teorema
Vol. XXVI/3, 2006, pp. 177-189
ISSN: 0210-1602

Transcendental Self-Deception

Sami Pihlström

RESUMEN

Este artículo ofrece un nuevo enfoque del concepto de autoengaño al conectarlo con el de yo trascendental. Se discuten brevemente las falacias de la metafísica especulativa (especialmente los "Paralogismos") que resultan de la "ilusión trascendental" de la razón, tal como es analizada por Kant, presentándolas como casos de autoengaño trascendental metafísico. Se introduce una distinción entre formas "inflacionistas" y "deflacionistas" de tal autoengaño. Finalmente, se sugiere que el autoengaño trascendental puede tomar también una forma ética.

ABSTRACT

This paper provides a new approach to the concept of self-deception by connecting this concept with that of the transcendental self. The speculative metaphysical fallacies (especially the "Paralogisms") arising from the "transcendental illusion" of reason, as analyzed by Kant, are briefly discussed as cases of metaphysical transcendental self-deception. A distinction between "inflated" and "deflated" forms of such self-deception is introduced. Finally, it is suggested that transcendental self-deception may also take an ethical shape.

I. INTRODUCTION

Although there has been considerable debate over the concept of self-deception in recent (primarily analytic) philosophy of mind and epistemology, this concept has hardly ever been connected with the distinction between the *transcendental* and the *empirical* self. The purpose of the present essay is to establish such a connection and to argue that a transcendental approach —an attempt to place the self in the "transcendental tradition" inaugurated by Immanuel Kant [see Carr (1999)]— will significantly enrich the conceptual issues of self-deception, though it will hardly make them easier. In particular, I will explore what I call *transcendental self-deception*, the kind of self-deception the transcendental self may be said to engage in. This kind of self-deception is, we will notice, closely connected with metaphysical —both "inflated" and "deflated" [cf. Putnam (2004)]— accounts of the self.

Through my proposal to link the issues of self-deception and the transcendental self, I also hope to be able to, implicitly, throw new light on the elusive concept of *transcendental knowledge*. If transcendental thinkers like Kant and Husserl are correct, we should view ourselves as "knowing", or at least as being capable of knowing, at a transcendental level, that we *are*, in addition to our empirical, psychological selfhood, transcendental selves (and thereby responsible for constituting the structure of reality we are able to cognize and meaningfully represent); however, we should beware of self-deceptively misinterpreting this piece of transcendental knowledge as a piece of metaphysical knowledge about the existence of a substantial self as a "thing in itself".

II SELF-DECEPTION AND THE TRANSCENDENTAL SELF

Philosophers have been extremely puzzled by the phenomenon of self-deception, as is demonstrated by the almost 400 contributions to this issue listed by the *Philosopher's Index* (in August, 2006). How is the rather everyday phenomenon known as self-deception possible at all, as it seems to require that one and the same self both believes that *p* and believes that not-*p*? All the major theories of self-deception propose different answers to this problem, and all of them involve problems of their own. Some philosophers are willing to announce self-deception to be impossible and/or conceptually incoherent, while many others work toward a coherent understanding of this notion. [For diverging contributions to these issues, see, e.g., Bach (1980), Martin (1986), and the essays collected in McLaughlin and Rorty (1988) and Dupuy (1998). For a comprehensive bibliography of philosophical writings on self-deception, see http://www.philosophy.stir.ac.uk/old/cnw/self-deception.htm.]

For our purposes here it is not necessary to adopt any particular theory of self-deception. We might, for example, rest satisfied with the following working definition proposed by Robert Audi:

> A person, *S*, is in a state of self-deception with respect to a proposition, *p*, if and only if:
>
> (1) *S* unconsciously knows that not-*p* (or has reason to believe, and unconsciously and truly believes, that not-*p*);
> (2) *S* sincerely avows, or is disposed to avow sincerely, that *p*; and
> (3) *S* has at least one want that explains, in part, both why *S*'s belief that not-*p* is unconscious and why *S* is disposed to avow that *p*, even when presented with what he sees is evidence against *p* [Audi (1988), p. 94].

