

Audun Øfsti · Peter Ulrich · Truls Wyller (eds.)

Table of Contents

## Indexicality and Idealism II

### The Self in Philosophical Perspective

*[Faint, illegible text from the table of contents, including author names and page numbers.]*

mentis  
Paderborn

Audun Qvortved · Peter Ulrich · Ingrid Wyster (eds.)

## Indexicality and Idealism II

The Self in  
Philosophical Perspective

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Ein Titeldatensatz für diese Publikation ist bei  
Der Deutschen Bibliothek erhältlich.

Gedruckt auf umweltfreundlichem, chlorfrei gebleichtem  
und alterungsbeständigem Papier © ISO 9706

© 2002 mentis, Paderborn  
(mentis Verlag GmbH, Schulze-Delitzsch-Straße 19, D-33100 Paderborn)  
[www.mentis.de](http://www.mentis.de)

Alle Rechte vorbehalten. Dieses Werk sowie einzelne Teile desselben sind urheberrechtlich  
geschützt. Jede Verwertung in anderen als den gesetzlich zulässigen Fällen ist ohne vorherige  
Zustimmung des Verlages nicht zulässig.

Printed in Germany  
Herstellung: Druckhaus Plöger  
Einbandgestaltung: Anna Braungart, Regensburg  
Druck: Elbe-Werkstätten GmbH, Hamburg  
ISBN 3-89785-194-6

## Table of Contents

Preface .....	7
Contributions	
<i>Geert Keil</i> Der Nullpunkt der Orientierung .....	9
<i>Truls Wyller</i> Reference, Reflexivity, and Transcendental Idealism .....	31
<i>Olav Asheim</i> Indexicals and Direct Reference .....	41
<i>Siri Granum Carson</i> The Dependency of Non-Indexical Languages .....	55
<i>Anton Friedrich Koch</i> Hegel über Indexikalität und sinnliche Gewißheit .....	65
<i>Peter Ulrich</i> Strawson über Kant .....	85
<i>Jens Saugstad</i> "I think ....". Kant on Self-Consciousness .....	103
<i>Louise Röska-Hardy</i> Indexicality and Mental Concepts .....	127
<i>Helmut Pape</i> What Indexical Signs Really Do .....	145
Bibliography .....	161
Index .....	167

# "I think ...."

## Kant on Self-Consciousness

Jens Saugstad

The profound insights associated with Wittgenstein's rejection of a private language set a standard in the philosophy of mind. In this paper I argue that Kant's doctrines of pure and empirical self-consciousness incorporate these insights while avoiding Wittgenstein's too negative conclusions.

### 1. The later Wittgenstein on "I"

According to Peter Hacker, the discussion of the self from Descartes to the present is nurtured by a grammatical misunderstanding of the first-person pronoun that was exposed by the later Wittgenstein. Let me begin by presenting a fragment of Hacker's story.

Whereas Descartes took the word "I" to refer to a thinking substance and Locke assumed it to refer to a self constituted by consciousness, "Hume sank deeper into the quagmire." For his project of finding out whether there is an impression that gives rise to the idea of a self only makes sense on the pre-supposition that "I" is a referring term. Hacker generously accredits Kant with pinpointing "the source of error as the confusing of the purely formal character of 'transcendental self-consciousness' with consciousness of a pure soul-substance." However, this only led to his obscure theory, "according to which 'I' on the one hand signifies the form of all experience, and on the other mysteriously refers to a noumenal object, the moral self." How did Kant reach this conclusion? That is not entirely clear, Hacker admits, but it has to do with transcendental idealism:

This empirical 'self' is indeed aware of fleeting experiences and conscious of itself as so aware. Being an object with an autobiography, a history, it must be part of the temporal, phenomenal world. Hence the 'self' that appears to itself is itself mere appearance. So the 'self' as it is in itself, the 'noumenal self', must underlie it, as noumena underpin phenomena. This entity must be both unknowable and also the known subject of the moral law. For the moral 'self' must be free, subject to the laws of practical reason, rather than to the constraints of causality. So it must be super-sensible and atemporal and must also have a moral history, being the subject of temporal moral judgements. 'I have an obligation to ...' seemingly refers to the ethical subject, the bearer of good and evil; but how, within the bizarre framework of transcendental idealism, that is possible is never made clear.<sup>1</sup>

<sup>1</sup> P.M.S. Hacker, *Wittgenstein. Meaning and Mind*, Part 1, Essays (Oxford: Blackwell, 1993), pp. 210-211; cf. pp. 221-222.

So in the end Kant was unable to free himself from the muddled thinking of the tradition from Descartes. The cure for all the above mistaken philosophical views, according to Hacker, is to adopt the later Wittgenstein's position that "I" is not a referring term.

In *The Blue Book* Wittgenstein says: "To say 'I have pain' is no more a statement *about* a particular person than moaning is." And when the interlocutor objects: "But surely the word 'I' in the mouth of a man refers to the man who says it; it points to himself; and very often a man who says it actually points to himself with the finger," Wittgenstein replies: "But it was quite superfluous to point to himself. He might just as well only have raised his hand. It would be wrong to say that when someone points to the sun with the hand, he is pointing both to the sun and himself because it is *he* who points; on the other hand, he may by pointing attract attention both to the sun and to himself." And "If, in saying 'I', I point to my own body, I model the use of the word 'I' on that the demonstrative 'this person' or 'he.'" However, "The difference between the propositions 'I have pain' and 'he has pain' is not that of 'L.W. has pain' and 'Smith has pain'. Rather, it corresponds to the difference between moaning and saying that someone moans." Thus, "In 'I have pain', 'I' is not a demonstrative pronoun."<sup>2</sup> In a lecture from the early 1930ies he puts the point as follows: "The word 'I' does not refer to a possessor in sentences about having an experience,"<sup>3</sup> and in *Philosophische Untersuchungen* he remarks succinctly: "Ich' benennt keine Person, ...."<sup>4</sup> It seems plausible that Wittgenstein persisted in rejecting that the word "I" is a referring term, for in a lecture from 1946 or -47 he said: "If you consider substituting a signal for 'I suffer' you see that the first mistake is to take 'I' as *standing* for something."<sup>5</sup>

