique* is to "save" physics, then this is exactly the type of claims that have to be considered irrelevant, because they claim nothing about the empirical reality, but only about our inner states. This type of claims could even be called "Humean" since the discussion about them is an obvious allusion to Hume, who believed that these assertions are the only ones that can be used to express any sensible thought about the world.[[31]](#footnote-32) Thus dealing with this problem seems to perfectly fit into Kant's interest of saving physics, and it seems that this is the context in which this aspect of his analysis of objectivity should be considered.

Of course, the influence of pure concepts of understanding should not be overestimated when talking about physical laws. Pure concepts, as well as the synthesis of transcendental apperception as their root, are a necessary, but not a sufficient condition for physical laws. They just give our minds the possibility of ever making any connections between disparate intuitions, concepts and representations, but they only give us the most general rules of experience that nature necessarily obeys. But in order to establish particular physical laws it is necessary to go beyond that and to perform actual empirical observations upon which we can ground our actual knowledge about nature. "*Particular laws, because they concern empirically determined appearances, cannot be completely derived from the categories, although they all stand under them. Experience must be added in order to come to know particular laws at all; but about experience in general, and about what can be cognized as an object of experience, only those a priori laws offer instruction*."[[32]](#footnote-33)

## 1.3. The Second Project

As I pointed out before, in this paper I follow the idea of Andrew Brook about the "two projects" of Kantian psychology. The first one was saving physics as a body of necessary and universal truths, and it was developed in the previous chapter. The idea of the second project, according to Brook, is to "*insulate religion and morality, including the possibility of immortality and free will, from the corrosive effects of this very same science*."[[33]](#footnote-34) I do not agree completely with this claim - Kant does see this job of insulating primarily as something that should ensure independence for morality and religion, but science also benefits largely from this process, and this is also of great interest for Kant. Nevertheless, I do agree that the two projects seem to be concerned with different objectives, and that the *primary* concern of the second one is of practical nature.

 Now, what does this "insulation" mean? It means that the person has to be somehow partly isolated from the natural world if we are to conceive of it as having free will, as being autonomous. Kant explains in the *Third Antinomy[[34]](#footnote-35)* how freedom of will and necessity of natural laws are not necessarily in collision and that they can exist together. However, to make this plausible first he has to restrict the order of nature from being able to answer certain questions about our soul and personality. Only after this is done he can propose the concept of "*kingdom of ends*"[[35]](#footnote-36) and support it with the concept of free will, because otherwise actions of human beings are nothing more than any other event in the chain of causally determined events and thus they take no responsibility for their actions.

Kant's way to put this restriction to work are the *Paralogisms of Pure Reason*. Paralogisms are one of three types of transcendental illusions that necessarily appear in human mind and have to be discovered and controlled if we are to properly delimit the authorities of physics and metaphysics, or theoretical and practical philosophy. "*One can place all illusions in the taking of a subjective condition of thinking for the cognition of an object*."[[36]](#footnote-37) Therefore, it could be said that all illusions happen because we confuse the thing-in-itself and the phenomena - as it was said before, the phenomenon itself is not an illusion, but taking the phenomenon to be some kind of a thing-in-itself that is not subject to the a priori rules of the human mind is an illusion. And the three most important, necessary illusions of the human reason are the paralogisms, antinomies and the ideal of pure reason, all of them in some way or another related with our mind being convinced that the conditions of our experience tell us something about the objects themselves or the way they are, in themselves, independently of our experience, ordered.

What is of our concern here are the paralogisms, since they tell us something (or rather, prohibit us from saying anything) about the human soul and the identity of the subject or a person. The first paralogism is the paralogism of substantiality, the second of simplicity, the third of personality and the fourth of ideality. The first one claims that "*the soul is a substance*", the second one that it is "*in its quality, simple*", the third one that it is "*in the different times in which it exists, numerically identical i.e., unity (not plurality)*", and the fourth that it is "*in relation to possible objects in space*".[[37]](#footnote-38) All four paralogism basically come down to the same thing - the fact that there is an "I" that connects all our representations and unifies them into one sensible experience gives us no knowledge about this "I"; not only that we cannot claim anything about its properties, but we cannot even claim that it is a substance, a *thing*. Clearly, of most interest among paralogisms for this paper is the third paralogism, because it claims something about identity and unity of the soul and this discussion will take the most of this section of the paper.

 The first paralogism, however, seems basic compared to other ones and is also very convenient for the illustration of what the machinery of paralogisms is all about. Let's take a look at its structure:

"*I, as a thinking being, am the absolute subject of all my possible judgments, and this representation of Myself cannot be used as the predicate of any other thing.*

*Thus I, as thinking being (soul), am substance*"[[38]](#footnote-39)

It is clear where the conflation happens. From being the subject of all my judgments, from the fact that all my representations are united due to the "I" of transcendental apperception as the focal point of my experience, I infer that this "I" has to be a thing, has to be something that has the ontological status of a substance as well as definable properties.[[39]](#footnote-40) But this is certainly not true. For there is no point in time in which I observe myself *as subject* as an object, which might again be a trivial claim, but it again has immense consequences on the theory. The fact that I cannot observe this transcendental "I" that is simply an act of synthesis *as an object* then gives me no authority to claim something about it. For it is only possible to have knowledge about something through intuitions, through sensory experience. Do we have any intuition of this "I"?

 The answer to this question is the basis for Kant's distinction between the transcendental and empirical self, already present in the *Deduction* chapter. Since this is the paragraph central for the understanding of the paralogisms in general, I quote it in full: "*Now this original and transcendental condition is nothing other than the transcendental apperception. The consciousness of oneself in accordance with the determinations of our state in internal perception is merely empirical, forever variable; it can provide no standing or abiding self in this stream of inner appearances, and is customarily called inner sense or empirical apperception. That which should necessarily be represented as numerically identical cannot be thought of as such through empirical data. There must be a condition that precedes all experience and makes the latter itself possible, which should make such a transcendental presupposition valid.*"[[40]](#footnote-41)

The logic of this distinction is, thus, very similar to the logic applied when showing how our experience cannot be connected if there is no synthetic act that precedes and organizes the reception of outer sensations. The difference is that here we talk about *inner* sensations or affections. In the similar way to that in which we observe the outer world in space, we observe ourselves, the always-changing affections of our soul in time. On the same token, when we make the soul the object of our experience, in the intuitions of soul alone there is nothing to convince us that there is any kind of unity among the affections of our inner self. The soul, as a mere object of empirical analysis, is a just a series of affections with no objective relations among them. This self, the empirical self, is an object of the empirical apperception - it is simply a conscious representation of a series of my own mental states, something we might today call introspection.

However, the empirical apperception, in order to be even possible, requires some formal principle of unity among intuitions of the inner sense, and this unity is given by the transcendental apperception. If we could gain any intuition about the transcendental apperception, then it would need another unifying principle to make connections between intuitions, and so on - this regress would have to stop somewhere eventually. On the contrary, we have no perception of the transcendental apperception since it is not given to us as an object, and thus we have no knowledge of it.[[41]](#footnote-42) Nevertheless, we can conclude that there is such a principle from the very possibility of a connected, sensible and communicable experience gained through both inner and outer sense. Transcendental apperception is the self-awareness of the subject as the subject, while empirical apperception is the self-awareness of the subject as an object.

Now we go back to the conclusion that from the transcendental "I think" that connects all our representations we cannot infer to the knowledge of its mode of existence or its properties. This mistake is so easy to make (and was made numerous times) simply because it is very easy to conflate the two modes of self-awareness and claim that the conditions of our experience - e.g. existence and unity of transcendental "I" - mean that this "I" is a thing that is unified, which we observe in the inner sense, and about which we can obtain knowledge. Other paralogisms seem more-less similar to the first one and can partly be explained upon the explanation of the first, with more specific applications - they are not about substantiality of the "I", but about its simplicity, unity and relation to the objects.

When it comes to the third paralogism, what we can say about what is the meaning and scope of the unity of the apperceptive "I" and whether it guarantees personal identity in the more modern sense is a widely discussed issue among modern interpreters of Kant. First of all, let's see the exact structure of the paralogism:

"*What is conscious of the numerical identity of its Self in different times, is to that extent a person.*

*Now the soul is etc.*

*Thus it is a person*."[[42]](#footnote-43)

Beatrice Longuenesse gives essentially the same but a more comprehensible version of the argument in which the implicit but omitted part of the argument is nicely explicated. This is essentially the same argument but is more intelligible and easier to work with:

"*What is conscious of the numerical identity of itself in different times is to that extent a person*

*Now I, as thinking, am conscious of the numerical identity of myself in different times.*

*So, I, as thinking, am a person*."[[43]](#footnote-44)

Now the paralogism here hides in the ambiguity of the middle term which is "*being conscious of the numerical identity of oneself"*. The ambiguity consists in the fact that in one of the premises the term is used in a transcendental sense, and in the other one in the empirical sense. Namely, from my awareness of the unity of transcendental apperception, no consequences follow regarding my empirical self. From the unity of the transcendental self, from the fact that there is one "I" that unites all my representations, we cannot derive any kind of empirical identity. Thus, logically speaking, the conclusion - that from the transcendental "I" we can infer to the objective identity of the empirical "I" - cannot be valid, since the middle term is not well-defined.

Now what this objective, "empirical identity" consists in is a subject of a great debate. I will present two main interpretations: for the first one I will use the term *substance* *interpretation*, and for the other one *personal identity interpretation*, since those who stick with the first one think that what is denied is the unity of the person/subject as a unity of a certain substance, while those who stick with the second one hold that what is denied is personal identity in the modern sense.

Although the entire First edition chapter about the third paralogism (A362-A366) is extremely dense and full of very important and/or problematic claims, for now I will quote just an essential part of the probably most problematic passage of this part of the *Critique,* especially given that it is very convenient to make use of to explain the substance interpretation: "*An elastic ball strikes another one in a straight line, communicates to the latter its whole motion, hence its whole state (if one looks only at their position in space). Now assuming substances on the analogy with such bodies, in which representations, together with consciousness of them, flow from one to another* [...] [the last one] *would share the states of all previous ones, together with their consciousness and its own. The last substance would thus be conscious of all the states of all the previously altered substances as its own states* [...] *and in spite of this it would not have been the very same person in all these states*."[[44]](#footnote-45) The very first problem with interpreting this, as well as the other problematic passages in this chapter is the ambiguity of the terms person, subject, Self and soul. This seems to be the starting point of the most of the debates - if Kant had had clear-cut definitions he stuck to the whole time, the interpreters would have probably had much better ideas about what is his position on this problem. However, he did not, so he left them guessing. The first guess, as I mentioned, is that what he had in mind is similar to the point of the first paralogism, but with the special emphasis on the problem of unity.

###  1.3.1. The Substance Interpretation

The substance interpretation holds that "being the very same person" as used in the aforementioned paragraph means that "'*The soul is an enduring substratum of mental states' or 'Across-time identity-statements about the soul assert the persistence through time of a substratum.'*"[[45]](#footnote-46) Therefore, according to them, when Kant uses the term "person" he refers to an underlying substance that is supposed to hold together all the mental states of a person. Usually it is held that Kant had in mind an immaterial substance, although even for material substances the point remains the same.

And the point is that if we imagine our representations (together with the sense of "I" and awareness of these representations being ours) being transferred through a series of soul-substances, the "I" would not stop being aware of those representations as its own as soon as it was transferred from one substance to another. According to Kant, we are able to imagine our whole mental life being transferred through a number of these substances with the sense of self-awareness intact (the same way motion is transferred from one elastic ball to another, but it remains the same motion). Therefore, we cannot claim anything about any necessary connection between the "I" and representations that fall under "I" on one side and any kind of substance on the other. The unity and stability of "I" do not mean that it "draws" this unity and stability from a unified and stable substance, since it is absolutely not necessary in this picture.

Strawson, who believes in the substance interpretation, reckons that Kant is not seriously proposing or considering such a possibility. He simply thinks that this is a particular way of dealing with the futility of a belief in an immaterial substance that would be necessarily related to the transcendental "I" and our representations. The same way we can imagine these immaterial substances as underlying the soul, we can also imagine our whole mental life being transmitted from one immaterial substance to another - in fact, we could not be able to make a difference between these two situations. How could we know anything about the underlying immaterial substance if we have no perception of it, and how do we gain knowledge about whether our soul "changed" its underlying immaterial substance or it is still the same? It is exactly the futility of a Cartesian-like rational psychology that Kant wants to show, not the actual possibility or plausibility of such events.[[46]](#footnote-47) On the same token, Bennett is suspicious that this whole passage could be plausibly read as "ironic" or even "mocking" if we stick to the substance interpretation.[[47]](#footnote-48)

Bennett claims the same about Locke's account of the personal identity, which strikingly resembles the previously quoted passage from the *Critique*. Among other things, Locke writes that "*for the same consciousness being preserved, whether in the same or different substances, the personal identity is preserved."[[48]](#footnote-49)* In the same section Locke treats the problem Kant also mentions - the interruption of consciousness.[[49]](#footnote-50) Someone who claims a soul to be a thinking substance would have problems establishing what happens when this soul's consciousness is interrupted - does it cease to exist, and when it "comes back", does it remain the same substance? Kant and Locke do not have to bother with this - Kant would claim that it is the same person as long as there is the same "I" accompanying its representations while it is conscious, while Locke proposes a similar but a somewhat more elucidated explanation with less metaphysical burden - in order to conceive of someone as the same person the only condition is that there it is the "same consciousness" in this person, which in practice means that complete oblivion or amnesia mean the interruption of this identity.