A lot depends here, of course, on what sort of "unconscious" knowledge and reasons to believe, or what sort of wants (desires, interests, wishes), we are dealing with. As will be explained in what follows, the relevant kind of wants that lead to self-deceptive accounts of the nature of the self itself are primarily twofold, both equally ideological: (i) the wish to maintain a (perhaps religiously inspired) belief in the immortality of the soul, on the one hand (leading to the fallacies Kant called Paralogisms), and (ii) the wish to maintain a strictly scientific picture of the world, including the mind, self, or subject, on the other (leading to the attempt to eliminate the self from the ultimate scientific metaphysics).

The notion of the transcendental self, and the distinction between the empirical and the transcendental selves, can be introduced with reference to Kant and Wittgenstein, for instance, though the phenomenological tradition, especially Husserl, should not be forgotten, either [cf. Carr (1999) and Allison (2004); see also Pihlström (2003) and (2004)]. This distinction might even be thought to solve the conceptual problems of self-deception in one fell swoop. One might simply suggest that the transcendental self deceives the empirical, or vice versa. This is hardly promising, however. The transcendental and the empirical "selves" are merely *aspects* of a single self, as both Carr (1999) and Allison (2004) maintain on the basis of their "one world" ("double aspect") interpretations of Kant; the original issue concerning how a self can deceive itself will thus immediately arise anew. There is only one world, not two; a fortiori, there are not two selves situated in two different worlds but only one to which we (ourselves, as the kind of selves we are) may adopt different perspectives. If the transcendental (metaphysical) self is construed, as in Wittgenstein's *Tractatus*, as a "limit" of the (one and only) world rather than an object in the world [see Wittgenstein (1921), §§5.6ff.], it is much easier to hold the kind of "aspectual" interpretation scholars like Allison and Carr propose. Then, the empirical self, the object of psychology, would be a thing in the world, while the transcendental self, to whom the world of objects is given, would be something quite different, though *not* another "thing" but, rather, a limit of the world of things, a condition for the possibility of there being such a world at all.

It must be noted, however, that even if we seek a way of understanding transcendental self-deception as something the transcendental self (whatever it ultimately is) engages in, we need not deny that socio-cultural aspects of self-deception, understood as a habit forgetful of the "generalized other" thematized in pragmatism, in particular [see Mitchell (2000)]. We can easily maintain such socio-cultural features even in transcendental self-deception, if we are prepared to "naturalize" and "historicize" transcendental philosophy, and thereby the transcendental self and its world-constituting role, in terms of pragmatism, as has been suggested elsewhere [Pihlström (2003)]. The broader implications of such a pragmatic reinterpretation of transcendental philoso-

phy for the phenomenon of (transcendental) self-deception must remain implicit here, however. The talk about the transcendental self as a "limit" of the world or as a "point of view", perspective, on the world may sound as entirely anti-naturalist. This, however, is not the intention of the pragmatically oriented transcendental philosopher. It is an important insight in pragmatic, naturalized transcendental philosophy that transcendentally relevant issues, including the self and self-deception, can be seen from a double perspective (or even multiple perspectives): we are dealing with something (i.e., the self and its reflexively deceptive tendencies) that is *both* entirely natural, empirical, psychological, or socio-cultural *and* implicated in the very constitution of the natural world.

Moreover, the project of investigating self-deception as a transcendental phenomenon (one such phenomenon among many others) is itself pragmatically motivated. The idea is to move forward in the practice of transcendental reflection, to connect crucial Kantian ideas with the need to cope with, and understand, psychologically problematic phenomena such as self-deception.

III Self-deception and transcendental illusion

Instead of attempting a direct solution to the puzzles of self-deception by means of the distinction between the transcendental and the empirical self, it is more promising to examine self-deception in terms of *transcendental illusion* and the metaphysical errors this unavoidable, natural illusion of human reason yields. Here, it is crucial that transcendental illusion, or (as I will go on to say) transcendental self-deception, arises *not* as a conflict of beliefs but as a conflict regarding the *epistemic status* of the concepts or ideas (of reason) that apply to the self and the world — a conflict between, say, substantial and merely formal, or constitutive and regulative, concepts or ideas. The concept of self-deception may, then, be illuminated through a study on the nature of transcendental illusion, and the metaphysical errors reason tends to arrive at on the basis of such inevitable illusion, as discussed in Kant's "Transcendental Dialectic" [see Grier (2001) and Allison (2004)].