According to Wittgenstein's diagnosis, the idea that "I" is a referring term stems from our proneness to confound psychological (and more generally, mentalistic) sentences in the first and the third person. In the third person the question of *identifying* who is in pain can indeed be raised, but it cannot arise in the first-person singular present tense. Thus, when I exclaim, "I am in pain," Hacker says, "I am not selecting or picking out one person among others; .... Rather, what I say *enables my hearers* to identify who is in pain." The confusion consists in thinking that by identifying myself *to my audience* I thereby also identify myself *to myself*. And as "The role of the referring expression," according to Hacker, "is to specify which person one is talking about," he concludes that the first-person pronoun "I" does not qualify as a referring term.<sup>6</sup>

<sup>2</sup> Ludwig Wittgenstein, *The Blue and Brown Books* (Oxford: Basil Blackwell, 1958), pp. 67-68.

<sup>3</sup> Quoted in John W. Cook, *Wittgenstein, Empiricism, and Language* (New York: Oxford University Press, 2000), p. 28.

<sup>4</sup> Wittgenstein, *Philosophische Untersuchungen* (Frankfurt am Main: Suhrkamp Verlag, 1971), § 410.

<sup>5</sup> Quoted in Cook, *Wittgenstein, Empiricism, and Language*, p. 31.

<sup>6</sup> Hacker, *Wittgenstein. Meaning and Mind*, p. 226.

This view has struck many people as counterintuitive. Actually, Wittgenstein himself found his own position paradoxical, Hacker tells us:

Wittgenstein's endeavour was to draw our attention to differences between the word 'I' and proper names, descriptions, and other personal pronouns, differences in function, in identification of the bearer, in grammatical combination with other expressions, in verification, etc. These differences are compelling, and they led Wittgenstein to deny – paradoxically, as he admitted – that the word 'I' is a referring expression at all.

Still, nothing prevents the Wittgensteinian, Hacker believes, from saying that the first-person pronoun refers:

What matters crucially is that one be aware of the differences; and if thereafter, one still wants to say that 'I' is nevertheless a kind of referring expression or, better perhaps, a degenerate referring expression, nothing need hang on that preference as long as one does not assimilate the function of the word 'I' to an inappropriate paradigm of reference.<sup>7</sup>

I believe Wittgenstein's position contains profound truths. And I grant Hacker that many of the views about the nature of the self in our philosophical tradition stem from construing the first-person pronoun as a referring term according to inappropriate paradigms. Wittgenstein does, however, deny too much. Although it is true that "I" does not, in an important sense, refer to the person who uses the word, it seems to me that Wittgenstein did not – and could not – allow *any* sense in which it does.

## 2. Wittgenstein's critique of a private language

The comparison of "I have pain" to moaning testifies to the fact that Wittgenstein's view that the first-person pronoun is not a referring term is intimately connected with his rejection of a private language. According to Wittgenstein, the belief in a private language arises when grammatical rules that exclude doubt from sensation language-games in the first-person singular present tense are falsely interpreted as evidence for infallible knowledge of oneself having sensations. The picture that holds us captive<sup>8</sup> is the idea that a sentence like "I have pain" is an introspective judgment about my own pain just as "The table is square" is an observation judgment about an external object. The result is that pain and inner states in general are turned into private objects about which only I can have knowledge. As an aid to get outside the picture, Wittgenstein suggests as a possibility that words that are later used to refer to sensations are introduced as replacements for the natural expressions of the relevant sensation.

<sup>7</sup> Hacker, *Wittgenstein. Meaning and Mind*, p. 228.

<sup>8</sup> Wittgenstein, *Philosophische Untersuchungen*, § 115.

For instance, a child who has hurt himself and cries learns to replace crying with verbal exclamations and later with sentences.<sup>9</sup>

There can be no doubt that we often describe our pains – to the physician we consult about it, for instance.<sup>10</sup> Wittgenstein, however, is often criticized for not allowing first-person *reports* or *descriptions* of one's own mental states.<sup>11</sup> Although the comparison of "I have pain" to moaning have suggested this to many readers,<sup>12</sup> it was hardly his intention. At the most primitive level "I have pain" functions as a mere verbal exclamation of pain replacing the natural expression – as a verbal cry. But Wittgenstein did not believe that such sentences necessarily remain mere verbal exclamations. As Hacker emphasizes, his point is rather that describing one's inner states is *very different* from describing physical objects. We should not be deluded by the fact that both cases are called descriptions, for the word "description" has a different meaning in the two. Pointing to the expressive *roots* of self-descriptions is a powerful way of counteracting our tendency to confound these different kinds of description. Thus, reports or descriptions like "I have a throbbing pain in my knee" are not introspective judgments based on calm "inspection" of my pain, but "more refined forms of verbal expressive behaviour ... rooted in the more primitive ones."<sup>13</sup>

But although we can describe our own sensations, Wittgenstein is famous for his rejection of the possibility of *self-knowledge* of pains and other sensations: "Von mir kann man überhaupt nicht sagen (außer etwa im Spaß) ich wisse, daß ich Schmerzen habe." His main reason is that "es hat Sinn, von Andern zu sagen, sie seien im Zweifel darüber, ob ich Schmerzen habe; aber nicht, es von mir selbst zu sagen."<sup>14</sup> However, the situation is different for the emotions and motives for acting, Hacker explains:

A highly self-conscious person is one who attends to his emotional and conative life, who registers the ebb and flow of his passions, reactions, and attitudes. Such a person reflects upon his responses, analysis them, and searches for patterns that inform them. His descriptions of his mental states will typically be sensitive, detailed, and articulate. But even here, at this end of the spectrum of first-person psychological propositions, the relevant descriptions are very unlike descriptions of the 'outer'. They may be observant, but do not generally rest on observation. They may, in various ways, be inadequate or defective, but not because of misperception. Their typical flaws are likely to be forms of *self-deception*, rooted in a defect of the will rather than of the intellect, let alone of the senses. ... There is no

<sup>9</sup> Wittgenstein, *Philosophische Untersuchungen*, §§ 243ff.