Kant's and Locke's conclusions about some of the points on this issue are highly similar and there is no terminological confusion in Locke's theory between concepts of a person, a man and a soul - it is clear that when he speculates about "whether the consciousness of past actions can be transferred from one thinking substance to another" he is contemplating on something very similar to what Kant writes about, according to those who hold the substance interpretation to be true. No matter whether any of them or both of them are ironic, mocking or serious, it seems that if we accept Kant's problematic passage, at least partly, as an allusion to Locke, substance theory becomes more plausible.

Moreover, Corey Dyck claims that the third Paralogism is a clear reference to Kant's own *Lectures on Metaphysics* and the views he used to hold true before writing the *Critique*. He even claims that to Kant's remark about inability to infer from the transcendental "I" to the "objective persistence of self", a rationalist psychologist could reply "*along the lines of Kant's previous expositions in the ML1 notes that my own consciousness of the I of the I think proves the identity of my substance since, if I was comprised of consecutive substances, I could not be conscious of my numerical identity.*"[[50]](#footnote-51) All these possible allusions by Kant - to Descartes, to Locke, even to a younger self, seem so support that the substance reading might be credible, since all those are quite possibly references to the problem of the transfer of consciousness among multiple substances.

However, there are at least two obvious problems with the interpretation. The first one is the "third person" problem, the fact that Kant deals with the notion of an outside, third-person authority on the matter of identity of the self. "*For we cannot judge even from our own consciousness whether as soul we are persisting or not, because we ascribe to our identical Self only that of which we are conscious; and so we must necessarily judge that we are the very same in the whole of the time of which we are conscious. But from the standpoint of someone else we cannot declare this to be valid, because, since in the soul we encounter no persisting appearance other than the representation "I", which accompanies and connects all of them, we can never make out whether this I (a mere thought) does not flow as well as the other thoughts that are linked to one another through it.*"[[51]](#footnote-52) Here the term "soul" seems to be used in the sense of an immaterial substance underlying our experiences. The claim about us not being able to know anything about its persistence is still nothing problematic, but it is hard to make sense of the claim that a third person can help us resolve this.

Namely, if the soul really was an immaterial substance, how could anyone tell whether I am the same soul as yesterday, given that there can be no knowledge without intuition? How does another person then know whether I am the very same in the whole of time I am conscious of or not, in any of the two senses? On the other hand, if Kant had a material substance in mind here, then I am as good of an authority as anyone else. For if I - my entire mental life and my sense that it is mine - am transferred from one body to another (which is probably not what Kant had in mind anyway), I imagine I would be easily aware of that fact. At least as well as anybody else. It could be, then, that the claim about the "persisting soul" or identity is not a claim about substances after all, but about something else.

A somewhat strange terminology used, for instance in A363, where Kant writes about the "replacement of the subject", in which the identical-sounding I "*still keeps in view the thought of the previous subject, and could also pass it along to the following one*"[[52]](#footnote-53) also supports this claim. For throughout the Paralogisms, Kant usually refers to the thinking substance as the "soul" or "person", so when writing about the replacement of the "subject", that might suggest that he is not referring to a mere transfer of my representations and the "I" that connects them from one substance to another. Let's see what the proponents of the personal identity interpretation have to say about this problem.

###  1.3.2. The Personal Identity Interpretation

The essence of the personal identity interpretation is the claim that the unity of the transcendental "I" does not entail personal identity. Namely, the proponents of the interpretation think that it is possible to refute someone who claims to have had a certain experience or to have a certain memory, by saying that is not his or her experience or memory, but someone else's. This then would be the explanation of what Kant is implying when writing about the "I" that even in the "replacement of the subject still keeps in view the thought of the previous subject".

The claim amounts to the imaginable possibility that if someone's memories are transferred into my consciousness I will think of them as mine. Or, to be more precise, Kant claims that "*despite the logical identity of the I - a change can go on that does not allow it to keep its identity.*"[[53]](#footnote-54) Moreover, he writes one more thing about the infamous "observer" - "*This from the I that accompanies - and indeed with complete identity - all representations at every time in my consciousness, although he admits this I, he will still not infer the objective persistence of my Self*."[[54]](#footnote-55) The substance interpreters read the "persistence of Self" as the persistence of a substratum, but the personal identity interpreters read it as the identity of my subject throughout the entire temporal sequence that is encompassed by the transcendental "I". The lack of the "persistence of Self" thus designates the impossibility of being certain that, for instance, a memory I have is *mine*, although I have a representation of this memory as if it was mine. Unity of self does not mean the identity of the self throughout time - I could have someone else's memory and feel that it is mine. The same-sounding "I" that encompasses all our memories does not guarantee that a certain memory is ours, although it makes us feel that it is and we cannot help it. This would then do some harm to what Parfit calls the "psychological continuity".[[55]](#footnote-56)

If personal identity interpretation is correct, then it turns out that Kant somehow foresaw the discussion about the possibility of the q-memories, memories that we have but are not ours. In the thought experiments of the modern philosophy of mind it is usually considered that these memories are artificially implanted in our consciousness, either through brain surgery or with the help of different softwares that would enable this. What is claimed here then is that the transcendental "I" would in no way help me distinguish between memories of events that happened to me (thus memories initially *obtained by* *me*) on the one side and memories of events that actually were not obtained by me, but when implanted into my consciousness they were encompassed by the transcendental "I" and I was *made to believe* that they were obtained by me on the other side.[[56]](#footnote-57) Hence my personal identity would be questioned - I could not be sure that any of my representations originally belonged to me, apart from the ones I am aware of at the moment. I would have a representation of the memory that only occurs to me as a memory *right now,* but was obtained by another subject and incorporated into my consciousness afterwards. The function of the transcendental "I" would thus be limited to the function of synchronic unity, but nothing more than that.

This is the view held by Brook, for instance - "*for unity of consciousness, I must q-remember earlier representations and actions as though they were mine - but they need not have been.*"[[57]](#footnote-58) Q-remembering a state is having a memory of a state that was not originally mine but was transferred to my consciousness and, according to Brook, something like it is not only allowed by Kant's theory, but can be seen as a real possibility. He claims that his unity is always only synchronous and can guarantee the identity of myself only at the given moment: "*Suppose a memory of something I saw yesterday enters a global object. Neither yesterday's object nor yesterday's perception of it enters. What enters is the content or object of a memory; the memory exists now, not then*."[[58]](#footnote-59) At any given moment we are conscious it is plausible to believe that all the memories were merely fabricated as ours, because whatever content enters our mind we cannot help but conceive of it as ours since it falls under the synthesizing power of the transcendental apperception. However, a third-person authority could give us empirical proofs that it is not ours - for instance, I could have memories about being at the other side of the world yesterday, but someone can give me very good proofs that I never got out of my apartment. Thus the mistake is not only in me making a mistake whether I was *here* or *there* yesterday - I am also subject to a different type of error with a different emphasis: whether or not it was *I* who was somewhere yesterday, or is it *someone else* (Shoemaker calls this "misidentification relative to the first person pronouns"[[59]](#footnote-60)). This is perfectly imaginable in the personal identity interpretation.

However, this interpretation is subject to serious problems as well. First of all, let's see how it deals with the same type of argument that was used to test the substance interpretation, the third-person argument. Kant claims that I necessarily judge that it is I throughout all my representations and only a third-person observer can claim differently. Nevertheless, on the personal identity interpretation, as well as on the substance interpretation, it is perfectly imaginable that I can find out myself that it was not me who had a certain representation yesterday, although I feel as if it was me and I have a memory of an event as if it was obtained by me. If another person can obtain empirical evidence that it is not the case, why would it be impossible for me to figure it out myself when faced with the evidence, without any help from others? However, the defense of a personal identity interpretation proponent is also imaginable - I could never actually accept this kind of evidence and nothing can convince me that something is not my actual memory although it seems as if it is. As Mendus remarks: "*I cannot but regard myself as identical through time, but this judgment is not an inference from an awareness of a continuing subject of experience: this is not an inference from an awareness of an object of any sort*."[[60]](#footnote-61) An observer can notice that, maybe even convince me with strong enough pieces of evidence that something peculiar indeed happened, but from a first-person perspective I can never escape the notion of that memory as mine, since it remains under the influence of the synthesizing "I". Even when we reasonably believe (from a "third-person" perspective) that the evidence is strong, I still remember (or rather, q-remember) an event as happening to me or an object as being actually observed by me. But if this is so then there are more serious consequences, those concerning the problem of objectivity.

Brook claims that all this means that "*represented representations would be as distinct from representations in themselves as represented objects are from objects in themselves*."[[61]](#footnote-62) It seems that Brook introduces another layer of experienced reality that Kant hardly had in mind - firstly, there are things-in-themselves, then secondly, representations of objects that have nothing to do with the things-in themselves, and finally, representations of representations obtained through our inner sense that have nothing to do with the second layer of experience. This means that we have representations of objects only as they appear to us right now, while about representations of past objects we know nothing regarding the question whether they were originally ours or not.

Two immediate problems arise instantly. As always in similar theories, the first one is the problem of defining how long is "right now" and the second one why should what is happening "right now" have a kind of ontological priority over all other things that happen in time. The first problem is more rhetorical because there is no reasonable border that would help us demarcate the "right now" from "just an instant ago", but is a definitely a serious problem when we try to make an ontological difference between the two (and I believe this is exactly what Brook is doing). The second problem is, however, even more serious.

If we could only be sure about what is happening right now, but could not make any objective claim as to what happened ten minutes ago (objective in the sense that I suppose my memory is not simply my q-memory and refers only to my representation, but also directly refers to objects of appearance and can be compared and communicable to objects of experience of others), then most (if not all) of our claims about nature would be invalid and absurd. Brook quotes that Kant writes how "*time cannot be outwardly intuited*"[[62]](#footnote-63) and is just a form of our inner sense, which is true, but it does not mean that it has no role in the knowledge of the outside world. We can cognize anything about the world of appearance exactly because we sort the intuitions of the outer sense according to the rules of the inner sense of time. The material we get from the outer sense has to be ordered before we can obtain any knowledge about it, and these two processes ("obtaining" sensory material and ordering it in time) are parallel processes and cannot be separated from one another. As Kant notes in the *Aesthetics:* "*It* (time) *is only of objective validity in regard to appearances, because these are already things we take as objects of our senses*"[[63]](#footnote-64). It is important to have in mind that *only* in this sense objectivity is employed in the *Critique* regarding the objectivity of natural science and our knowledge in general. Time is objective in regard to appearances, but this is the only sensible meaning of objectivity for Kant.

Kant also claims that "*Our assertions accordingly teach the empirical reality of time, i.e., objective validity in regard to all objects that may ever be given to our senses. And since our intuition is always sensible, no object can ever be given to us in experience that would not belong under the condition of time*."[[64]](#footnote-65) Furthermore, any plausible interpretation of the *Schematism* chapter would support the claim that without personal identity, objectivity seems destroyed as a concept. In *Schematism* Kant explains how, in order to apply categories to the sensations and gain any sensible knowledge about the world, we need to use the schemas of pure reason, schemas that create a temporal dimension in the experience and give temporal depth to the categories. This is seen as a function of transcendental imagination, which is necessary for this job. For instance, "*the schema of the cause and of the causality of a thing in general is the real upon which, when it is posited, something else always follows. It therefore consists in the succession of the manifold insofar it is subject to a rule*".[[65]](#footnote-66) This has important consequences.

Representations in time in which I observe both myself and the world are not "representations of representations", but representations of objects.[[66]](#footnote-67) Schemas are rules of our minds exhibited *in time* and insofar as they are rules of all of us they amount to objective rules. If I was not sure about whether my state just a moment ago is really *my* state or is just a representation of a representation that has nothing to do with it, how are we supposed to have any belief in the transcendental bases of the objective knowledge? How do I perform experiments? How do I write the results of the experiment down? None of these things happen in an instant. All we would be saying about the world in general would be said about representations of representations that do not necessarily have anything to do with those of other people or with any particular object. As Bennett puts it "*the play with 'my time' and 'his time' seems to credit the observer and myself with different and non-interacting ontological frameworks.*"[[67]](#footnote-68) Depending on how we interpret this remark, more precisely, the phrase "ontological framework", this sentence has very different implications.