A particularly relevant case of self-deception occurs in the dialectical fallacies Kant labels "Paralogisms" [see his chapter, "Von den Paralogismen der reinen Vernunft", Kant (1781/1787), A341/B399ff.]. As is well known, Kant, in his "critical" phase, gave up a number of "pre-critical" metaphysical doctrines he had himself maintained, including the Cartesian conception of an immaterial, substantial, simple, persisting, personal, and immortal soul. The simplicity of the soul, in particular, had been regarded as crucial for its immortality, because an absolutely simple soul cannot be thought to be disintegrate into parts [see Grier (2001), p. 165].

Let us briefly study how Kant abandons the metaphysics of the soul in the sections of the *Prolegomena* (1783) corresponding to the treatment of the Paralogisms in the *Critique* [§§46-49]. He argues that the persistence of an ultimate subject of thought, or the soul, can only be demonstrated within the world of appearance: "[P]ermanence can never be proved of the concept of a substance, as a thing in itself, but for the purposes of experience only" [§47]. He goes on:

> If therefore from the concept of the soul as a substance, we would infer its permanence, this can hold good as regards possible experience only, not [of the soul] as a thing in itself and beyond all possible experience. But life is the subjective condition of all our possible experience, consequently we can only infer the permanence of the soul in life; for the death of man is the end of all experience which concerns the soul as an object of experience, except the contrary be proved, which is the very question in hand. The permanence of the soul can therefore only be proved (and no one cares for that) during the life of man, but not, as we desire to do, after death; and for this general reason, that the concept of substance, so far as it is to be considered necessarily combined with the concept of permanence, can be so combined only according to the principles of possible experience, and therefore for the purposes of experience only [§48].

Accordingly, the principle stating the permanence of substances only holds for things insofar as they are appearances, not non-empirically, and thus not *post mortem* [§48n]. In an empirical sense, a permanent self (or "soul") is a part of the world of appearances, exactly as all other natural things. Just as external bodies, the appearances of the outer sense, do not (in contrast to things in themselves) exist independently of my thoughts, transcendentally speaking, nor do I, as an appearance of the inner sense, as a soul in the sense of empirical psychology, exist independently of my representational capacities [§49]. Here we arrive at key issues of Kant's transcendental idealism, which, in contrast to Berkeleyan dogmatic and Cartesian skeptical idealisms, views spatiotemporal phenomena, including those of inner sense, as elements of a transcendentally constituted natural, empirical world [cf. Kant (1781/1787), B274ff., A377ff.].

Exactly as in the other illusions discussed in the Transcendental Dialectic (viz., the Antinomy and the Ideal of Pure Reason), in the case of the Paralogisms, our reason tends to infer from a transcendental concept necessary as a formal presupposition of experience — in this case, from the original unity of the transcendental apperception, i.e., the fact that the prefix "I think" must be able to accompany all my representations [B131-132], which is a starting point for the transcendental deduction of the pure concepts of understanding, or the categories —to a substantial, metaphysical thesis about things in themselves— in this case, to a statement about the existence and fundamental properties of a substantial, persisting, and therefore (in princi-

ple) immortal soul independent of the phenomenal world and the conditions of experience. Such inferences are illegitimate. The "ideas of reason" invoked in them (soul, freedom, God) can function only regulatively, not constitutively (as the categories do). We can have no cognitions but only a problematic concept of the objects corresponding to those ideas [A339/B397]; they do not have "objective significance" for us.