<sup>10</sup> I owe this illustration to Robert Spaemann.

<sup>11</sup> John Searle, *The Rediscovery of the Mind* (Cambridge, Massachusetts: The MIT Press, 1992), p. 146.

<sup>12</sup> Cf. Louise Röska-Hardy, "Idealism and the 'I' of Self-Ascriptions," in Øfsti, Ulrich and Wyller (eds.), *Indexicality and Idealism* (Paderborn: mentis 2000), p. 58.

<sup>13</sup> Hacker, *Wittgenstein. Meaning and Mind*, pp. 85ff, esp. pp. 91-92.

<sup>14</sup> Wittgenstein, *Philosophische Untersuchungen*, § 246.

such thing as self-deception with respect to sensations, but our emotional life and avowals of motivation are run through with the distorting influence of the will and fantasy. So here, unlike the case of sensations and spontaneous avowals that are expressions of the 'inner', there is room for knowledge and error, i.e. self-knowledge properly speaking, and self-deception or failure to realize the pattern of one's reasons and desires. But this kind of knowledge, ignorance, and error is altogether unlike knowledge, ignorance, and error regarding what is 'outer'.<sup>15</sup>

I find this passage instructive. It shows how Wittgenstein's thesis about first-person sentences combines with a fruitful perspective on self-knowledge. Nevertheless, it is hard to accept the exemption of sensations from self-knowledge. Wittgenstein's claim that I cannot *know* that I have pain seems just as counterintuitive as his claim that the first-person pronoun is not a referring term at all.

In the next section I present Kant's critique of rational psychology, arguing that it shares the essence of Wittgenstein's claims about the first-person pronoun, while section 4 argues that Kant's theory of pure self-consciousness allows for a *practical* interpretation of the referential function of the first-person pronoun. In section 5 I argue that Kant's doctrine of inner sense anticipates Wittgenstein's thesis about the public nature of psychological concepts – while still permitting self-knowledge about sensations. In the final section I indicate why Wittgensteinians are mistaken if they believe, with Hacker, that the rejection of inappropriate ways of construing the reference of the first-person pronoun amounts to a wholesale refutation of metaphysical views about the person, including Kant's.

### 3. Kant's critique of rational psychology

One of the purposes of Wittgenstein's grammatical remarks about the first-person pronoun is to avoid metaphysical conceptions of the person like Cartesian substance dualism. In *The Blue Book* he discusses the temptation to see the word "I" used as subject (in sentences like "I see-so-and-so," "I think it will rain" and "I have toothache") as referring to something immaterial – the mind – which inhabits the body:

We feel then that in the cases in which "I" is used as subject, we don't use it because we recognize a particular person by his bodily characteristics; and this creates the illusion that we use this word to refer to something bodiless, which, however, has its seat in our body. In fact *this* seems to be the real ego, the one of which it was said, "Cogito, ergo sum".<sup>16</sup>

This comment may have inspired Hacker's criticism of Kant too, but in fact there is a striking parallel between Wittgenstein's view here and Kant's critique

<sup>15</sup> Hacker, *Wittgenstein. Meaning and Mind*, pp. 95-96.

<sup>16</sup> Wittgenstein, *The Blue and Brown Books*, p. 69; cf. p. 66.



of rational psychology in *Kritik der reinen Vernunft* (the Paralogisms)<sup>17</sup>. Rational psychology is a classical metaphysical doctrine of the nature of the self with strong affinities to Descartes, and summed up in the view that the thinking subject is an immaterial, incorruptible, numerically identical, immortal substance, that is, a spiritual substance interacting with bodies in space (including the human body) (B 403). How does Kant argue against this view?

According to the Transcendental Deduction of the Categories our capacity to know the world is made possible by our self-consciousness as thinking subjects – by pure apperception. In the Transcendental Aesthetics we were told that pure apperception is the simple representation (*einfache Vorstellung*) of the I (B 68), and in § 16 of the Deduction Kant makes the famous statement: "Das: *Ich denke*, muß alle meine Vorstellungen begleiten können" (B 131-132). The "*Ich denke*" is the sole *Satz* or *Text*, Kant says, upon which the alleged science of rational psychology builds. Notice also that it is an incomplete sentence to be completed with various concepts,<sup>18</sup> at the most abstract level with the categories: "*Ich denke die Substanz, die Ursache usw*" (B 400-401).

Thus, rational psychology attempts to determine the nature of the thinking subject solely on the basis of inferences from premises generated by the general expression of pure self-consciousness: "I think ...." The argument has the following form:

[P<sub>1</sub>:] Was nicht anders als Subjekt gedacht werden kann, existiert auch nicht anders als Subjekt, und ist also Substanz.

[P<sub>2</sub>:] Nun kann ein denkendes Wesen, bloß als ein solches betrachtet, nicht anders als Subjekt gedacht werden.

---

[C:] Also existiert es auch nur als ein solches, d. i. als Substanz (B 410-411, deductive form added).

The major premise (P<sub>1</sub>) states a formal rule for the application of the substance category to objects. Roughly, we can judge about e.g. a table that it is square, but not about a square that it is table. A table cannot be conceived of as a property predicated about something else. It is therefore a substance. The minor premise (P<sub>2</sub>) states that in conceiving of oneself as a thinking being, one cannot think of oneself otherwise than as subject. The basis for this claim is that in sentences of the form "I think ...," the word "I" always has the role of subject. Presumably, therefore, the substance category applies to oneself *qua* a thinking being; that is, one exists as a substance *qua* a thinking being, which

<sup>17</sup> All references to *Kritik der reinen Vernunft* are placed in brackets in the main text, e.g. A 23 or B 38, where A and B refer to the first and second edition, respectively, of that work. All references to other works of Kant are to the Weischedel edition (before the slash), and to the Akademie edition (after the slash).