They definitely are different ontological frameworks insofar as they are, what we would call today, "private". We do have some kind of privileged access to our states and a first-person and a third-person view have their separate realities since *I* am under the authority of the transcendental "I", and *she* is also under the authority of the transcendental "I", but these transcendental "Is" are not the same and represent two ontologically different frameworks - between my "I" and someone else's "I" there is no point of contact in a numerical sense. We are numerically two different "Is". On the other hand, an ontological "point of contact" or a "common denominator" which guarantee interaction are the same rules, same categories, same machinery of schematism, which allow us to communicate our ideas about the world and to claim something about objects *in time,* and thus about ourselves in time. Because objects do not exist in time unless they are intuited by *me* in time, they do not exist if time cannot be "outwardly intuited" in the sense that we apply the pure intuition of time and its rules, schemas (based on the rules of categories) to gain objective knowledge. I do not see how there can be knowledge (of course, inside Kant's theory) if personal identity is not assumed. If the first project was to save physics, destroying it completely in the second project seems unreasonable.

Invoking the problem of states of hallucinations and faded memory, for instance, also could not help this interpretation, for it is not Brook's point to prove that we can forget or hallucinate and thus have a belief that something has happened to us while in fact it has not. In the case of hallucinations it is safe to say that these claims are corrigible, but only in the context of the actual content of these representations, not regarding the problem whether I was the one who had them. This is a special state of the human mind that occurs rarely, and is easily proved to be wrong by the others, by the community. But my hallucinations are still *my* hallucinations and not somehow imported into my mind from someone else's mind, thus it is safe to say that we can be mistaken about whether the contents of my representations are somehow, only *contingently*, different from those of the others. But in the personal identity interpretation lack of objectivity of our representations is not just a contingent occasional fact, it is a transcendental basis of our experience, because this interpretation is not concerned only with the problem of the concrete empirical objects of representations. It is concerned with the authority and identity through time of the subject in general, thus having much more serious consequences for the workings of our experience.

Additional counter-argument can be that at two very important places Kant uses unity and identity synonymously[[68]](#footnote-69). However, even Brook himself admits that "*Kant* [...] *did not articulate any view that unity lacks temporal depth*."[[69]](#footnote-70) The whole interpretations seems a bit like it is attributed to Kant without enough concern for the historical context of Kant's philosophy. Problems of q-memories, transfer of memories, psychological continuity and so on are modern philosophical problems. Kant did not seem to have them in mind. Susan Mendus also writes that Kant "*never mentions the criteria for correctness for such judgments and never says that they cannot be wrong*"[[70]](#footnote-71) - of course, here she has in mind judgments of ascribing my former mental states to myself instead of to ascribing them to someone else. And this seems to be the whole truth. The problem of corrigibility and misidentification through misinterpretation of the I is one of the focal issues in the modern treatment of the problem of personal identity, but for Kant it is not straightforward that this is even an issue.

To be fair, it should be said that Kant did not deny the possibility of misidentification through misinterpretation of the I. Although he never writes about this explicitly, the concept seems to follow from some of his ideas. The emptiness of transcendental apperception seems to be the main problem here. Here, Brook has a point - it would make absolutely no difference for us whether our memories are representations of representations or representations of objects - for our consciousness it is completely the same. For how could we ever tell? For all we know, all our representations stand under the authority of the transcendental "I", but whether those are *really* about objects, whether they *really* stretch in time or are just synchronous ideas that seem as if they stretch in time and refer to objects - we could never know. However, I believe that Kant's attitude can be read from the *aims* of his philosophy - if one of two main objectives of his psychology is to "save" physics, it makes no sense that pursuing the second main objective entails destructive implications for physics and any objectively valid claims in general. This seems to me to be the most serious problem of the personal identity interpretation.

Both interpretations, substance and personal identity, have its pros and cons. I will use the concept of unity of the self again in the next chapter when dealing with the dissociative identity disorder. In this exposition, I will not stick to neither of the interpretations, but will try to show that Kant's theory can incorporate the phenomenon of DID equally well on both interpretations. Before that though, I would like to explore another aspect of Kant's theory of the unity of the self, namely the possible empirical interpretations and implications that his theory might have.

## 1.4. Potential Empirical Aspects of Transcendental Apperception

Kant was definitely clear about the problem of the potential empirical cause or counterpart of the transcendental apperception - there is none and even if there was, we could not know about it. In fact, that is the pillar of the second project - if Kant is to isolate the realm of morality and religion from the realm of nature then he cannot allow an empirical root or cause of transcendental "I". The "I think" of the transcendental apperception is "*an empirical proposition*"[[71]](#footnote-72) because it has an empirical application and we gain our awareness of it through senses, and in fact without its empirical application it would not be possible to talk about it or have knowledge of it or have any kind of experience. But searching for its empirical background or counterpart is necessarily futile, according to Kant. The uselessness of the search for something "empirical" here means not just that proving transcendental apperception by physical continuity or identity is ineffective, but it also holds for psychological continuity, since our mental states are objects of our senses as much as any physical body. Thus comparing Kant's results with those of some of the most famous thought experiments about personal identity, such as Parfit's[[72]](#footnote-73), William's[[73]](#footnote-74) or Shoemaker's[[74]](#footnote-75), is probably not the way to go. Although I conceive of the Kant's transcendental apperception as something that already assumes personal identity, it is not because we infer it from any objective fact, or look for it in physical or psychological continuities or entities, or any kind of particular experience. On the contrary, it is because the concept of personal identity seems to be supported by the act of original synthesis, which in turn is responsible for me having any kind of particular experience. Nevertheless, investigating some other possible physical or empirical aspects of Kant's transcendental apperception seems both interesting and relevant for Kant's theory.

In this section I will first address an interesting article by Semir Zeki in which he tries to sketch what might a neurological explanation of a (essentially) Kantian unified consciousness look like. After that I will turn to Cassam's and Strawson's claim that Kant, had he been completely consequent, would have found that a concrete physical body is something that necessarily underlies any concept such as transcendental apperception. In the third part of the section I will briefly consider Longuenesse's remark that the practical aspect of the self in Kant demands this self to be at least partly subject to empirical analysis.

Let's first consider Zeki's article then. The very introductory sentence of the article is already highly interesting: "*In this article, I propose that there are multiple consciousnesses that constitute a hierarchy, with what Kant called the 'synthetic, transcendental' unified consciousness (that of myself as the perceiving person) at the apex*."[[75]](#footnote-76) Zeki intends to do this by suggesting an interesting interpretation of a few experiments concerned with the functional specialization in the visual brain. He refers specifically to the V4 complex and V5 complex, V4 being in charge of colour recognition and V5 being in charge of visual motion. His assumption is that these two structures are both spatially and temporally uncorrelated and relatively independent, V4 being responsible for the "micro-consciousness" of colour, and V5 being responsible for the "micro-consciousness" of perceived motion. According to Zeki, processing sites are also perceptual sites[[76]](#footnote-77), which is confirmed by the experiments in which subjects with damage to V1 area (the primary visual cortex) but with a healthy V5 complex still have some consciousness of motion.

Furthermore, colour processing and motion processing are temporally unrelated - colour is perceived around 80 ms before motion. Also, locations are perceived before colours, and colours before orientations.[[77]](#footnote-78) These micro-consciousnesses, although distributed in time and space and relatively uncorrelated, still amount to a single visual picture and over longer periods of time (around 500 ms) we do not feel this discrepancy, but have a feeling that these attributes are bound together. Zeki calls such a compound association of micro-consciousnesses a "macro-consciousness", which would in this case be a unified visual representation of our surroundings, slowly "built up" by more primitive and separate sensations of colour, motion or location, which are themselves already present in our consciousness before the act of binding.

Moreover, he suggests that in the final instance, all the different macro-consciousnesses amount to a single unified consciousness: "*micro- and macro-consciousnesses, with their individual temporal hierarchies, lead to the final, unified consciousness, that of myself as the perceiving person*[[78]](#footnote-79)." So the latter ones actually *depend* on the former ones: macro-consciousness depends on micro-consciousnesses, and the final unified consciousness depends on macro-consciousnesses. Zeki believes that this unified consciousness is something quite similar to Kant's transcendental apperception, while micro-consciousnesses seem related to the concept of the empirical apperception. The only difference is, of course, the causal order. In Kant unified consciousness has the priority over the others and is responsible for making them possible, while in Zeki's paper it is exactly the other way around.

There are at least two major problems with Zeki's idea. First one being the fact that a further build-up, from a unified visual representation to "perceiving myself as the perceiving person" might be a remarkably complex job, if not impossible. The second problem, closely related to the first one, is that even if it was done, it would still remain a question whether Kant's transcendental apperception is still qualitatively more than just a unified consciousness of all the micro-representations acquired by my senses. Kant would definitely think so, exactly because in Kant's theory the causal chain is set the other way around - as he claimed numerous times when arguing against Hume, the very fact of the synthetic unity of our experience cannot be due to any kind of collection of initially scattered sensations. That is why it has to be the other way around. However, the fact that Kant would not go along with such a theory does not mean that such theories cannot help us gain better insight in Kant's philosophy of the self.

I see the theories such as Zeki's as a very constructive and fruitful way of deepening our understanding of Kant. One of the most common objections to Kant's philosophy is that he develops its main foundations as something given, necessary and absolute, with no interest in a potential genesis or evolution that might have lead to these foundations, and also with no regard to the cultural or linguistic differences and peculiarities that might exist among different peoples and cultures. The final remark of Zeki's article thus seems to me as implying something very important for anyone dealing with Kant's philosophy: "*even though in adult life the unified consciousness sits at the apex of the hierarchy of consciousnesses, ontogenetically it is the micro-consciousnesses that occupy this position.*"[[79]](#footnote-80) We could believe in some of the main lessons of Kant's philosophy, such as the priority of the transcendental apperception over any particular experience, but still reconsider the ontogenesis (and, why not, phylogenesis) of such concepts and tools used in Kant's philosophy. How does transcendental apperception (or 'unified consciousness') evolve from the moment of birth until the moment when it is fully developed, what are its components, what are the neural counterparts of it and its components - these are all perfectly sensible questions.[[80]](#footnote-81) Kant takes transcendental apperception as given, because in a normal adult it is given and it is a very useful and ingenious way of explaining our everyday experience. But explaining the genesis of transcendental apperception, explaining physical phenomena that underlie it or its cultural or linguistic basis, are all very good ways of revitalizing Kant's theory. Same goes for some other key concepts of Kant's philosophy, both theoretical and practical. I will try to do something of the sort in the last chapter when I attempt to interpret Kant's concept of the unity of self from Dennett's perspective, that takes into account the linguistic and evolutionary aspects of this concept.

Now we move on to the second problem. Strawson's and Cassam's theories of the unity of self are based on Kant's theory but add one interesting element to it.[[81]](#footnote-82) Basically, they accept Kant's idea that one has to be aware of itself as the subject if the objects are to be taken as existing independently - there has to be a focal point of all perceptions, a point that connects and orders them all into one sensible experience. The line of reasoning is somewhat similar to Kant's: "*Without being unified by means of their actual or potential ascription to a single subject, diverse perceptions would be independent units, and no one perception would have any bearing upon any other.*"[[82]](#footnote-83) What is different in these theories is the demand that this awareness of oneself as the subject is the awareness of one's physical body.

One of Strawson's main objections to Kant is, as I wrote before, that Kant's theory is not radical enough and that it is not entirely consequent - had it been, he would have concluded himself that this self-awareness necessary for any experience is a bodily awareness. Strawson's problem is the tension between temporality and supersensibility of the transcendental subject. As it was mentioned before, although the transcendental "I" is not an object in time nor we can obtain any particular knowledge of it, it is somehow supposed to be responsible for all our sensory experience and it is supposed to be a representation that *always* accompanies all my representations, thus having a temporal dimension after all. Strawson's problem is that "*that being has no history*"[[83]](#footnote-84), and this is where the paradox lies. There is a peculiar discrepancy between a man as a natural being and his "supersensible", transcendental and apperceptive basis, a discrepancy that seems too large at the first glance. In Strawson's words, "*everything which can be ascribed to a man as a case or instance of such self-consciousness must be something which occurs in time and second, it must be conscious of himself [...] as intellectually engaged at some point, or over some stretch, of time."[[84]](#footnote-85)*. Hence it seems that any knowledge of the self and any self-ascription must be temporal, and if it cannot be so, this supersensible "I" can never be connected with the natural person, the person actually experiencing reality, since the distance between the two seems insurmountable. Therefore, if we are to believe that there is the identity of the self responsible for the unification of our entire experience, it should be the self that exists in time. Strawson's and Cassam's solution is to identify this self with the person's physical body, which is then the self to which all our experience is ascribed to, a focal point of our experience that, in a more-less Kantian manner orders and makes sense of our perceptions.

I will present one of the ways in which Kant's theory is defended from such objections, suggested by Lewis Baldacchino. He claims that there is an important sense in which we have the awareness of our transcendental "I" as temporal: "*because we are conscious of the logically necessary numerical identity of the self throughout experience in time it must be conceived as an enduring being through time*."[[85]](#footnote-86) The key, according to Baldacchino, is to be careful to distinguish between *knowledge* and mere *awareness* here. Knowledge of the transcendental "I" is definitely not obtainable, since we cannot have any perception of it in the full sense of the word. Yet we can still have an awareness of it, awareness that comes down to me being conscious of not what I am, but simply *that* I am. For this Kant needs to employ a strange and rarely used category of an "indeterminate perception".