The metaphysician engaged in paralogistic fallacies thus conflates things in themselves with mere appearances. Illusion and confusion result, when our a priori knowledge about the subjective conditions of thought (e.g., the "I think") is erroneously taken to provide us with metaphysical knowledge about a peculiar object, the soul. As Henry Allison emphasizes, the pure form of a thinking subject, the logical subject of thought, is in the Paralogisms conflated with a noumenal subject (in Kant's "positive" sense of "noumenon") taken to exist as a "transcendent" substance (instead of a transcendental principle) in the world, a subject as a peculiar kind of object that is taken to be describable by means of non-sensible predicates referring to the transcendent; here, a purely formal subject is illegitimately hypostatized into a substantial thing [see Allison (2004), ch. 12, especially pp. 347-348, 354-356]. This "hypostatization" of the self can now be understood as an essentially self-deceptive manoeuvre (though this is not how Allison, or any other Kant scholar I am familiar with, expresses the matter). The self, wishing that immortality be true, deceives itself into a metaphysical speculation about its own substantial properties, although all it is justified in postulating, from the perspective of "critical" philosophy, is a formal unity of apperception.

In a recent study, Michelle Grier has with great care analyzed Kant's doctrine of transcendental illusion. She reminds us that illusions and dialectical fallacies must be distinguished from each other: while the transcendental illusions —i.e., the ideas of pure reason, including the idea of the soul— as such are inevitable for human reason, dialectical metaphysical inferences, such as the Paralogisms, are not. We can liberate ourselves from the latter but not from the former [Grier (2001), pp. 10, 143, 263, 303-304.] The transcendental illusions lead to fallacies only when connected with a transcendental misuse of concepts (or categories). The pre-critical doctrine Kant labelled "transcendental realism" is primarily responsible for the fallacies.

Grier offers a detailed treatment of the "pseudo-rational" idea of the soul, functioning as a premise of the Paralogisms [*ibid*., ch. 5]. The error is to suppose that one could move from the transcendental subject to metaphysical theses about a substantial, simple, self-preserving person [*ibid*., p. 144]. It is, again, the metaphysical, constitutively intended use of an idea of reason that yields erroneous metaphysics: the condition of thought (the self, or the soul) is regarded as objectively real, even though we cannot possess a corresponding concept of an object [*ibid*., pp. 147, 152]. Grier basically agrees with Allison when she notes [*ibid*., p. 169] that the illusory move will have been

made when one steps from the concept of a transcendental self to an idea of reason (the absolutely unconditioned unity of the conditions of thought) and when this logically absolute unity is hypostatized into the real absolute unity of a metaphysical entity (the soul). It is a separate matter whether Kant accuses the metaphysician of regarding the soul as a noumenal or a phenomenal object. Grier maintains that both accusations may be relevant. The self is no object at all, neither noumenal nor phenomenal [*ibid*., pp. 159-160].

We need not here dwell on the details of Kant's argumentation, nor on the conflicting interpretations that have been offered on the relevant chapter. [See the literature cited by Allison (2004) and Grier (2001), as well as Ameriks (2006).] It is, rather, important to note that the kind of metaphysical errors the metaphysician involved in paralogistic inferences arrives at can be seen as self-deceptive in their basic nature (though this is clearly *not* the way in which either Kant himself or the commentators I have cited articulate the problem). Here, unlike in the cosmological or theological illusions, the self, or more specifically its capacity for reason, is metaphysically misled to confused and illegitimately maintained views about *itself*. Moreover, it leads itself to this disaster. It *should* know better. This is why its errors are, transcendentally speaking, self-deceptive. The "want" motivating the self-deceptive way of thinking, in the case of the Paralogisms, is obviously the desire to maintain a religiously relevant conception of an immortal soul. This desire prevents the self-deceptive metaphysician from realizing the merely formal status of the unity of her/his own subjectivity (apperception) as a transcendental condition of the cognitive experience s/he is capable of.