<sup>18</sup> Cf. W.H. Walsh, *Kant's Criticism of metaphysics* (Edinburgh: Edinburgh University Press, 1975), p. 178.

is the conclusion C. I take it that the full doctrine consists of specifications of C that follow from the formal properties of "I think ...." For example, since the word "I," as used in the "I think ..." to express any thought that I might have, represents one and the same singular self – me – not a plurality of selves, the rational psychologist infers that the self is indivisible (B 407-408, cf. A 351ff). In this manner, rational psychology professes to prove that the person is a union of an immaterial, thinking substance with a physical body by mere analysis of the "I think ...."

Against this Kant objects that in the two premises of the argument "thinking" (*das Denken*) is taken in different senses: "im Obersatze, wie es auf ein Objekt überhaupt (mithin wie es in der Anschauung gegeben werden mag) geht; im Untersatze aber nur, wie es in der Beziehung aufs Selbstbewußtsein besteht, wobei also an gar kein Objekt gedacht wird, sondern nur die Beziehung auf Sich, als Subjekt, (als die Form des Denkens) vorgestellt wird." More succinctly:

Im ersteren wird von *Dingen* geredet, die nicht anders als Subjekte gedacht werden können; im zweiten aber nicht von *Dingen*, sondern vom *Denken* (indem man von allem Objekte abstrahiert), in welchem das Ich immer zum Subjekt des Bewußtseins dient; .... (B 411n).

Let us refer to these two senses of "thinking" as thinking in the objective and thinking in the subjective sense. Thinking in the objective sense is expressed in sentences of the form "Object A has property B," while the corresponding thinking in the subjective sense is expressed in sentences of the form "I think substance and accident." In the latter case, Kant explains, we do not apply the categories to ourselves (B 406-407). Consequently, I do not ascribe a property to myself, viz., that I have a thought (by analogy to the way in which I ascribe properties to objects). But if I do not *ascribe thoughts to myself* in pure apperception, then it seems to follow that the "I think ..." does not offer any justification for the correlated ontological idea that the thinking subject is a kind of *substance in which thoughts inhere*.

The key to the subjective sense of "thinking" is Kant's doctrine that thinking is an *activity*. As he says, the consciousness of myself in the representation "I" is a mere intellectual representation of the self-activity (*Selbsttätigkeit*) of a thinking subject (B 278). More clearly, pure self-consciousness is "ein Bewußtsein dessen, was der Mensch tut."<sup>19</sup> So when I use the first-person pronoun "I" in the "I think...", *I just express what I do*. The relevant actions here are the actions of the understanding. In so far as thinking in the objective sense yields knowledge (*Erkenntnis*), it is uttered in judgments passed according to the rules of the actions of the understanding through which we apply the categories to objects of experience. In the corresponding subjective sense, on the other hand, the thinking is uttered in the thinking being's verbal expression of the action of the understanding that he is performing.

This gives the following diagnosis of the rational psychologist's argument. In premise P<sub>1</sub> "thinking" is used in the *objective* sense of thinking about objects in

<sup>19</sup> Kant, *Anthropologie in pragmatischer Hinsicht*, BA 57/161.

general; in premise P<sub>2</sub>, on the other hand, "thinking" is used in the *subjective* sense of expressing what the thinking being *does* when thinking about objects. The rational psychologist has tacitly shifted from the objective to the subjective sense of "thinking." This implies that he cannot validly deduce from the use of the word "I" in the sentence "I think ..." that the thinking subject is a spiritual substance. His inferences are all paralogistic, i.e., logical fallacies in which the middle term is used ambiguously in the premises (B 411 and B 411n).

Formulating Kant's diagnosis of rational psychology in Wittgensteinian terms seems warranted, for he says: "Also ist der so berühmte psychologische Beweis lediglich auf der unteilbaren Einheit einer Vorstellung, die nur das Verbum in Ansehung einer Person dirigiert, gegründet" (A 355). One might thus say that the grammatical rules for the use of the term "I" in the sentence "I think ..." are mistakenly interpreted as evidence for *a priori* knowledge of the self. Analogous to the misunderstanding of psychological sentences in the first-person singular present tense that Wittgenstein diagnosed, the rational psychologist commits the error of inferring from the "Oberflächengrammatik"<sup>20</sup> of the "I think ..." as a subject-predicate judgment to the conclusion that the subject-term "I" refers to a substance to which thoughts are ascribed by means of predicate-terms of the form "think ...." There is, of course, a difference between expressing what I *do* and expressing a sensation that I *have*, but in both cases pointing to the expressive roots of first-person mentalistic sentences serves to expose the confusion of treating them as judgments about private objects – as introspectionist counterparts to sentences predicating about external objects.

But even though Kant clearly believes that pure apperception does not provide any *evidence* for the view that the self is a spiritual substance, one may wonder whether the agent that performs the actions expressed by the "I think ..." still is an immaterial subject for Kant, and thus that his main difference from Descartes is one of argument, not of conclusion. Kant's actions of the understanding are usually understood as private, mental acts or even as unconscious processes that somehow make experience possible, and these mentalistic or internalist interpretations (as I prefer to call them) naturally lead to the standard reading that the transcendental subject, the noumenal self, is some kind of hidden, underlying self (just as Hacker presupposes). This internalist reading seems further supported by Kant's claim that "Ich unterschiede meine eigene Existenz, als eines denkenden Wesens, von anderen Dingen außer mir (wozu auch mein Körper gehört) ist ebensowohl ein analytischer Satz"; all he seems to hold is that the Cartesian has given no conclusive argument that I really am the way in which I think of myself: "Aber ob dieses Bewußtsein meiner selbst ohne Dinge außer mir, dadurch mir Vorstellungen gegeben werden, gar möglich sei, und ich also bloß als denkend Wesen (ohne Mensch zu sein) existieren könne, weiß ich dadurch gar nicht" (B 409). However, it is crucial that Kant