Kant writes: "*In the transcendental synthesis of the manifold of representations in general, on the contrary, hence in the synthetic original unity of apperception, I am conscious* *of myself not as I appear to myself, nor as I am in myself, but only that I am. This representation is a thinking, not an intuiting*."[[86]](#footnote-87) As I just mentioned, in order to make sense of the awareness of the self that I have in the synthetic original unity, without appealing to the inner sense and intuitions in the inner sense, Kant uses the notion of the indeterminate consciousness, a kind of perception we have about ourselves in the transcendental "I think": "*An indeterminate perception here signifies only something real, which was given, and indeed only to thinking in general, thus not as appearance, and also not as thing in itself (a noumenon), but rather as something that in fact exists and is indicated as an existing thing in the proposition 'I think'*"[[87]](#footnote-88). Therefore, from the transcendental "I think" I can infer to "I exist" without appealing to an intuition of this "I", but rather to experience in general. This "I exist" is not an assertion about the state of affairs in the world that is "filtered" through categories and should be somehow supplied by the perception of the "I" (as would be necessary in a "determinate" perception of the thing), but just a general conclusion about the "I" that is inferred from the fact of possibility of our experience in general. This conclusion is reached only by thinking, and yet it still does not signify something noumenal.

This concept seems strange in the context of the *Critique of Pure Reason* and is kind of sticking out of the cognitive machinery Kant uses to explain our experience. Baldacchino turns our attention to this concept because this is a key point in the *Critique* where the temporality of the transcendental "I" is apparently supported by Kant. It is a sort of "backdoor" temporality, certainly not as full-blooded as the one suggested by Strawson and Cassam, but maybe just good enough. Namely, if it could be said that the assertion "I exist" is implied just by the very application of the transcendental "I" to our sensations, then there is a connection between the transcendental self and an actually existing person after all.[[88]](#footnote-89) "I" has to be conceived as "enduring in time", since it is safe to say it actually exists at every moment of time I am conscious of. This consciousness of the transcendental "I", although indeterminate, is still not any less real - as Baldacchino suggests, "*the expression 'indeterminate consciousness' is to be understood epistemologically, not ontologically*."[[89]](#footnote-90)

There is also a more general objection that can be used against Strawson's and Cassam's criticism of Kant, and it is expressed by Beatrice Longuenesse.[[90]](#footnote-91) To the objection of Kant "not being consequent", she implies that Kant is actually more consequent than them, and goes a step further with his conclusion, compared to, particularly, Cassam. It is a very common-sensical objection: Cassam claims that it is our body that ensures a point of reference for all our experience, but Kant's theory goes a step beyond that and tries to find out an even more primitive "point of reference", one that would be responsible for having a representation of our body at all. The objective, spatio-temporal conditions of our experience as laid out by Cassam need a further point of reference that would relate them to a numerically identical subject, which is in this case the subject's body and which is always in the proximity of the experienced objects. But to even conceive of the body as the focal point of experience, or of spatio-temporal conditions and relations at all, a "step back" is necessary: "*For Cassam, it is an empirical subject whose experiential route through the world determines spatiotemporal enabling conditions of perception. For Kant it is the agent, whatever that agent might be, of the act of combining and comparing representations by way of which spatiotemporal enabling conditions for recognizing independently existing objects become available for cognition in the first place*."[[91]](#footnote-92) Therefore, transcendental "I" should not be considered as a mere "abstraction" from the self as a living, spatio-temporal being. Invoking concepts such as experiential route of the subject's body as well as spatio-temporal relations already assumes the cognitive machinery of the *Critique* - to have such concepts we already need a thinking subject that creates them and recognizes them at the same time.

Now we move to the last problem of the section - I will briefly present one or two points about the potential need for an empirical self when dealing with the issues of practical philosophy. The passage from the *Third paralogism* concerned with this problem that gives troubles to the interpreters of Kant is this one: "*Meanwhile, the concept of personality, just like the concepts of substance and the simple, can remain (insofar as it is merely transcendental, i.e., a unity of the subject which is otherwise unknown to us, but in whose determinations there is a thoroughgoing connection of apperception), and to this extent this concept is also necessary and sufficient for practical use."[[92]](#footnote-93)* The concept of personality mentioned here, is of course the one from the *Third Paralogism,* the one in which it is said that the identity of the person is basically due to the same-sounding "I" that accompanies all my representations in all times I am conscious of, but of which not much more than that could be said.

Longuenesse thinks that this is far from "sufficient" for practical use, and supports her view with excerpts from *Metaphysics of Morals*, from which she concludes that "*it is with* *respect to persons so understood that the degree of praiseworthiness or blameworthiness of an action is evaluated, depending on the force of the drives that had to be overcome to engage into a morally legitimate action*."[[93]](#footnote-94) The line of reasoning goes like this: if one is to be considered a morally and legally imputable person, it is actually presupposed that it is a living, empirical being, since the value of an action is judged upon how difficult it was to overcome one's natural instincts and impulses in the process. This seems like a very good point. And if this is the case, then the model Kant proposed in A365 is not sufficient since it does not include any mention of the natural, empirical aspect of personality.

Corey Dyck has a different view of the problem. For him, the "practical use" is not as much about moral, legal and forensic issues, but about the problem of immortality of the soul. This brings about a drastic change of focus. The question here is not primarily whether a person can be found accountable for her actions upon the concept of person presented in the *Paralogisms*, but whether or not this concept can be consistent with the immortality of the soul. It is a question of whether, after death, the soul "*will remain conscious of its numerical identity*"[[94]](#footnote-95), and whether it will "*retain its distinctive capacities"[[95]](#footnote-96)*, in a word, whether it will remain the same person. This is the worry of rational psychology, but, as Dyck shows again by pointing out to certain passages of *Lectures on Metaphysics*, it is also something Kant was worried about himself in the earlier stages of his philosophy. So how does the concept of the self from the *Paralogisms* help us in this respect?

It does that by assuring us that there is no bodily influence, in fact not just bodily influence, but 'nothing manifold' in Kant's concept of the transcendental "I". This banishes the biggest worry of them all - if the soul is somehow partly dependent on the body, then what aspects of the personality (if any) survive bodily death? On the other hand, if the "I" is completely independent of the body, then *if* soul survives the death of the body, it will survive it undamaged and unchanged. Of course, immortality of the soul is of great importance in Kant's practical philosophy because it is a postulate of practical reason required to be believed in, as a condition for potential reward or punishment for our behaviour that is not necessarily achieved in this life. For Kant, a "sufficient" concept thus simply means that it is not in collision with the concept of soul surviving the death of the body and preserving its distinctive features, and to have this it is only important that the "I" is not composed of or dependent of anything corporeal. Dyck's idea is in accord with what Brook writes about "the second project" - complete insulation of the "I" from the realm of natural science enables practical philosophy to retain its authority in the questions of morality.

There is one more aspect of Kant's theory regarding the relation of the "I" and the body that is worth considering, and it is the problem of relation of the body and the soul. However, it seems to be more convenient to analyze it in the next chapter since it gives a helpful insight when one considers the problem of possible incorporation of the dissociative identity disorder into the framework of Kant's theory. Therefore, I leave it for chapter 2 where I will briefly deal with this problem as well.

# 2. Transcendental Apperception and Dissociative Identity Disorder

## 2.1. Features of DID

In the next two chapters I will use one concrete example in order to attempt to briefly show how Kant's theory of the self can be revitalized with a few important modifications. As it was said before, I will use the example of dissociative identity disorder, and in this chapter I will show that, despite the apparent paradox, the phenomenon of DID fits into Kant's theory without major problems. In the third chapter it will also turn out that a few small Dennettian-style modifications of the theory can offer a much more productive solution for such a task, although even with these modifications we can consider the solution to be Kantian to a large extent.

I will start with a few paragraphs about the disorder itself. DID is, in layman's terms, a phenomenon of having more persons, or more selves in one body. This disorder was formerly known as multiple personality disorder, and as such was defined in the Diagnostic and Statistic Manual of Mental Disorders (DSM) III: "*The essential feature is the existence within the individual of two or more distinct personalities, each of which is dominant at a particular time. Each personality is a fully integrated and complex unit with unique memories, behavior patterns, and social relationships that determine the nature of the individual's acts when that personality is dominant.*"[[96]](#footnote-97) In DSM-IV it was renamed to dissociative identity disorder: "*There is an inability to recall important personal information, the extent of which is too great to be explained by ordinary forgetfulness [...] Each personality state may be experienced as if it has a distinct personal history, self-image and identity, including a separate name [...] The number of identities reported range from 2 to more than a 100*."[[97]](#footnote-98)

The amnesia mentioned is often asymmetrical. This means that the more passive persons (among which is usually the "original" person, the one that was the only one there before the dissociation took place) usually do not have memories of the actions of more hostile and "protective" identities, and the more hostile identities do have memories about the doings of the passive persons - although this does not have to be the case. [[98]](#footnote-99) Usually at least one or more identities are not even aware of some or any of the other existing - they do feel that "something is wrong", they are aware of their amnesia and of other people's reports of them behaving strangely, but very often they do not have the idea that this is due to having other persons taking control. There are all kinds of variations - those in which identities share memories, those in which they do not, those in which they know of each other or do not, those who are symmetrically and asymmetrically amnestic, those who communicate with each other and so on and so on. The disorder has a vast number of different types of manifestations.

What "qualifies" this different "patterns of behaviour" to be called "persons" is their shocking consistency and a sense of self that transcends the "imaginary playmates" and amounts to actual different persons inside a body. Whenever amnesia occurs it seems genuine and also does not appear to allow "leakage" of even implicit memory.[[99]](#footnote-100) However, DID is not all about lack of memory or awareness of the other identities - "only" 72 percent of the cases had one identity who denied the existence of all the others, and 86 percent of the cases had at least one identity that reported to be aware of *all* the others.[[100]](#footnote-101) It is more about completely different (even contrasted) personality traits that are exhibited when different persons "take over" - they have different interests, temperaments, skills, they are of different age or sex. We cannot take all these identities to be a one and united person, because if we do that we cannot rationalize this person's behaviour. "*The idea that there is a unity in the conscious mental life of a person allows us to ascribe to that person certain beliefs, attitudes and traits that enter into the kinds of explanation we offer for behaviour*".[[101]](#footnote-102) Therefore, if we just define the person *ad hoc* as being necessarily tied to the idea of a body, then we have to refuse that these are "real" persons. But if we try to rationalize their behaviour, if we seek for consistency patterns and explanations of person's thoughts and actions, it is much more convenient to think of these identities as distinct persons.

However, there are still those who do not think that DID should be considered as a distinct mental disorder, nor that different identities should be regarded differently than being parts of just one person. Indeed, there is a big overlap with other personality disorders - people who suffer from DID exhibit all kinds of features that are characteristic of other disorders that do not include interpreting them as having multiple persons inside one body. The usual suspect is the borderline personality disorder and therapists dealing with DID patients are sometimes accused of actually convincing their patients that there are different selves inside them, whereas they are single persons who simply exhibit bizarrely inconsistent behaviour. Proponents of this "single person thesis"[[102]](#footnote-103) thus think that cause of the disorder is merely *iatrogenic*, that the narrative of "different persons" is triggered in the patients by the therapist and that it is not "real". Historical facts about the disorder seem to support the claim: for instance, in an article written in 1972 Horton and Miller comment on how rare this disorder is: they report that there were only a 100 cases reported in the "past century", with only a dozen in "the last 50 years".[[103]](#footnote-104) The disorder was not unheard of in the past - there are cases recorded in the 19th century, and double personality first became a medical concept in France in 1875.[[104]](#footnote-105) But it was very uncommon, so even in the DSM-II, published in 1968, there is no mention of the DID. What is strange is that in the years to come after Horton's and Miller's article was published there was an enormous rise of the number of reported cases that is on the rise still to this day. For instance, Humphrey and Dennett write that in 1989 there were already 4000 known cases, with only 1000 before 1984.[[105]](#footnote-106) Also, a vast majority of the cases that are recorded are in North America, with the disorder still being very uncommon in other parts of the world.

These data do suggest that a number of cases have in fact been triggered by the therapists. But this still does not mean that mean that it is *not real* - whether it was triggered by a "real" cause or by a therapist, it still does not mean that the "reality" of multiple persons in a body is not a useful way to explain what in fact is happening with people who suffer from the disorder. It might be that the difference between the proponents of the "single person thesis" and "multiple persons thesis" is just a matter of terminology, but there are at least two spheres in which this question plays an important role. The first one is legal[[106]](#footnote-107), and the second one is the sphere of the philosophy of mind.

When it comes to philosophy of mind, it is clear that we cannot have an absolute decision on the matter and that the thesis we choose to support depends on our initial premises and definitions of what we are to consider a person or a self. If we stick to Gillett's concept, quoted earlier in this chapter, I believe taking the view of "multiple persons thesis" is the only way. Additionally, the question of etiology can also help us make a decision. If there was a way to find that the phenomenon of multiple persons was somehow paralleled with the workings of the brain and that there are different neural patterns that could in theory "fit" the narrative about different persons, this might be a good additional argument for the multiple persons thesis.