Given the religiously inspired aspiration for immortality, as a source of metaphysical self-deception, a side issue might also be invoked here: is religious self-deception in general analogous to, or different from, the kind of transcendental errors speculative metaphysics commits? Possibly, the notion of superstition, for instance, could be analyzed in terms of self-deception. Here one is tempted to think about concepts such as pseudo-religion, hypocrisy, superstition, etc. The argument of the present paper does not depend on any specific view on religious self-deception, however. Nor am I suggesting that religious views would inevitably be metaphysical, or as vulnerable to criticism as the kind of metaphysics attacked by Kant and others. I only want to leave room for the possibility that in this area, too, one may find cases of self-deception that can be transcendentally analyzed.

IV. Self-deception and metaphysics

Is metaphysics (religious or not), then, inevitably an exercise in self-deception, because based on transcendental illusion? I hope to be able to defend a more positive, yet critical, picture of metaphysics [cf. Pihlström (2007b)].

The use of a transcendental method in metaphysical inquiry, and the postulation of a transcendental self, *need not* be committed to self-deception; or, at least, we may draw a distinction between "bad", speculative, self-deceptive metaphysics entangled with transcendental illusion and "good", non-speculative, critical metaphysics that avoids such entanglement. Yet, some forms of (transcendental) metaphysics, such as metaphysical solipsism [Pihlström (2004)] or the kind of hypostatization of the self we have seen Kant criticize in the Paralogisms, may be essentially self-deceptive. Accordingly, should we say that, for example, the solipsist, like the metaphysician in the Paralogisms, self-deceives when imagining her-/himself to be the only genuine subject in the world, whose existence would, then, be ontologically dependent on her/his experiences or mental states? Perhaps only a self-deceptive maneuver can lead one to such a wild belief that one is alone in the world, its one and only "real" subject. Or is the solipsist just wrong? If we see her/him as engaging in self-deception, we must claim her/him to believe deep down, "unconsciously" perhaps, that the world is *not* simply reducible to her/his own construction but yet to believe to the contrary, because of a strong wish to elevate her/his own subjectivity to a privileged position in comparison to everything else.

In some cases we can, though in some others we perhaps cannot, analyze the metaphysical illusions reason itself produces (for it, itself, to get entangled in) as forms of self-deception. As a therapy, what we would then need is something like the "self-discipline" of reason discussed by Kant in the chapter on the "Discipline of Pure Reason" in the *Methodenlehre*, Part II of the First Critique. Is such a discipline a reliable guard against reason's tendency to self-deceive? The duty to construct, reflexively, such a self-discipline, and to let oneself to be guided by it, is both intellectual and moral (with no sharp distinction between the two). But it is, as any enterprise of reason, eminently fallible. It can never guarantee that self-deceptive errors of metaphysics do not occur.

The illusory Paralogisms, in particular, can be understood as involving a specific kind of metaphysical self-deception, because the self here misconstrues its own capacities, possibilities, and ultimate nature. Mistakenly, reason postulates a substantial, immortal self, though only a formal unity of transcendental apperception can be legitimately reached as a transcendental condition for the possibility of objective cognition. The remedy in this special case, as in the more general case of reason's self-deceptive tendency as such, is —and can only be— the self's own (self-)critical, transcendental, disciplined use of reason. Only by means of a self-reflective discipline that will reveal transcendental illusions can we hope to avoid the speculative metaphysician's self-deceptive mistakes.

Another, very important kind of transcendental self-deception, equally metaphysical (and equally in need of transcendental critique) is what may be

labeled *fictionalism* or *eliminativism* about the self, the mind, or subject(ivity). Such a position has famously been defended by Daniel Dennett, among others [e.g., Dennett (1991)], and it has been elaborated on in some of the best recent discussion of the nature of the self [e.g., the essays collected in Strawson (2005)]. For the fictionalist, there is no "real" self at all; we simply tell ourselves a fictional story about the existence of such a unified self. There may be some truth in this view, given the criticisms of the self as an object launched by transcendental philosophers from Kant to Wittgenstein (see above). However, the fictionalist theorists of the self usually also eliminate the transcendental self, treating it as a mysterious, ghostly entity that should not be postulated by any scientifically respectable philosopher. This is a grave mistake, as has been observed [for a vigorous transcendental criticism of Dennett's view, in particular, see Carr (1999), pp. 123-124]. It is very difficult to maintain the position that the self is a mere fiction, because then there would be no one by whom or to whom the fictitious story would be told. Someone needs to be able to tell, hear, and interpret such a fictional story – and thus we seem to come back to the transcendental subject as the "always already" presupposed condition for the possibility of the world as a world of objects (for us), a world within which the distinction between fiction and non-fiction makes sense. However, we must then again be careful to avoid hypostatizing that self into a special kind of object in the world.