<sup>20</sup> Wittgenstein, *Philosophische Untersuchungen*, § 664.

does not say that it is analytic that I distinguish my own existence, as a thinking being, *from my own body*, but rather that it is analytic that I distinguish my own existence, as a thinking being, from *other* things outside me: "denn *andere* Dinge sind solche, die ich als von mir *unterschieden* denke." Thus, Kant's claim here is *not* that I distinguish myself, as a thinking being, from my own body, but rather that my body belongs to the category of things outside me and that I distinguish myself, as thinking being, from *other* things outside me than my body. This is perfectly consistent with the view that I, as a thinking being, am a corporeal being. The interpretation is confirmed by a comment on the advantages of transcendental idealism regarding this question:

Auf solche Weise würde ebendasselbe, was in einer Beziehung körperlich heißt, in einer andern zugleich ein denkend Wesen sein, dessen Gedanken wir zwar nicht, aber doch die Zeichen derselben in der Erscheinung, anschauen können. Dadurch würde der Ausdruck wegfallen, daß nur Seelen (als besondere Arten von Substanzen) denken; es würde vielmehr wie gewöhnlich heißen, daß Menschen denken, d.i. ebendasselbe, was, als äußere Erscheinung, ausgedehnt ist, innerlich (an sich selbst) ein Subjekt sei, was nicht zusammengesetzt, sondern einfach ist und denkt (A 359-360).

If the thinking subject is *the very same* being as what, *qua* object of outer sense, is called a body, then clearly the thinking subject itself must be a corporeal being – a human being. I adopt two perspectives upon myself as thinking, human being: as appearance and as thing in itself. In the perspective in which I regard myself as appearance (as a member of the sensible world) I treat myself as an object on par with any other object in space. In the perspective in which I regard myself as a thing in itself (as a member of the intelligible world) I do not relate to myself as an object of experience (say, as an object of medicine or physics). But surely, my body may still in this perspective be part of the subjective conditions of experience (as indeed it already is in Kant's pre-critical paper about space, "Von dem Ersten Grunde des unterschiedes der Gegenden im Raume").

In order for Kant to arrive at this conclusion, it is essential that pure self-consciousness apply in the first instance to free, *embodied* actions. This would entail that the actions of the understanding basically are manifest (although it often suffices to carry out their mental counterparts). This claim is a basic tenet of the externalist interpretation (not to be confused with what is commonly labeled "externalism" in contemporary philosophy) that I have begun developing elsewhere.<sup>21</sup> The principal point of my interpretation is that all the formal conditions of experience that Kant attributes to the knowing subject are *public*: The pure intuitions of space and time are ostensive displays for the senses, such as the pointing which accompanies utterances like "Look over there!" and the

<sup>21</sup> Jens Saugstad, "Kant on Action and Knowledge," *Kant-Studien* 83 (1992), and "Sensibility, Space and Public Display," in Øfsti, Ulrich and Wyller (eds.), *Indexicality and Idealism. The Self in Philosophical Perspective*.

waving of a flag that accompanies "Now runner A crossed the finishing line"; the actions of the understanding are ordinary, manifest actions which make possible our knowledge of objects given to sense. There are two sorts of such actions: figurative and intellectual syntheses (B 151). To the first sort belong the manifest actions through which we produce paradigmatic, public representations of concepts, such as the construction of a circle on a piece of paper with a pair of compasses (B 287, B 299, B 742); to the second belong the manifest actions of comparing objects given to sense with the former self-produced representations of concepts, such as the comparison of a table with a yardstick in the measurement of length (B 456n).

There is indeed strong textual evidence for the externalist interpretation. For instance, Kant clearly holds that the successive synthesis attached to the concept of magnitude, at the fundamental level, consists in the manifest action of counting fingers or points on paper (B 15, B 299). A further confirmation of the externalist interpretation is that a judgment for Kant by its very nature is linguistic.<sup>22</sup> *Ich denke* is for the same reason a linguistic sentence, which is why Kant, as we have seen, characterizes it as *Satz* and *Text*. In this light, it seems that part of the point of the thesis in § 16 that the "I think ..." must be capable of accompanying all my representations is to show that sensible intuitions must either be *produced* through the agent's own, manifest intentional actions or that sensible intuitions, if they are empirical intuitions passively given to sense (B 41), must *match* such self-produced public representations. For in this way, the "I think ..." would indeed be capable of accompanying all my sensible intuitions, which would therefore all belong to my pure apperception, i.e., be intrinsically related to my action-consciousness as a free agent.

The externalist interpretation implies that Kant's view of the person cannot be Cartesian. For my consciousness of what I *do* in actions of the understanding, externalistically interpreted, is clearly not a consciousness of a *spiritual* substance. When I construct a circle on paper or count the coins in my pocket – such actions are thoughts no less than their mental counterparts<sup>23</sup> – the relevant specifications of "I think ..." ("I construct a circle" and "I count") express what I do as a free, *embodied* agent.

Kant supports his case by appealing to his theory of knowledge. In order to have the kind of knowledge which the rational psychologist professes to have, two elements must be available to him: a concept and an intuition of the self (B 74-75, B 158 and B 158n). In knowledge of the external world as well as in psychological knowledge of inner life, intuitions are *sensible*. But this will not do if one attempts to demonstrate the *spirituality* of the soul. A *sensible* intuition would immediately transform rational psychology into empirical

<sup>22</sup> Kant, *Über eine Entdeckung*, BA 16n/193-194n and *Logik*, A 170/109; cf. *Anthropologie*, BA 48-49/155 and BA 55-56/159.