And as for the etiology, apart from being caused iatrogenically, DID is caused by extensive and intense childhood trauma - Humphrey and Dennett report that of all cases "*perhaps* *95 percent - have been sexually or physically abused as children*".[[107]](#footnote-108) Whether the "real" cause is this or the influence of therapists or a combination of both is a question that is hard to answer, but the correlation between child abuse and DID is very high and child abuse is generally taken to be the most important and most common cause of the disorder. Amnesia and split personalities are an automatic and unconscious mechanisms of protection from bad memories and bad experiences that became too much for the "original identity" to deal with. In traumatic situations the original person simply "switches off" unconsciously and lets another identity take all the memories and stress of traumatic situations. If the abuse lasts for long periods of time, these identities develop and acquire the organization and consistency of a "real" person. In different kinds of abuse, there is a separate personality for every type of stressful set of circumstances, who switch on and off at appropriate moments. In other words, appropriate parts of the mental life are handed over to another "self" because "*the emotional cost of access to certain psychological states and attitudes is too high*".[[108]](#footnote-109)

This would be a good basis for a more extensive evolutionary and physical explanation of the disorder. At the surface, it is clear that this is an "evolutionary" mechanism - our consciousness prevents itself from accessing parts of our mental life that would actually harm us. This is a good starting point, although still superficial. But Kelly Forrest, for instance, has a further and more elaborate explanation: splitting personality is originally a matter of survival and it starts because there is an internal confilct in early childhood between "*surviving the abuse and maintaining the attachment with the parent*".[[109]](#footnote-110) The first causes of it, according to this approach, happen very early, when dealing with the negative feedback and abuse by a parent leads to an inevitable conflict of experiencing abuse on one hand and the need for attachment on the other, which then leads to a child exhibiting a very sharp difference in reactions depending on the context. The child then has a good chance of developing this state into a more general dissociation of the personality, in which its whole personality becomes exclusively context-oriented, with no long-term integration whatsoever. What is important about her approach is that she considers that these phenomena influence the neural networks in the brain, and suggests they lead to a lateral inhibition that, simply speaking, temporarily or permanently isolates certain neural networks from one another, and these networks are then activated depending on the context. "*The development of amnesia barriers in persons with DID results from the OFC [orbitofrontal cortex] isolating subsets of Me concepts and in turn results in multiple, conflicting, context-dependent Me conceptual systems [...] the aberrant development of the lateral inhibition between complex neural networks subserving higher order organizational functions fragments the behavioral structures*".[[110]](#footnote-111)

In any event, it could be said that the multiple persons approach could, in theory, be argued for by appealing to "real", physical underlying structures that could support explaining DID in the multiple-person fashion. If this turns out to be plausible enough, we have very little reason to think of DID as an "illusion" of having different persons in one body. Anyhow, the multiple person approach is the mainstream approach right now and it seems more convenient, practical and reasonable to think of these different identities as actual different selves.[[111]](#footnote-112)

## 2.2. DID and Transcendental Apperception

How is all this related to Kant and how is this seemingly contradictory to his theory? A quote from the *Deduction* might tell the story: "*The I think must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all [...] thus all manifold of intuition has a necessary relation to the I think in the same subject in which this manifold is to be encountered.*"[[112]](#footnote-113) So the apparent contradiction lies in the *same* subject of the "I" that has to be able to accompany all my representations. How does the *same* subject correspond to the multiple subjects of a person with DID?

As it was mentioned, the solution is easy and almost trivial. Every person in the body would have a different "I", and it could easily be said that there are multiple different "Is" accompanying representations of different selves of someone who suffers from DID. Transcendental apperception is the "I" that organizes and orders experiences in an unconscious act of synthesis, not the "I" that was cognized and experienced. Having in mind the ambiguity of the latter "I", if I went with the single person thesis and necessarily related the "I" to my body, I could say that this "I" of mine is related to different subjects, or different selves that are contained in my body. Then this would be contradictory to Kant's theory. But what Kant's theory implies is the direct opposite of that, although, of course, he never explicitly addresses the problem of multiple "Is", subjects or selves in one body.

Let's imagine being one of the selves in the multiple-person body. Then the "I that accompanies all my representations" could not be the I of some other self in the body, since those are not my representations, but his or hers. On the multiple persons thesis, his or hers memories are not the same as mine and vice versa, my "I" and the "I" of any other self inside the same body are not any more correlated than any other two "Is" in the world - although we use the same transcendental rules to connect sensations into a coherent experience, we have *different* experiences, different memories, different representations, accompanied by a different "I". Just like selves that occupy different bodies. It makes no difference whether we go with the substance interpretation or the personal identity interpretation, whether my representations really *are* representations in time, or appear *as if* they are such, while in fact they are all merely representations of me at this moment of time. For in both cases I do not feel, organize or cognize the representations of another self in the body as my own, I have "*my time*" and the other self has "*his time*", "*so the identity that is necessarily combined with my consciousness is not therefore combined with his consciousness"[[113]](#footnote-114)*, just like in the case of any two other selves, uncorrelated via the same body. Whether my memory is "objectively mine" (substance interpretation) or merely "subjectively mine" (personal identity interpretation) makes no difference, because the memory and representations acquired by a different self in the same body are neither objectively nor subjectively mine.

Therefore, I claim that Kant's theory of apperception is perfectly compatible with the phenomenon of DID. There are two problems that have to be separately addressed to make this even more clear - the problem of co-consciousness and the problem of the relation between the body and the soul. I believe none of these pose a serious threat to this claim.

To deal with the first problem, it is necessary to briefly consider Kant's account of other minds. This is a more general problem in Kant's philosophy that goes beyond his psychology. How do I even know other minds exist? There are several ways, the most prominent being inferring to their existence by the fact that other human beings *empirically* exhibit features of rational beings, beings that are self-conscious. From that, I infer that other human beings also possess minds.[[114]](#footnote-115) "*There are certain patterns of behaviour - 'signs' in the physical world, that we may take as evidence for a thinking being*."[[115]](#footnote-116) This can be seen in the willful acts of a human being, and entails the autonomy of the will - that it is possible for the person to act differently from the way it had actually acted, and whether this is the case is again to be decided empirically. It means that a person can be driven differently than just by laws of nature or mere natural impulses and instincts, which is, according to Kant, not the case with inanimate objects or animals. Obviously, none of these are stable criteria, but can only give us a hint whether a thing is rational or not, whether a mind can be ascribed to it or not.

Furthermore, an imperative of Kant's practical philosophy that we should always treat other people, other rational beings, as ends and never as means[[116]](#footnote-117) clearly suggest that it is possible to talk about other minds, whose existence is, also, one of the pillars of Kant's concept of objectivity. Now all these other minds use the same cognitive machinery as myself - I basically just extrapolate what I know about myself as a rational being to other rational beings.[[117]](#footnote-118) This is *all* I know and am ever able to know about other minds.

What implications does this have to our two problems? Let's first consider co-consciousness. Co-consciousness is the name for a phenomenon in which there are multiple selves in a body that are conscious of a certain event, the same event, although only one of the selves is the agent. So to say, other selves can "see" what is happening, but do not act. So if there are multiple "Is" in the same body that share the same experience, does that mean that there is after all a point of contact of the experiences of these selves?

It does mean exactly that, but this is still not contradictory to the claim that these are different "Is". Selves that occupy different bodies also experience the same events and there is a point of contact between their experiences as well. Seeing a car crash by two different selves from two different bodies is nothing different from seeing a car crash by two different selves in the same body. Those can be seen as *numerically* different experiences as much as different experiences of any two subjects. What Stevenson calls "the individuation of the experience" holds as much for selves in the same body as for any other: "*given mutual communication, two people can know that they are enjoying experiences with the same content, but we can still distinguish A's experiences from B's.*"[[118]](#footnote-119) There is absolutely no reason why this should not hold for the multiples, at least seen from the perspective of Kant's theory. And to show this, we must analyze Kant's concept of relation between the body and the soul. Is it allowedby Kant's theory that multiple souls, or multiple "Is" occupy one body?

As for the "soul", if it signifies the immaterial substance that underlies my consciousness, as we already analyzed, I can know nothing about it, neither about its characteristics nor its relation to the body. As for the "I" of others, as was mentioned in the analysis of other minds, I can know something about it only indirectly. I can see empirical "signs" that there is a rational being occupying a body, and I can infer that this rational being can employ the same rules of ordering sensations and experiences as I do. Whether there could be more "Is" occupying one body or only one is not a problem that was investigated in Kant's philosophy. In practical philosophy this would of course be an important issue, because of the problems of imputability and responsibility, but since the dilemma about multiple selves is not considered a dilemma at all in Kant's time, Kant never addressed it.

There are multiple ways in which we can interpret Kant's silence on this issue, but the most probable one is that Kant, if he had actually been presented with the problem, would have considered it somewhat absurd. However, that does not mean he would reject it as a problem - it might be that he would have considered this problem as unsolvable. In any event, we can only analyze what is written in Kant's works and what is implied by what is written, not what might have been written had some other questions been asked as well. And as for the relation of the self (or soul) and the body, very little has been said in Kant's works and everything that has been said points to the impossibility of any knowledge about the issue. Regarding the problem, he even remarks that "*concerning how a community of substances is possible at all [...] this without any doubt lies outside the field of all human cognition*".[[119]](#footnote-120) Therefore, even saying that "I" or a soul occupies a body in any way is very problematic in Kant's framework. The only point in the *Critique* where he seems to be saying something about this relation is A213=B260, although even that seems to be said without an intention to explain such a relation, and such an implication is apparently just a result of somewhat careless language used: "*The light which plays between our eyes and the celestial bodies produces a mediate community between us and them, and thereby shows us that they coexist. We cannot empirically change our position and perceive the change unless matter in all parts of space makes perception of our position possible to us*."[[120]](#footnote-121)

 Thus most we can say is that Kant considered, at least implicitly, that the temporal order in our empirical consciousness "*is related in some specifiable way to the temporal order of the positions of a particular body*"[[121]](#footnote-122). All we know, empirically, is that there is *some* relation between ourselves and our bodies. But since we know nothing about this relation, there is an open possibility that it is such that there can be multiple selves in a body. We could never confirm that with certainty, but we could never refute that as well.

To conclude, I believe this is enough evidence that the phenomenon of DID fits into Kant's framework and is allowed by his theory. However, there are a few reasons for which I believe that at this level this result is not particularly important. The first one is that from the Kantian perspective we can deal only with the unification of memories, perceptions and knowledge, and these are just a part of the equation concerning DID that also includes skills, attitudes, beliefs, temperament and a number of other personality traits that cannot be accounted for by Kant's theory of the self.

The second reason is that concluding that it merely *fits into* the framework does not necessarily mean that this framework offers a fruitful or constructive solution and a good starting point for *explaining* or gaining better insight into the phenomenon. Especially given that Kant's theory of transcendental apperception, as well as his philosophy, is in this sense quite flexible and refuses to both confirm and deny a lot of things, but claims ignorance and demands silence on such issues. Nothing could disprove the claim that there are multiple selves in a single body, but nothing could prove it beyond doubt as well. The explanatory force of the theory that claims absolutely nothing on the matter but "*it could be this way, but it could be otherwise as well"*  is not particularly promising. My job in this chapter was simply to prove that if we fill in the holes (left intentionally open) in Kant's philosophy the way we want, we can show that the multiple-persons approach to DID is possible inside it. But we cannot say anything further about this phenomenon if we stay inside the theory the way it was originally conceived, let alone gain a deeper insight into DID.

Nevertheless, I think that Kant's general framework and a number of his arguments, if disentangled from some of their metaphysical burden and put to work in another context can be very useful for this purpose. In the last chapter of the paper, I will be looking into Dennett's explanation of the DID, and I will attempt to show that some of the valuable lessons we can learn from Dennett's theory are very much Kantian, pointing out to the fruitfulness of the appropriate reassessment of Kant's ideas in general.

# 3. Self as a Centre of Narrative Gravity

## 3.1. Dennett's Theory

I will open this chapter with a part of the conversation between a DID patient and a therapist, as reported by Hardcastle and Flanagan. At a certain point of the conversation the patient says how she always thought her disorder is what sets her apart from other people, but has just realized that it is not that simple: *"[I think] That I'm not different at all - just honest about it. That they're not continuous, either. That everyone is organized the same way I am, with the compartments and differences, all that weirdness, but they can't admit it! That they have this fiction that they're both continuous and immutable. Because that's all it is - a fiction. What would have Ibsen called it? A saving lie - that's it! People are not the same today, tomorrow, any day, ever!*"[[122]](#footnote-123) This observation made by a DID patient is strikingly similar to some of the things Dennett claim in his theory of the unity of the self. The central point - that the unified self is due to some kind of fiction, that it is a mere abstraction, remains the same, and Dennett builds his whole theory upon this key assumption.

First of all, Dennet claims that self cannot be due to some kind of bodily continuity nor it can be in the brain. He establishes this with the help of a thought experiment in the essay titled "Where am I?"[[123]](#footnote-124). In this thought experiment Dennett wonders whether our self is something that is "attached" in one way or another to the body or the brain. Then he proceeds to show how we would react in an imaginary situation in which a person committed a crime. Would we punish the brain or the body? If we punish the body by sending it to prison, we can imagine the brain being transplanted to another body, thus "leaving" unpunished. On the other hand, if we punish the brain, leave it in the cell, but still connect it to a remote body that would move freely, that would not do the justice either. The experiment goes even further. Imagine I wander around with the body connected by some kind of remote connections to the brain, the brain itself being in the vat. I do not feel as if I am in the vat but that I am "behind my eyes", where the body is. However, as soon as the connection is interrupted, as soon as I lose contact with the body, my self is back to the vat, in an instant, "faster than the speed of light". What does Dennett conclude from this?