The self, then, clearly deceives her-/him-/itself by claiming, however "scientifically", argumentatively, and intellectually honestly, that s/he/it does not exist. *Who* actually deceives or is deceived here? Well, the self, of course — the transcendental self itself, which, to paraphrase Wittgenstein — mistakes its own position at (or as) the limit of the world to that of a part of the world, or an object in the world, which might (as any contingent object) either exist or fail to exist. As Walker (2006) also reminds us, there is a very simple transcendental argument proceeding from the fact that there is experience to a necessary condition for the possibility of such experience, i.e., the existence of the self or subject of experience. The transcendental philosopher must only, as argued, beware of hypostatizing this self, which is her/his "own", into a substantial entity. The trick is to walk the middle path between substantial, Cartesian metaphysics of the self, on the one side, and fictionalist, eliminativist (usually scientistic) accounts, on the other. Both of these extremes, in their distinctive ways, are entangled in self-deception. While the Paralogistic fallacies lead to inflated metaphysical theories of the self, the fictionalist or eliminativist views are no less metaphysical; they are, however, not *inflated* but *deflated* metaphysical accounts of the self. [For the distinction between inflated and deflated "Ontology", see Putnam (2004), ch. 1.]

At a transcendental level, then, we —insofar as we are able to view ourselves as selves at all— "know", though may not realize that we are even able to know, that our subjectivity is not only phenomenal but also transcendental,

world-constituting, and it is only through a kind of self-deception that we can hide this piece of (potential) philosophical knowledge from ourselves. This is by no means ordinary self-deception, but the transcendental perspective enriches the traditional picture of self-deception; and, conversely, the interpretation of both the paralogistic fallacies and the fictionalist attempts to eliminate the self (especially the transcendental self) as forms of self-deception in turn enriches our conception of transcendental philosophical knowledge.

It was noted above that the reason why paralogistic metaphysics amounts to self-deception is the underlying want to preserve a religiously relevant immortal soul. In the case of the self-deceptive metaphysics the fictionalist or eliminativist is entangled with the underlying want or desire playing an analogous role in the emergence of self-deception is, of course, different. It is, however, no less ideological, and, as I suggested, the resulting theory is no less metaphysical. In this latter case, the self-deceiver wants to preserve a fundamentally physicalist picture of the world, a "scientific image" that subjectivity, let alone transcendental subjectivity, does not seem to fit easily. Again, the self-deception that arises on these grounds can be corrected by means of a transcendentally self-disciplined argument leading to the insight that the transcendental self must always already have been presupposed in any attempt to draw a contrast between fictional and non-fictional images or stories — presupposed not as a metaphysical entity but as a limit, principle, or perspective.

It must also be kept in mind that metaphysical self-deception results not merely from fictionalist, eliminativist accounts of the self, insofar as these are based on a scientistic ideology (an underlying wish to maintain a scientific conception of the world and to place the self in it). More broadly, the scientistic mainstream paradigm of contemporary analytic philosophy of mind can be regarded as self-deceptive in a profound sense — though it would require a much longer and more ambitious paper to substantiate this thesis. It is always, at least to an extent, self-deceptive to view oneself as a mere object in the world, an object whose existence (or objective nature) would be so much as a problem, a problem which might invite fictionalist or eliminativist suggestions. Even if the metaphysician defends a realist account of the mind, either reductive or non-reductive, s/he might be guilty of the kind of self-deception that forgets that one's self is essentially a subjective point of view rather than a peculiar kind of object in the world (whose existence as a world of objects is, again, made possible only by the transcendentally constitutive subjective point(s) of view).