<sup>23</sup> Cf. Saugstad, "Kant on Action and Knowledge," section II, esp. pp. 386-387, and section IV.

psychology. But we do not have, according to Kant's theory of knowledge, the ability to give content to our concepts through non-sensible intuitions. So there is no intuition to which the rational psychologist can appeal in order to make sense of his epistemic claims. All the rational psychologist has to go on is the "I think ...." But actually the "I" of pure apperception does not even function as a concept; it is the emptiest and poorest of all our representations, Kant famously says (B 404-413; cf. A 382).

#### 4. In what sense does "I" refer?

In an earlier paper I said that "Although there are innocuous uses of 'refer' on which one may say that 'I' refers, the idea that it is an indexical encourages the confusion of treating it as an ordinary referring term."<sup>24</sup> I shall now try to make clearer what the illegitimate and legitimate ways of construing the reference of the first-person pronoun are.

We have seen that for Kant the first-person pronoun as used in the *Text* "I think ..." is not a concept, "sondern ein bloßes Bewußtsein, das alle Begriffe begleitet." This consciousness in itself (*Bewußtsein an sich*) is not a representation "die ein besonderes Objekt unterscheidet" (B 404, cf. A 382). Presumably, Kant is saying that the first-person pronoun as used to express what I do in so far as I engage in thinking does not distinguish a particular person. Thus, he seems to agree with Wittgenstein's view, in Hacker's rendering, that "When I use the first-person pronoun, I employ no principle of differentiation to select one person from among others."<sup>25</sup>

But at least once, Kant refers to the "I think ..." as a judgment: "Dieses ist der Begriff, oder, wenn man lieber will, das Urteil: Ich denke" (B 399). And he repeatedly takes the "I" as a representation. According to W.H. Walsh, he then encourages the very mistake he tried to avoid.<sup>26</sup> However, it follows from the critique of rational psychology that the "I think ..." cannot be a judgment in the sense in which "The table is square" is a judgment, i.e., a *theoretical* judgment. In so far as it is a judgment, we shall shortly see, it can only be a *practical* judgment. This leads to an alternative view on the referential nature of the "I" according to which it is legitimately called a representation of the thinking subject (not a *name* of representation, as Walsh, having missed the linguistic nature of thoughts for Kant, says).

Before pursuing this idea, a contemporary confusion related to the idea that the "I think ..." is a judgment should be exposed. Suppose, for the sake of argument, that Kant only means to say that it is *formally* a judgment, that is, a sentence that can be treated as a judgment from the point of view of deductive

<sup>24</sup> Saugstad, "Sensibility, Space and Public Display," p. 141.

<sup>25</sup> Hacker, *Wittgenstein. Meaning and Mind*, p. 223.

<sup>26</sup> Walsh, *Kant's Criticism of metaphysics*, p. 179.

logic. In a similar vein, "I have toothache" can be treated as a judgment from the logical point of view. The first-person pronoun can accordingly be treated as a referential term (which needs to be substituted by less ambiguous terms) for logical purposes. However, many contemporary philosophers believe that this fact proves the referential nature of the first-person pronoun. Louise Röska-Hardy puts what we might call the argument from formal logic in this way: "As the grammatical subject in first-person mentalistic ascriptions, 'I' occupies a position in which we can substitute other singular referring expressions, e.g. names and descriptions, as well as quantifiers, e.g. 'I have a toothache' entails 'Someone has a toothache'. In addition, the statement expressed by 'I have a toothache' resembles 'He has a toothache' or 'Jones has a toothache' in that they all contradict 'Nobody has a toothache' and entail 'Someone has a toothache', as Shoemaker points out. For these reasons 'I' does seem to function as a referring expression in the use as subject, even though it does not fit the Fregean model of reference for singular terms."<sup>27</sup> Such facts about formal logic cannot, however, be taken as conclusive evidence for the view that the first-person pronoun is a referring term. For as Hacker observes, one can make inferences from first-person psychological sentences that are *mere* verbal exclamations, and these utterances are surely the least plausible candidates for statements with a referential function of "I." The danger of the argument from formal logic is that it encourages the confusion of construing the reference of "I" after "he" – and thus of smuggling the third-person perspective in the back door. The same goes for the related argument from veridical symmetry, i.e., from the fact that "wenn der Satz 'ich  $\varphi$ ' in meinem Munde wahr ist, ist der Satz 'er  $\varphi$ ' in dem Munde eines Sprechers wahr, der sich mit 'er' auf mich bezieht."<sup>28</sup> Again I believe that Wittgenstein, in Hacker's paraphrase, has offered a decisive counterargument: "One can indeed say that in the mouth of a particular person 'I' refers to that person *in the sense that* if he rightly says 'I  $\varphi$ ' then others can say truly 'N.N.  $\varphi$ s'. But this transformation from first to third person is, so to speak, *for others*."<sup>29</sup> Hence, while encouraging the assimilation of the function of the first-person pronoun to inappropriate paradigms of reference, the arguments from formal logic and veridical symmetry do not prove that it actually is a referring term.

At the end of the first section I doubted Hacker's claim that Wittgenstein would agree that "I" can be said to refer to the speaker, if only one carefully avoids the grammatical misunderstandings so saying typically leads to.<sup>30</sup> Recall that Wittgenstein does not even allow that it *stands for* anything. This is connected to his view that the first-person pronoun can always be *eliminated*.

<sup>27</sup> See Louise Röska-Hardy, "Idealism and the 'I' of Self-Ascriptions," p. 59; cf. p. 63.

<sup>28</sup> Anton Friedrich Koch, "Der Selbstverlust des Begriffs," in Øfsti, Ulrich and Wyller (eds.), *Indexicality and Idealism*, p. 70.

<sup>29</sup> Hacker, *Wittgenstein. Meaning and Mind*, p. 225.

<sup>30</sup> Hacker, *Wittgenstein. Meaning and Mind*, p. 228.