He concludes that the self is neither necessarily attached to the body nor is it somewhere in the brain. It is an abstraction whose "location" is determined by a complicated set of circumstances. The further conclusions are drawn in the "Self as a Center of Narrative Gravity".[[124]](#footnote-125) Dennett claims that the self is "somewhere" as much as the abstraction of the center of gravity is "somewhere". We *imagine* it is in a certain place, for instance in our head, but just as we will not find a small lump of matter in the physical space for which we could say '*this* is the centre of gravity' - for that would be a categorical mistake - we will not find a structure in the brain to which we can point and say 'this is the self'. Centre of gravity, as well as the self, are just useful abstractions, fictions, that we make use of in physics, psychology, philosophy or everyday life and we should not confuse them with what might seem to be their physical counterpart.

The analogy of self and the center of gravity is important in another respect, because for Dennett self is also a kind of an "attractor". What "circles" around this attractor are our narratives. This attractor, as well as the center of gravity, changes through time together with the changes of circumstances we adapt to. What we *know* about ourselves is not really about knowledge, it is about interpreting what we do and what we say, creating a web of narratives, a *story* in which all our actions and thoughts fit together. We constantly create these stories to make sense of and rationalize our behaviour by a reference to a single "thing" that connects our experiences, memories, actions, beliefs, attitudes. It is not about cognizing this self as an objective thing, but about *creating* it as an abstraction that supports the consistency of all these different narratives, and also changing narratives in order to fit the idea of a single self. Dennett writes: "*it does seem that we are all virtuoso novelists, who find ourselves engaged in all sorts of behavior, more or less unified, but sometimes disunified, and we always put the best "faces" on it we can. We try to make all of our material cohere into a single good story. And that story is our autobiography. The chief fictional character at the center of that autobiography is one's self*."[[125]](#footnote-126) Dennett conceives of this as an important evolutionary strategy, and I will turn to that very soon. But before that let's see how DID fits into this framework.

For Dennett, the phenomenon of DID is explained by the fact that there are different narratives told and conceived by a DID person (or, in fact, persons). In this respect, this person is not a single person, but has different selves - as many of them as there are centers of narrative gravity. And it should not be at all uncommon that this happens, because even to people with one stable center of narrative gravity this happens to some extent. Each one of us is what Flanagan and Hardcastle call the "multiplex" - we have different *roles* in our life and different sets of rules, desires, projects, that sometimes contradict each other - we can be children, parents, students, bosses, employees, citizens and so on, and each one of these roles demands from us a different pattern of behaviour we stick to. This sheds a new light to the ideas of a DID patient quoted at the beginning of this chapter. We are all confabulators, our selves are also *imaginary*, but there is still an important difference. What distinguishes a multiplex person from someone with a multiple personality disorder is the fact that the former has "*an integrated and unified narrative that contains, as proper parts, the different selves we display in isolation*."[[126]](#footnote-127) On the other hand, there is no such a unified narrative in people with DID. The narratives they create are truly different narratives, with different self-centers that are triggered by an appropriate context, that are brought to the "surface" depending on the situation and then can further develop. This happens in stressful situations when one cannot incorporate the event into the narrative due to it being too traumatic, and particularly severe and extensive abuse can trigger a completely new narrative to come into being and the production of this narrative continues every time the same type of stressful situation happens. The new narrative (the new 'cluster of speech') develops freely and in the end evolves to the point in which it has a relatively stable and rationalized "center of gravity" and amounts to the consistency and complexity of a person. This is what Humphrey and Dennett call "*the division of emotional labor.*"[[127]](#footnote-128)

Another question that needs to be answered is where these narratives come from, and *who* creates them. Selves are not brains or particular parts of the brain, but they are *created* by our brains. Dennett compares the creation of narratives with a spider web.[[128]](#footnote-129) Just like a spider uses different materials to produce a web, our brains use words and produce complete narratives of ourselves. This does not happen consciously, as well as spider does not "consciously" weave its web. Dennett would say that "*we're just wired that way".* *I* am not in control of creating these narratives, because *I* is just a fiction, just a lense through which we look back at everything we experience, do, say and think. This self "*structures and influences first-person experiences. It can recast our representational and cognitive structures as well as influence our motivations and behavior*."[[129]](#footnote-130) It is just an abstraction that pre-conditions our experience and is not in fact created consciously, by our *selves*, but by our brains that do this job automatically. They do not do this job completely by themselves but in combination with "*my parents and siblings and friends*"[[130]](#footnote-131), or in other words, with the environment. The brain takes the language as "the material" and creates a web of narratives with the ultimate goal to create a stable center of our experiences, influenced in this job by environment and social pressure.[[131]](#footnote-132) The most a self can be is a "spokesperson" of this whole machinery, while the real cause, the "instructions" for forming a self have to come from somewhere else - from the brain. *We* are all confabulators and novelists, but we just write down what our brains dictate. It is also worth mentioning that this does not mean that there is a "central controller" in our brain, but it is just a matter of its organization, of the way it is structured - we are simply 'wired' that way.

Now what would be the evolutionary reason for this complex enterprise? Dennett draws an analogy between the psychological self and the biological self. A biological self is the "*minimal proclivity to distinguish self from other in order to protect oneself*."[[132]](#footnote-133) These are basically the boundaries of a creature's body, but even these boundaries are sometimes indefinite and vague. However, the human environment is more than just our physical/biological environment - it is also linguistic. Thus distinguishing a self from the other in the human world means creating a narrative that does so, a narrative whose centre is the self as opposed to everything else that is not the self, while bodily boundaries are less important.

So how does the division in DID happen? Basically, in order to protect my psychological self from "bad memories", from experiences that could seriously harm me, my brain "draws" a new boundary between my self and the other, and designates the person to whom these things are happening as an "other self". What is not in relation to my self is not in relation to my world as well - the bad memories and perceptions are not mine anymore and thus *they do not exist*. The bounds change in order to protect us and isolate these experiences, and they change by the shift of focus of the self-narrative - actually, by creating an additional focus. Of course, this is not supposed to happen instantly, in a blink of an eye, but it represents a process that happens gradually. Further fragmentation can then be explained using the same line of reasoning, with the already divided persons being more vulnerable and liable to further separation.

## 3.2. Dennett and Kant

That is, in short, Dennett's account of the DID. I believe that there are several aspects of the theory in which we can recognize the Kantian legacy. First of all, in both theories, the self is the reference point that connects and makes sense of our experience. The self is something that "happens" prior to the conscious experience, it 'recasts our representational and cognitive structures' and forms our view of the world. Furthermore, this happens in an *unconscious* act, it is not something we plan or want. Our consciousness is not the source of the rules that control this act: for Dennett the source is the brain, while for Kant the source are the *a priori* categories that underlie the consciousness of every human being. But the point is this: for both of them, self (or transcendental apperception) is unconscious and abstract and it reveals itself in the *act* of ordering and making sense of our experiences. The experience is never without it, it is a prerequisite for our awareness of the world and for introspection. As well as in Kant, in Dennett we do not infer to a possibility or reality of a united self on the basis of introspection (or, in Kant, empirical apperception), but we connect, interpret and order our mental states on the basis of the already existing center of our narratives (transcendental apperception). Also, Kant's use of the notion of imagination is indicative in this respect (although it is not what we usually mean by "imagination" - a mere subjective game of associations). For Kant, point of transcendental imagination is connecting things that are not *originally* connected, reading something into the world that is not already there.[[133]](#footnote-134) When it comes to DID, both theories, by claiming that the self is an abstraction by means of which we interpret our experiences (or parts of it) enable a tidy multiple-person solution of the problem.

Claims about the ontology of the self are also strikingly similar. Dennett writes: "*Do we exist? Of course! [...] Are there entities, either in our brains, or over and above our brains, that control our bodies, think our thoughts, make our decisions? Of course not!*"[[134]](#footnote-135) As in Kant, we do exist, but that is all we can say "objectively" about the self. The self is not a thing or an entity, yet it does exist and it is the lense through which our experience is, unconsciously, conceptualized and rationalized. I take this to be the most important common point of these two theories.

However, the full explanatory strength of Dennett's theory can be showed only when we look at the differences and novelties he introduced, compared to Kant. First of all, Dennett does not put the emphasis on the self as the basis of objectivity, but as the basis of our experience of *ourselves*, of our mental states, actions, beliefs and memories. And this is the indirect source of what we might call an advantage of his theory - it accounts for all mental states and features (even for those in which Kant is not particularly interested), not just objective knowledge and memories. For Dennett, the thing is not so much about explaining the self as the source of objectivity, but as a trick that inverts the common-sense notion of self-consciousness, the very same way Kant inverted it by introducing the notions of transcendental and empirical apperception. In Kant, as it has just been said, the theory is focused on explaining objective knowledge, and although he does occasionally mention other affections of ours (like emotions, pains or moods), they are not important for his theory, since they are private and lack objectivity. His account could, in theory, deal with these as well, but he never develops it in this direction since it is an irrelevant direction for his theoretical philosophy. This might be seen as one of the problems of Kant's theory in this context - it is too theoretical, the spontaneous act of synthesis is too cognition-oriented.[[135]](#footnote-136)

Nevertheless, the biggest problem of Kant's theory of the unity of self is that it "hits the wall" too soon. As it was said before, transcendental apperception is something we could generally assume in adults, and it is a very useful epistemological idea. The problem is that the theory does not allow us to inspect its roots properly. Dennett's idea, however similar to Kant's in the aforementioned sense, goes a step further with the analysis, because it does not have the metaphysical burden - or, rather, transcendental burden, that Kant's theory has. Here is what I have in mind.

Kant's idea of the cognitive self and Dennett's idea of the narrative self are both automatic, spontaneous and have an unconscious source. Dennett's explanation of this unconscious source is that we are just "*wired that way*"[[136]](#footnote-137) while Kant's explanation is somewhat similar - those are the primitive a priori rules of our experience. Now what makes all the difference is that the fact that 'we are wired that way' is a starting point for further research, while 'a priori rules of experience', at least as Kant conceived them, are an ending point of the research. The fact that they are considered universal and necessary, the fact that it is not allowed (and according to Kant, it would make no sense) to inspect their genesis, their origin, cause or source is what I mean by a 'transcendental burden'. Kant does not allow this only because he takes the a priori rules as *given*, as primitive facts of the theory and of human experience, since that is how adults in this era and this culture function. From our perspective, with the experience of the theory of evolution and with the subsequent development of anthropology, it is easy to say that such a view is careless. But in Kant's time conceiving of these rules as transcendental probably made more sense. In any event, this is where the theory's potential stops and it cannot move anywhere from there.

On the other hand, we can let go of this metaphysical burden, which is in Dennett done by switching from the cognitive self through which the primitive a priori rules of cognition are introduced, to the narrative self that is a linguistic construct and that is left open to further analysis. This makes all the difference, and this is easily seen in the insightfulness of his analysis (as opposed to Kant's) when it comes to the phenomenon of DID. While DID merely fits into a theory with Kantian transcendental burden, a theory, like Dennett's, that gets rid of the burden, yet accept some of the important conclusions of Kant's philosophy has a lot of potential to offer insightful explanations of the phenomenon. This is not just the case with explaining DID, it is easily conceivable that this is also the road to be taken when more general issues of epistemology and psychology are discussed.

From the point where Kant's theory stops, if the transcendental burden is discarded, there are at least two ways that can be taken if we are to conduct further research on the topic. First one is the way of cultural studies, ethnology, anthropology, comparative linguistics, comparative mythology etc, in short, a kind of research that would focus on the self as a cultural phenomenon. This has, of course, been a very fruitful field of research, especially in the 20th century. [[137]](#footnote-138) The other way would be the way of neurological research, entangled with different evolutionary explanations. Although self is not in the brain, some specific neural organization is a necessary condition for generating the self as an abstraction, and it could be that some neural networks of the DID patients share specific common traits. For example, Forrest's and Zeki's articles, to an extent, point in such a direction.

By claiming that the self as my fiction is created by "*my brain, acting in concert over the years with my parents, siblings and friends*"[[138]](#footnote-139), Dennett points to both directions. He is not that much concerned with the workings of the brain or a cultural platform upon which the narratives are built, but with the very inner workings of the "linguistic" self and the way it influences our self-awareness. Dennett's theory is more down-to-earth than Kant's, using linguistic concepts and ideas, and it has no transcendental burden, while Kant's theory is an epistemological analysis intertwined with a particular ontology that makes it much more rigid. Thus Dennett's theory has a more constructive application in the problem of DID, and it leaves much more room for further research, concerning the phenomenon of DID, the problem of unity of self as well as more general issues of the philosophy of mind. However, some of Kant's insights can be seen as the foundation stone for the theory of the self as a center of narrative gravity, and are of great importance for any theory that is dealing with the problem of unity of the self. The ideas that we are not passive receivers of sensations but active participants in the process of forming of our experience, that we are always reading something into our external and internal reality, and that my self-awareness plays the key role in the process of understanding this reality, are the lessons to be learned and built upon.