V. THE SELF-DECEPTIVE MORAL SELF

Self-deception has been actively discussed in relation to moral philosophy, too [see, e.g., Martin (1986)], with Jean-Paul Sartre's concept of "bad

faith" as a standard reference. However, the transcendental kind of ethical self-deception I briefly want to take up here is closely connected with metaphysical forms of transcendental self-deception. We might say that the self engages in ethical self-deception, if s/he/it forgets the human condition presupposed by its being a self in the first place, a condition which is structured by moral responsibility and thus by the ineliminable potentiality of moral guilt.

This is something we may label *transcendental guilt* (for a more detailed account, see Pihlström [2007a]). If the self lacks recognition of such fundamental (though primarily potential rather than actual) guilt, it ethically self-deceives. Here it is important to note that guilt, as an irreducible moral category, has a constitutive role to play in our ways of conceptualizing our relations to other people. Without experiencing guilt, or being able to do so, we would not be capable of employing the moral concepts and judgments we do employ. Elaborating on this transcendental argument, it is possible to arrive at a challenging "metaphysics of guilt" reminding us of our constantly potentially guilty existence in-relation-to-others. More generally, through such an account, we may perceive that an adequate moral theory can and should pay attention to the transcendental status of morally relevant emotions such as guilt — emotions that are, arguably, constitutive of our concept of moral seriousness. Otherwise, the moral self may easily indulge in self-deception. Instead of simply psychologizing moral emotions, however, an inquiry into transcendental guilt may employ, say, Raimond Gaita's Wittgenstein-inspired way of examining the place of the concepts of guilt and remorse in our ethical language-use [see Gaita (2004), as well as Pihlström (2007a)].

Moreover, if the pursuit of metaphysics itself ultimately has an ethical basis, as I would be prepared to argue on independent grounds [Pihlström (2005)], then the phenomenon of moral self-deception, transcendentally construed, is even more basic than the kinds of metaphysical self-deception examined in the previous section. Our whole metaphysics, not only our metaphysical account of our own place in the world's scheme of things, might fundamentally be based on self-deception, e.g., if we believe that it is the task of metaphysics to describe morally neutral, value-independent facts and categories.

From the ethical point of view, one of my results is the deep identity of the transcendental self and what we may call the ethically engaged "existential self". Both, in a suitably "naturalized" account of transcendental philosophy (and transcendental subjectivity), are "worldly", world-embedded and embodied [cf. Pihlström (2003)], *in der Welt*, to use Heideggerian jargon. Both are also capable of self-deception or "bad faith". Both are, moreover, primarily ethically oriented, as the discussion of transcendental guilt should bring to the fore. I hope I am not deceiving myself when I subscribe to the view that these two "selves", despite their insubstantiality, are ultimately one and the same — and that neither of them should be self-deceptively confused

with a Cartesian soul (or, for that matter, eliminated). The transcendentally guilty moral self whose true nature we should not self-deceptively hide from ourselves is, fundamentally, a point of view on the world rather than an object within the world; it is a point of view whose constant challenge is to see the world in an ethical lighting. This is the challenge of achieving a "moral vision", so to say. Insofar as this challenge is forgotten, what results is precisely the kind of forgetfulness that transcendental ethical self-deception consists in.

VI. Conclusion

I have identified two different varieties of transcendental self-deception, and explained why they are forms of self-deception. The first of these is the *metaphysical* variety, which comes in two basic forms: the self may be self-deceived about its metaphysical status (as in the Paralogisms) or about its alleged non-existence (as in fictionalist accounts of the self). We may, I suggested, call these the inflationary and the deflationary varieties of metaphysical transcendental self-deception, respectively. The second kind of transcendental self-deception is *ethical*. It involves a morally relevant forgetfulness of one's human condition, especially what I have called "transcendental guilt". Both the metaphysical and the ethical varieties are varieties of self-deception primarily in being forgetful of human limits and the human condition, to which the self must always already be committed in any case, simply in order to be a self.