Hacker assumes that he held this view only in his middle period,<sup>31</sup> but John Cook seems right that he persisted in it throughout the whole later period.<sup>32</sup> For Kant, of course, it would be impossible to eliminate the "I" as the "I think ..." is a necessary feature of human rationality. The eliminativist view may explain why Wittgenstein did not develop a positive account of the mode of reference of the first-person pronoun. But it also suggests that he lacked the right perspective for giving one. The lacking perspective, I want to propose, is Kant's doctrine of pure apperception as "der höchste Punkt" of all uses of the understanding (B 134n) and as the factor ultimately responsible for the unity of theoretical and practical reason.<sup>33</sup>

In what follows, I shall trace the expressive roots of the "I think ..." in order to see more clearly in what sense the first-person pronoun refers. But first a word about interpretation. Kant is famous for his claim that transcendental philosophy is concerned with the *quid juris*, the question of the legitimacy of something, not the *quid facti*, the question of facts, that is, the explanation of how we acquired something (B 116). I take it, however, that the *quid juris* does not exclude all accounts of acquisition; reflection – even speculation – on the formation of concepts may indeed shed light upon their status and legitimacy.

We have seen that pure apperception is a consciousness of what man *does*. But in general, the ability to express in language what one does is a component part of the ability to perform intentional actions. Wittgenstein's comment about nursery rhymes gives a hint as to how the connection between intentional action and its linguistic expression could have been set up.<sup>34</sup> In a Norwegian children's game resembling ring-a-ring o' roses, the participants are induced to jump, dance, and sit down as all sing in chorus "And then we all jump," and so on. In this way children may learn to answer "We jump" when asked what they are doing and, being taught that "we" is correctly used when they do things together with others, they may eventually learn to answer "I jump" as an expression of their own action (whether done alone or in company). The example points to the expressive beginnings of the kind of sentences through which a person expresses what he does and thinks. Since to think is to act intentionally, it follows that "I think ..." is a version of "I act ...." More specifically, "I think ..." is the linguistic form of the way in which we express what we do when we perform those free, intentional actions that make thinking about something possible. This subclass of intentional actions comprises all the rule-following actions defining our ability to pass judgment (which Kant calls the actions of the understanding). In other words, "I think ..." is the linguistic component part of the ability to perform the rule-following actions that comprise

<sup>31</sup> Hacker, *Wittgenstein. Meaning and Mind*, pp. 216ff.

<sup>32</sup> Cook, *Wittgenstein, Empiricism, and Language*, esp. pp. 28-31, 64-68.

<sup>33</sup> Kant, *Grundlegung zur Metaphysik der Sitten*, BA XIV/391.

<sup>34</sup> Cf. Wittgenstein, *Philosophische Untersuchungen*, § 7.



the activity of thinking. Thus, just as "I jump" is a component part of my ability to jump intentionally, "I think ..." is a component part of my ability to think.

Now, expressing what I do involves representing myself as the agent responsible for the action done, and this also applies to the actions of the understanding, the kind of actions that constitute thinking. Hence, the "I think ..." represents imputation, the primary relation of responsibility between the thinking subject and his thought-actions. When Kant says that the "I" as used in the "I think ..." represents the self-activity of the thinking being, he thus says that it refers to the human being as the responsible agent for his thinking. In other words, "I think ..." is a *proto-practical* judgment that makes theoretical judgment possible. But in first-person practical judgments the agent does not identify himself for himself; rather, he expresses what he does. If, for example, a teacher irritatedly asks class, "Who is whispering?", a pupil might admit guilt by saying "I do." The point of the answer is not to identify himself to himself as the pupil whispering, but to let the *teacher* identify him as such. In Hacker's paraphrase of Wittgenstein's position in *The Blue Book* (quoted in the first section), "Saying 'I ...' is more like raising my hand to draw attention to myself than it is like pointing to someone."<sup>35</sup> This, however, does not rule out that "I" is a referring term of a sort. To be sure, it is not "a super-referring term guaranteed against reference-failure" in the sense that it cannot fail to pick out the right person<sup>36</sup> – not because it can fail, but simply because it does not at all serve to pick out the agent from among several persons for himself. However, Wittgenstein goes too far when denying that "I" *stands* for something. By verbally signaling responsibility for the disturbance in class by means of the sentence "I do," the most plausible view is that the first-person pronoun *represents* the pupil as agent – analogous to how a lawyer represents a client. In this way, the first-person pronoun may be said to *stand for* the acting subject in the language of responsibility.

By the same token, I do not pick out for myself through the "I think ..." that I am the one who thinks. The "I" refers to the thinking subject merely in the sense that the first-person pronoun is that part of the verbal expression of "was der Mensch *ich*" which represents – stands for – the human being *qua* responsible agent for the exercise of the actions of the understanding. This is, admittedly, a very modest notion of reference, but it is all that we can possibly have in a pure first-person perspective. Still, it suffices for the purposes of empirical self-knowledge.

<sup>35</sup> Hacker, *Wittgenstein. Meaning and Mind*, p. 224.

<sup>36</sup> Hacker, *Wittgenstein. Meaning and Mind*, p. 228.

## 5. Inner sense

We have seen that self-knowledge – what Kant refers to as knowledge of the soul and its inner states (B 37) – requires that pure apperception is added concepts and their corresponding inner intuitions, i.e., the kind of sensible intuitions that pertain to inner sense, whose form is time.