# References

American Psychiatric Association (1980), *Diagnostic and Statistic Manual of Mental Disorders III,* Washington

American Psychiatric Association (1994), *Diagnostic and Statistic Manual of Mental Disorders IV*, Washington

Baldacchino, L. (1980), "Kant's Theory of Self-Consciousness", *Kant-Studien 71:* 393-405.

Bennett, J. (1966), *Kant's Analytic*, Cambridge

Bennett, J. (1974), *Kant's Dialectic,* Cambridge

Braude, S. (2003), "Counting Persons and Living With Alters: Comments on Matthews", *Philosophy, Psychiatry and Psychology, 10* (2): 153-156.

Brook, A. (1994), *Kant and the Mind,* Cambridge

Brook, A. (2001), "Kant, Self-Awareness and Self-Reference", *Self-Reference and Self-Awareness:* 9-30.

Brook, A. (2006), "Kant: A Unified Representational Base for All Consciousness", *Self-Representational Approaches to Consciousness:* 89-109.

Cassam, Q. (1997), *Self and World*, New York

Cassirer, E. (2006), *Kant - Life and Thought,* Belgrade

Cassirer, E. (1953), *Language and Myth*, New York

Dennett, D. (1981), *Philosophical Essays on Mind and Psychology,* Cambridge

Dennett, D. (1991), *Consciousness Explained*, New York

Dennett, D. (1992), "Self as a Center of Narrative Gravity", *Self and Consciousness: Multiple Perspectives:* 275-88.

Descartes, R. (1911), *Meditations on First Philosophy,* Cambridge

Dyck, C.W. (2010), "The Aeneas Argument: Personality and Immortality in Kant's Third Paralogism", *Kant Yearbook, 2* (1): 95-122.

Forrest, K. (2001), "Toward an Etiology of Dissociative Identity Disorder: A Neurodevelopmental Approach", *Consciousness and Cognition, 10:* 259-293

Gillett, G.R. (1986), "Multiple Personality and the Concept of a Person", *New Ideas in Psychology,* *4*: 173-184.

Hacking, I. (1991), "Two Souls in One Body", *Critical Inquiry, 17:* 838-867.

Hardcastle, V. G. and Flanagan, O. (1999), "Multiplex vs. Multiple Selves: Distinguishing Dissociative Disorders", *The Monist, 82* (4): 645-657.

Horton, P. and Miller, D. (1972), "The Etiology of Multiple Personality", *Comprehensive Psychiatry*, *13* (2): 151-159.

Hume, D. (1960), *Treatise of Human Nature,* Oxford

Humphrey, N. and Dennett, D. (1989), "Speaking for Our Selves: An Assessment of Multiple Personality Disorder", *Philosophical Explorations:* 68-98.

Kangrga M. (1989), *Ethics and Revolution*, Zagreb

Kant, I. (1909), *Critique of Practical Reason,* London

Kant, I. (1997), *Groundwork for Metaphysics of Morals,* Cambridge

Kant, I. (1997), *Prolegomena to Any Future Metaphysics,* New York

Kant, I. (1998), *Critique of Pure Reason*, Cambridge

Locke J. (1979), *An Essay Concerning Human Understanding*, Oxford

Longuenesse, B. (2006), "Self-Consciousness and Consciousness of One's Own Body: Variations on a Kantian Theme", *Philosophical Topics, 34* (1/2): 283-311

Longuenesse, B. (2007), "Kant on the Identity of Persons", *Proceedings of the Aristotelian Society,* *107*: 149-167.

Longuenesse, B. (2012), "I and the Brain", *Psychological Research*, *76:* 220-228.

Marina, J. (2011), "Transcendental Arguments for Personal Identity in Kant's Transcendental Deduction", *Philo, 14* (2): 109-136.

Matthews, S. (2003) "Establishing Personal Identity in Cases of DID", *Philosophy, Psychiatry &Psychology 10* (2): 143-151.

Mendus, S. (1984), "Kant's Doctrine of the Self", *Kant-Studien, 75*: 55-64.

Parfit, D. (1971), "Personal Identity", *Philosophical Review, 80:* 3-27.

Peters, M.L. et al. (1998), "Apparent amnesia On Experimental Memory Tests in Dissociative Identity Disorder: An Explanatory Study", *Consciousness and Cognition, 7:* 27-41.

Shoemaker, S. (1968), "Self-Reference and Self-Awareness", *The Journal of Philosophy, 65:* 555-567.

Shoemaker, D. (2002), "The Irrelevance/Incoherence of Non-Reductionism About Personal Identity, *Philo, 5* (2): 143-160.

Stevenson, L. (2000) "Synthetic Unities of Experience", *Philosophy and Phenomenological Research, 60* (2): 281-305.

Strawson, P.F. (1966), *The bounds of Sense,* London

van Kirk, C. (1986) "Kant and the Problem of Other Minds", *Kant-Studien, 77* (1): 41-58

Williams, B. (1970), "The Self and the Future", *Philosophical Review, 79:* 161-180.

Zeki, S. (2003) "Disunity of Consciousness", *Trends in Cognitive Science, 7* (5): 214-218.