These issues, particularly the relation between metaphysical and ethical transcendental self-deception, certainly require further scrutiny. This paper has only been able to set some tasks for reflexive transcendental inquiry into the self's deceptive tendencies and their metaphysical and ethical relevance. It is important to categorize the different forms of transcendental self-deception, in order to avoid the metaphysical speculations they may yield, just as it is, in more everyday (psychological) cases, important to be aware of the kind of self-deceptive strategies we habitually use, in order to lead a more self-reflectively coherent and integrated life. In a more comprehensive treatment of transcendental self-deception, both historical work (especially on Kant and other theorists of transcendental subjectivity) and systematic conceptual analysis and argumentation will be needed.

Department of Social Sciences and Philosophy
University of Jyväskylä
P.O. Box 35, FI-40014 Finland
E-mail: sami.pihlstrom@helsinki.fi

REFERENCES

ALLISON, H. E. (2004), *Kant's Transcendental Idealism: An Interpretation and Defense – A Revised and Enlarged Edition*, New Haven, CT and London, Yale University Press (1st ed. 1983).

AMERIKS, K. (2006), "The Critique of Metaphysics: The Structure and Fate of Kant's Dialectic", in Guyer, P. (ed.), *The Cambridge Companion to Kant and Modern Philosophy*, Cambridge, Cambridge University Press, pp. 269-302.

AUDI, R. (1988), "Self-Deception, Rationalization, and Reasons for Acting", in McLaughlin and Rorty (eds.), pp. 92-120.

BACH, K. (1980), "An Analysis of Self-Deception", *Philosophy and Phenomenological Research*, vol. 41, pp. 351-370.

CARR, D. (1999), *The Paradox of Subjectivity: The Self in the Transcendental Tradition*, Oxford, Oxford University Press.

DENNETT, D. C. (1991), *Consciousness Explained*, London: Penguin.

DUPUY, J.-P. (ed.) (1998), *Self-Deception and Paradoxes of Rationality*, Stanford, CA, CSLI Publications.

GAITA, R. (2004), *Good and Evil: An Absolute Conception*, rev. ed., London and New York, Routledge (1st ed. 1991).

GRIER, M. (2001), *Kan´s Doctrine of Transcendental Illusion*, Cambridge, Cambridge University Press.

KANT, I. (1781/1787), *Kritik der reinen Vernunft*, Schmidt, R. (ed.), Hamburg, Felix Meiner, 1990.

— (1783), *Prolegomena to Any Future Metaphysics*, Fieser, J. (ed.), Carus, P. (trans.), based on the 1902 English translation; online: http://philosophy.eserver.org/kant-prolegomena.txt (retrieved September, 2006).

MARTIN, M.W. (1986), *Self-Deception and Morality*, Lawrence, University Press of Kansas.

MCLAUGHLIN, B.P. & RORTY, A.O. (eds.) (1988), *Perspectives on Self-Deception*, Berkeley, University of California Press.

MITCHELL, J. (2000), "Living a Lie: Self-Deception, Habit, and Social Roles", *Human Studies*, vol. 23, pp. 145-156.

PIHLSTRÖM, S. (2003), *Naturalizing the Transcendental: A Pragmatic View*, Amherst, NY, Prometheus/Humanity Books.

— (2004), *Solipsism: History, Critique, and Relevance*, Tampere, Tampere University Press.

— (2005), *Pragmatic Moral Realism: A Transcendental Defense*, Amsterdam and New York, Rodopi.

— (2007a), "Transcendental Guilt: An Emotional Condition of Moral Experience", *Journal of Religious Ethics*, vol. 35.

— (2007b), "Transcendental Philosophy as an Ontology", forthcoming in Haaparanta, L. and Koskinen, H.J. (eds.), *Categories of Being: Essays on Metaphysics and Logic*.

PUTNAM, H. (2004), *Ethics without Ontology*, Cambridge, MA and London, Harvard University Press.

STRAWSON, G. (ed.) (2005), *The Self?*, Malden, MA and Oxford, Blackwell.

WITTGENSTEIN, L. (1921), *Tractatus logico-philosophicus: Logisch-philosophische Abhandlung*, Frankfurt am Main, Suhrkamp, 1961.