1. See Corey W. Dyck: "The Aeneas Argument: Personality and Immortality in Kant's Third Paralogism", *Kant Yearbook* (2010), p. 95-122 [↑](#footnote-ref-2)
2. Andrew Brook, *Kant and the Mind* (Cambridge, 1994), p. 1. [↑](#footnote-ref-3)
3. Andrew Brook, "Kant: A Unified Representational Base for All Consciousness", *Self-Representational Approaches to Consciousness* (2006), p. 90. [↑](#footnote-ref-4)
4. Grant R. Gillet, "Multiple Personality and the Concept of a Person", *New Ideas in Psychology* (1986), p.178. [↑](#footnote-ref-5)
5. Immanuel Kant, *Critique of Pure Reason*, (Cambridge, 1998), A365-6 [↑](#footnote-ref-6)
6. David Hume, *Treatise of Human Nature* (Oxford, 1960), p. 156. [↑](#footnote-ref-7)
7. Kant, *Critique of Pure Reason*, A125 [↑](#footnote-ref-8)
8. Ernst Cassirer, *Kant - Life and Thought,* (Belgrade, 2006), p. 153. [↑](#footnote-ref-9)
9. Kant, *Critique of Pure Reason,* A320=B377 [↑](#footnote-ref-10)
10. Full table in *Critique of Pure Reason,* A70=B95 [↑](#footnote-ref-11)
11. Full table of categories in the *Critique of Pure Reason,* A80=B106 [↑](#footnote-ref-12)
12. P.F. Strawson, *The bounds of Sense,* (London, 1966) 79-80. [↑](#footnote-ref-13)
13. Kant, *Critique of Pure Reason*, A93=B126 [↑](#footnote-ref-14)
14. *Ibid.,* B168-9 [↑](#footnote-ref-15)
15. That Kantian "self" is a kind of a spontaneous act is one of its central features and it will be of importance at three points in the next section: 1. when the difference is made between two kinds of apperceptions; 2. in the discussion about whether this "I" is some kind of a substance or not; 3. in the discussion whether some kind of knowledge about this transcendental "I" is possible. It will also be a significant point in the analogies that will be drawn between Kant's and Dennett's theories. [↑](#footnote-ref-16)
16. Susan Mendus, "Kant's Doctrine of the Self", *Kant-Studien* (1984)*,* 62. [↑](#footnote-ref-17)
17. Andrew Brook "Kant, Self-Awareness and Self-Reference", *Self-Reference and Self-Awareness* (2001), p.8-9 [↑](#footnote-ref-18)
18. *Ibid*., p. 6 [↑](#footnote-ref-19)
19. Kant, *Critique of Pure Reason,* A346-7. The problem of other minds demands a bit more attention when it comes to compatibility of Kant's theory with the fact of DID, and I will come back to it in the second chapter of the paper. . [↑](#footnote-ref-20)
20. *Ibid,* B69 [↑](#footnote-ref-21)
21. In the *Refutation of Idealism* Kant tries to present a somewhat strange proof as to why we should hold that there actually is the external world from which sensations are given to us, instead of believing that it all - both the material of sensations and the a priori forms that order it - comes from our mind. He tries to show that with idealism being true, we would have no way of "*grasping the successive existence of ourself in different states through outer intuition*" (B292). A cryptic proof gets even more complicated when we realize that the "ourself" Kant writes about is neither the empirical nor the transcendental self, but something like what Marina calls "*self that is the product of self-affection*" (see Jacqueline Marina, "Transcendental Arguments for Personal Identity in Kant's Transcendental Deduction", *Philo* (2011), p. 21). However, I will not pursue the problem of this proof further since it seems not to be of vital importance for this paper. [↑](#footnote-ref-22)
22. Transcendental imagination is another faculty of mind introduced by Kant as a kind of a bridge between senses and understanding. What Kant writes about it is quite different in the First and Second edition, but the point is that the imagination is the faculty that does the job of connecting previously unrelated material of the senses. It has its root and source in "*an original consciousness, in agreement with the categories*" (B161) and is essentially dependent on the faculty of understanding and the original synthetic unity of transcendental apperception. Together with the pure intuition of the time it plays the major role in the *Schematism*, doing the same job - connecting the understanding and the senses. [↑](#footnote-ref-23)
23. Immanuel Kant, *Prolegomena to Any Future Metaphysics,* (New York, 1997),paragraphs 18-20 [↑](#footnote-ref-24)
24. Kant, *Critique of Pure Reason*, B142. [↑](#footnote-ref-25)
25. Johnatan Bennett, *Kant's Analytic*, (Cambridge, 1966), p. 131. My whole reconstruction of this problem is based on the chapter *Objectivity and Transcendental Deduction* in Bennett's book. [↑](#footnote-ref-26)
26. Kant, *Critique of Pure Reason,* B142 [↑](#footnote-ref-27)
27. Bennett, *Kant's Analytic*, p. 132. [↑](#footnote-ref-28)
28. Kant, *Prolegomena,* paragraph 20. [↑](#footnote-ref-29)
29. See for instance: *Critique of Pure Reason,* A113, A108, B161, B 162, B164-5 [↑](#footnote-ref-30)
30. *Ibid.,* A108 [↑](#footnote-ref-31)
31. Compare this to Hume's concept of "cause" that he assumes can be used without implying any unnecessary and illusory metaphysical background - *Treatise of Human Nature*, p. 172-4. Kant's notion of conjunction by rules of mere association is undoubtedly akin to this. [↑](#footnote-ref-32)
32. Kant, *Critique of Pure Reason,* B165 [↑](#footnote-ref-33)
33. Brook, "Kant: A Unified Representational Base for All Consciousness", p. 90. [↑](#footnote-ref-34)
34. Kant, *Critique of Pure Reason,* A444-451, B472-9 [↑](#footnote-ref-35)
35. Immanuel Kant, *Groundwork for Metaphysics of Morals* (Cambridge, 1997), p. 41. [↑](#footnote-ref-36)
36. Kant, *Critique of Pure Reason*, A396 [↑](#footnote-ref-37)
37. *Ibid,* A344=B402 [↑](#footnote-ref-38)
38. *Ibid*, A348 [↑](#footnote-ref-39)
39. When Kant writes about "substance" here, he certainly has in mind mainly the immaterial substance, *res cogitans*, the neuralgic point of the Cartesian theory of the mind, and it is obvious from the remarks Kant makes on the "I think" when warning us that from "I think" we cannot conclude that "I am the thing that thinks", thus obviously referring to Descartes (*Meditations on First Philosophy,* Cambridge, 1911). This is the view commonly held in the interpretations. However, the fact that Kant also refers to the (im)possibility of interpreting the "I" as a material substance should not be underestimated - it is not the idea foreign to philosophy of his era, especially in the works of Holbach, La Mettrie and other French philosophers of Enlightenment. [↑](#footnote-ref-40)
40. Kant, *Critique of Pure Reason,* A 106-7 [↑](#footnote-ref-41)
41. On the other hand, Kant suggests that the fact we have no knowledge of it does not mean that we do not have any *awareness* of it. I will analyze this problem in the fourth section of this chapter. [↑](#footnote-ref-42)
42. Kant, *Critique of Pure Reason*, A361 [↑](#footnote-ref-43)
43. Beatrice Longuenesse, "Kant on the Identity of Persons", *Proceedings of the Aristotelian Society* (2007), p. 150. [↑](#footnote-ref-44)
44. Kant, *Critique of Pure Reason,* A363-4 [↑](#footnote-ref-45)
45. Jonathan Bennett, *Kant's Dialectic* (Cambridge, 1974), p. 105. [↑](#footnote-ref-46)
46. Strawson, *The bounds of sense,* p.168. [↑](#footnote-ref-47)
47. Bennett, *Kant's Dialectic,* p. 109-110. [↑](#footnote-ref-48)
48. Locke, *An Essay Concerning Human Understanding,* XXVII, 13. [↑](#footnote-ref-49)
49. Kant, *Critique of Pure Reason,* A365; Locke, *An Essay Concerning Human Understanding,* XXVII, 10. [↑](#footnote-ref-50)
50. Dyck, "The Aeneas Argument: Personality and Immortality in Kant's Third Paralogism ", p. 117. [↑](#footnote-ref-51)
51. Kant, *Critique of Pure Reason,* A364. [↑](#footnote-ref-52)
52. *Ibid,* A363 [↑](#footnote-ref-53)
53. *Ibid.* [↑](#footnote-ref-54)
54. *Ibid,* A362-3 [↑](#footnote-ref-55)
55. Parfit, "Personal Identity", *Philosophical Review* (1971) [↑](#footnote-ref-56)
56. Contra Strawson, who claims that it would make no sense to ask such a question. See Strawson, *The Bounds of Sense,* p. 165. [↑](#footnote-ref-57)
57. Brook, *Kant and the Mind,* p. 190. [↑](#footnote-ref-58)
58. *Ibid,* p. 195. [↑](#footnote-ref-59)
59. Shoemaker, "Self-Reference and Self-Awareness", *The Journal of Philosophy* (1968), p. 556. [↑](#footnote-ref-60)
60. Mendus, "Kant's Doctrine of the Self", p. 61-62. [↑](#footnote-ref-61)
61. Brook, *Kant and the Mind,* p. 197. [↑](#footnote-ref-62)
62. Brook, *Kant and the Mind,* p. 197. [↑](#footnote-ref-63)
63. Kant, *Critique of Pure Reason,* A35=B51 [↑](#footnote-ref-64)
64. *Ibid,* A35=B52 [↑](#footnote-ref-65)
65. *Ibid,* A144=B183 [↑](#footnote-ref-66)
66. It is interesting that Brook himself, as I mentioned before, holds that representation is an indivisible three-fold thing, consisting at the same time of the representation itself, a representation of object, and self-awareness. I must admit I have troubles to understand just how is this consistent with Brook's views I am presenting in this section. [↑](#footnote-ref-67)
67. Bennett, *Kant's Dialectic,* p. 111. [↑](#footnote-ref-68)
68. Kant, *Critique od Pure Reason,* A402, A404 [↑](#footnote-ref-69)
69. Brook, *Kant and the Mind, p.* 196. [↑](#footnote-ref-70)
70. Mendus, "Kant's Doctrine of the Self", p. 60 [↑](#footnote-ref-71)
71. Kant, *Critique of Pure Reason,* B422 [↑](#footnote-ref-72)
72. Parfit, "Personal Identity" [↑](#footnote-ref-73)
73. Williams, "The Self and the Future", *Philosophical Review* (1970) [↑](#footnote-ref-74)
74. Shoemaker, "The Irrelevance/Incoherence of Non-Reductionism About Personal Identity, *Philo* (2002) [↑](#footnote-ref-75)
75. Semir Zeki, "Disunity of Consciousness", *Trends in Cognitive Science* (2003), p. 214. [↑](#footnote-ref-76)
76. *Ibid* [↑](#footnote-ref-77)
77. *Ibid,* p. 215. [↑](#footnote-ref-78)
78. *Ibid,* p. 217. [↑](#footnote-ref-79)
79. *Ibid* [↑](#footnote-ref-80)
80. Longuenesse has a similar idea in mind when she suggests a way in which some kind of physical reality of free will could be shown. One of the requirements would be to distinguish the patterns of neural activities when a person is being conscious of oneself as a subject (in a Kantian sense) from other types of self-consciousness. See Longuenesse, "I and the Brain", *Psychological Research* (2012). [↑](#footnote-ref-81)
81. Quassim Cassam, *Self and World*, New York (1997); Strawson, *The Bounds of Sense* [↑](#footnote-ref-82)
82. Cassam, *Self and World,* p. 40. [↑](#footnote-ref-83)
83. Strawson, *The Bounds of Sense,* p. 249. [↑](#footnote-ref-84)
84. *Ibid*, p. 248. [↑](#footnote-ref-85)
85. Lewis Baldacchino, "Kant's Theory of Self-Consciousness", *Kant-Studien* (1980), p. 395. [↑](#footnote-ref-86)
86. Kant, *Critique of Pure Reason,* B157 [↑](#footnote-ref-87)
87. *Ibid,* B423 [↑](#footnote-ref-88)
88. This might be another way to go when we try to prove the personal identity interpretation wrong - here it appears that the transcendental "I" does have a temporal component and that Kant asserts that the existence of "I" *in time* is something that could be claimed to be objective to a certain extent. However, the indefiniteness and ambiguity of the term 'indeterminate consciousness' gives us serious trouble if we want to make sense of such an argument. It is not clear what (if anything) could reliably be said to follow from the awareness of something that we necessarily lack knowledge of. I believe that the vagueness of the phrase 'indeterminate consciousness' and lack of elaboration as to where it belongs and how exactly it fits into Kant's system prevents us from using it in arguing against the personal identity interpretation. [↑](#footnote-ref-89)
89. Baldacchino, "Kant's Theory of Self-Consciousness" 402. [↑](#footnote-ref-90)
90. Longuenesse, "Self-Consciousness and Consciousness of One's Own Body: Variations on a Kantian Theme", *Philosophical Topics* (2006), p. 283-311. [↑](#footnote-ref-91)
91. *Ibid,* p. 299. [↑](#footnote-ref-92)
92. Kant, *Critique of Pure Reason,* A365 [↑](#footnote-ref-93)
93. Longuenesse, "Kant on the Identity of Persons", p. 161. [↑](#footnote-ref-94)
94. Dyck, "The Aeneas Argument: Personality and Immortality in Kant's Third Paralogism", p. 119. [↑](#footnote-ref-95)
95. *Ibid,*  p. 120. [↑](#footnote-ref-96)
96. American Psychiatric Association, *DSM-III,* Washington(1980), p. 257. [↑](#footnote-ref-97)
97. American Psychiatric Association, *DSM-IV*, Washington (1994), p. 484-5 [↑](#footnote-ref-98)
98. *Ibid* [↑](#footnote-ref-99)
99. Peters et al., "Apparent amnesia On Experimental Memory Tests in Dissociative Identity disorder: An Explanatory Study", *Consciousness and Cognition* (1998), p. 28-41 [↑](#footnote-ref-100)
100. *Ibid,* p. 28. [↑](#footnote-ref-101)
101. Gillett, "Multiple Personality and the Concept of a Person", p. 177. [↑](#footnote-ref-102)
102. "*According to the single person thesis, individuals with DID have a disorder that has the effect of fragmenting one's existing personality. The patient is to be regarded morally and legally as a single person whose psychiatric symptoms, among other effects, cause delusions of identity.*" - Steve Matthews, "Establishing Personal Identity in Cases of DID", *Philosophy, Psychiatry &Psychology* (2003), p. 144 [↑](#footnote-ref-103)
103. Paul Horton and Derek Miller, "The Etiology of Multiple Personality", *Comprehensive Psychiatry* (1972), p. 151. [↑](#footnote-ref-104)
104. Ian Hacking, "Two Souls in One Body", *Critical Inquiry* (1991), p. 840. [↑](#footnote-ref-105)
105. Nicolas Humphrey and Daniel Dennett, "Speaking for Our Selves: An Assessment of Multiple Personality Disorder", *Philosophical Explorations* (1989), p. 69. [↑](#footnote-ref-106)
106. More on this point in Matthews, "Establishing Personal Identity in Cases of DID", and Stephen Braude, "Counting Persons and Living With Alters: Comments on Matthews", *Philosophy, Psychiatry and Psychology* (2003) [↑](#footnote-ref-107)
107. Humphrey and Dennett, "Speaking for Our Selves: An Assessment of Multiple Personality Disorder", p. 69 [↑](#footnote-ref-108)
108. Gillett, "Multiple Personality and the Concept of a Person", 182. [↑](#footnote-ref-109)
109. Forrest, "Toward an Etiology of Dissociative Identity Disorder: A Neurodevelopmental Approach", *Consciousness and Cognition* (2001), p. 280. [↑](#footnote-ref-110)
110. *Ibid*, p. 281. [↑](#footnote-ref-111)
111. As for the still-puzzling dependence of DID of the historical and cultural context, Humphrey and Dennett offer an unconventional answer to the puzzle - there might have been much more cases of DID throughout history and around the world which were simply not recognized as such. They give a few interesting examples that should suggest what they mean by that and in which directions we should look for the "hidden" cases of multiple personality. These examples include demagogues, actors, "UFO abduction" survivors, as well as a few important historical figures. See Humphrey and Dennett, "Speaking for Our Selves: An Assessment of Multiple Personality Disorder", p. 96. [↑](#footnote-ref-112)
112. Kant, *Critique of Pure Reason,* B131-2 [↑](#footnote-ref-113)
113. *Ibid*, A363 [↑](#footnote-ref-114)
114. *Ibid,* A359, A546-7=B574-5 [↑](#footnote-ref-115)
115. Carol van Kirk, "Kant and the Problem of Other Minds", *Kant-Studien* (1986), p. 12. [↑](#footnote-ref-116)
116. Kant, *Groundwork for the Metaphysics of Morals,* p. 37 [↑](#footnote-ref-117)
117. Kant, *Critique of Pure Reason,* A 346-7 [↑](#footnote-ref-118)
118. Leslie Stevenson, "Synthetic Unities of Experience", *Philosophy and Phenomenological Research* (2000), p. 283. [↑](#footnote-ref-119)
119. Kant, *Critique of Pure Reason*, B428 [↑](#footnote-ref-120)
120. *Ibid,* A213=B260 [↑](#footnote-ref-121)
121. Longuenesse, "Self-Consciousness and Consciousness of One's Own Body: Variations on a Kantian Theme", p. 303. [↑](#footnote-ref-122)
122. Valerie Gray Hardcastle and Owen Flanagan, "Multiplex vs. Multiple Selves: Distinguishing Dissociative Disorders", p. 643. [↑](#footnote-ref-123)
123. Daniel Dennett, "Where am I?", in *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge, 1981) [↑](#footnote-ref-124)
124. Daniel Dennett, "Self as a Center of Narrative Gravity", *Self and Consciousness: Multiple Perspectives* (1992), p. 275-288. [↑](#footnote-ref-125)
125. *Ibid*, p. 287. [↑](#footnote-ref-126)
126. Hardcastle and Flanagan, "Multiplex vs. Multiple Selves: Distinguishing Dissociative Disorders", *The Monist* (1999), p. 651. [↑](#footnote-ref-127)
127. Humphrey and Dennett, "Speaking for Our Selves", p. 76. [↑](#footnote-ref-128)
128. Daniel Dennett, *Consciousness Explained* (New York, 1991), p. 417. [↑](#footnote-ref-129)
129. Hardcastle and Flanagan, "Multiplex vs. Multiple Selves: Distinguishing Dissociative Disorders", p. 650. [↑](#footnote-ref-130)
130. Dennett, *Consciousness Explained*, p. 429. [↑](#footnote-ref-131)
131. This is in accord with what Forrest claims about neural networks responsible for the sense of the self - the neural reorganization of these networks is experience-dependent and is largely influenced by the environment. See Forrest, "Toward an Etiology of Dissociative Identity Disorder: A Neurodevelopmental Approach" [↑](#footnote-ref-132)
132. Dennett, *Consciousness Explained*, p. 414. [↑](#footnote-ref-133)
133. It is worth mentioning, in this respect, that Hardcastle and Flanagan even suggest developing story-telling skills in the treatment of DID. See "Multiplex vs. Multiple Selves: Distinguishing Dissociative Disorders", p. 653. [↑](#footnote-ref-134)
134. Dennett, *Consciousness explained,* 413. [↑](#footnote-ref-135)
135. Altough, for instance, Crotian philosopher Milan Kangrga (Kangrga, *Ethics and Revolution,* p. 64-65) interprets the faculty of spontaneity as trespassing into the realm of practical philosophy and having an essential practical component that Kant apparently was not aware of. Kangrga writes that Kant, by claiming that understanding prescribes the laws to the nature, does not realize that he is making an analogy between practical and speculative reason: just as the practical reason prescribes the laws to itself, in the same way speculative reason prescribes laws to the nature in an act of freedom. Kangrga considers this apperception to be an "historical act". [↑](#footnote-ref-136)
136. Dennett, *Consciousness Explained*, 428. [↑](#footnote-ref-137)
137. In the context of Dennett's "linguistic" explanation it is intriguing to mention Cassirer, a neo-Kantian, who writes about people of a number of primitive tribes relating the notion of the self to a specific linguistic construct - their names. This relation is much tighter than in "developed societies". These peoples strongly believe that the change of the name entails the change of personality - thus for instance boys change their names when undergoing rituals that introduce them into the world of adults; and just pronouncing a name of a dead person or a god is believed to cause them to appear. Some of these beliefs are echoed in superstitious traditions of the "developed societies" as well. See Cassirer, *Language and Myth,* chapter IV. [↑](#footnote-ref-138)
138. Dennett, *Consciousness Explained*, p. 429. [↑](#footnote-ref-139)