It is commonly believed among philosophers that Kant's doctrine of inner sense and psychological experience cannot be reconciled with Wittgenstein's position on the public nature of psychological concepts. It is indeed natural to interpret Kant's inner intuitions as introspective "glances" into the soul and its inner states, or perhaps as those private objects with which we are immediately acquainted in introspection. Thus, Hacker lumps Kant together with Locke in the belief that "'Internal sense' or 'inner sense' is the source of our knowledge of the subjective objects in our private inner world, and it gives us an *immediate, non-referential acquaintance* with them, which contrasts with the inferential, perception-mediated knowledge that we strive to attain of objects in the public domain."<sup>37</sup>

Hacker seems not to take into account the fact that Kant in "Widerlegung des Idealismus" adopts the position that *inner experience is mediated by external objects* (B 274-279). Kant's proof makes evident that he is opposed to the very Cartesian conception of introspection that Wittgenstein attacks. Of course, Kant did not deny that we can and do scrutinize our own thoughts and feelings; so doing, while extremely difficult, is indeed pivotal to moral life.<sup>38</sup> But he cannot be charged with the view that the concepts appealed to in such scrutinizing have a private content formed introspectively. For as he says in the comment on *Widerlegung* in the preface to the second edition of *Kritik der reinen Vernunft*: "das Dasein der Dinge außer uns (von denen wir doch den ganzen Stoff zu Erkenntnissen selbst für unseren inneren Sinn her haben)" (B XXXIXn).

In this section I shall argue that there is in Kant a counterpart to Wittgenstein's private language-argument – not *Widerlegung*, but his doctrine of inner sense. Properly understood, Kant has here anticipated Wittgenstein's thesis: "Ein 'innerer Vorgang' bedarf äußerer Kriterien."<sup>39</sup> Actually, I believe Kant's critical reminders (cf. B 45) about time as the form of inner sense provide us with a *more effective argument* to this externalist (in my sense of the term) conclusion than Wittgenstein's various arguments against the possibility of establishing privately a rule for the use of "S" as a name for sensations.<sup>40</sup> Kant has admittedly left us with a mere sketch, the completion of which unavoidably takes us beyond what he explicitly says. I want to argue, though, that the

<sup>37</sup> Hacker, *Wittgenstein. Meaning and Mind*, pp. 17-18; cf. p. 210.

<sup>38</sup> Kant, *Die Metaphysik der Sitten*, Tugendlehre, § 14; *Anthropologie*, §12/§14.

<sup>39</sup> Wittgenstein, *Philosophische Untersuchungen*, § 580.

<sup>40</sup> Wittgenstein, *Philosophische Untersuchungen*, §§ 256-270.

externalist interpretation of the doctrine of inner sense is a logical development of the central doctrines of Kant's epistemology.

For Kant, time is the "great externalizer." In *Widerlegung* he appeals to the idea that the determination of anything in time, including my own existence, presupposes something *permanent* against which to measure change, and that this permanent is only found in material objects. In the transcendental aesthetic he reminds us that although time is "keine Bestimmung äußerer Erscheinungen," but rather determines "das Verhältnis der Vorstellungen in unserem inneren Zustande," we compensate for the lack of shape in inner life by representing time in analogy to space "und stellen die Zeitfolge durch eine ins Unendliche fortgehende Linie vor" (B 50). (Of course, this does not reduce time to space, for the temporal relations represented by means of the line are not themselves spatial.)

As time is the form of inner sense, what holds for time holds also for inner states. Thus, in "Allgemeine Anmerkung zum System der Grundsätze," Kant argues:

um uns nachher selbst innere Veränderungen denkbar zu machen, müssen wir die Zeit, als die Form des inneren Sinnes, figürlich durch eine Linie, und die innere Veränderung durch das Ziehen dieser Linie (Bewegung), mithin die sukzessive Existenz unser selbst in verschiedenem Zustande durch äußere Anschauung uns faßlich machen, ... (B 292).

In the comment about inner sense in § 24 of the transcendental deduction of the categories, Kant argues in addition that the determination of inner change depends on the changes that external *objects* display. And from these facts about time he argues that I know myself "gleich anderen Phänomenen, nicht wie ich vor dem Verstande bin, sondern wie ich mir erscheine":

Daß es aber doch wirklich so sein müsse, kann, wenn man den Raum für eine bloße reine Form der Erscheinungen äußerer Sinne gelten läßt, dadurch klar dargetan werden, daß wir die Zeit, die doch gar kein Gegenstand äußerer Anschauung ist, uns nicht anders vorstellig machen können, als unter dem Bilde einer Linie, sofern wir sie ziehen, ohne welche Darstellungsart wir die Einheit ihrer Abmessung gar nicht erkennen könnten, imgleichen daß wir die Bestimmung der Zeitlänge, oder auch der Zeitstellen für alle innere Wahrnehmungen immer von dem hernehmen müssen, was uns äußere Dinge Veränderliches darstellen, folglich die Bestimmungen des inneren Sinnes gerade auf dieselbe Art als Erscheinungen in der Zeit ordnen müssen, wie wir die der äußeren Sinne im Raume ordnen; mithin, wenn wir von den letzteren einräumen, daß wir dadurch Objekte nur sofern erkennen, als wir äußerlich affiziert werden, wir auch vom inneren Sinne zugestehen müssen, daß wir dadurch uns selbst nur so anschauen, wie wir innerlich von uns selbst affiziert werden, d.i. was die innere Anschauung betrifft, unser eigenes Subjekt nur als Erscheinung, nicht aber nach dem, was es an sich selbst ist, erkennen (B 156)

Kant clearly believes that there is a parallel between outer and inner sense: both provide knowledge of appearances, not of things in themselves. However, this parallel should not be understood as a variation of the traditional view that

self-consciousness, in Paul Churchland's phrase, "is just a species of perception: *self-perception*. ... It is just directed internally rather than externally."<sup>41</sup> Unfortunately, Kant-commentators tend to take for granted that Kant shares such a view. Thus, Henry Allison believes that "the manifest difference between outer and inner intuition stems from the fact that what we outwardly intuit are appearances with spatial forms and properties, while what we inwardly intuit is the appearance of these very appearances, along with mental states such as feelings, in consciousness."<sup>42</sup> However, internalist views like this do not square well with Kant's critical reminders about the dependency of the representation of time upon objects in space. If we must represent the change, duration and points of time of inner perceptions by means of *external objects in space*, it follows that the *inner intuitions* corresponding to our psychological concepts about the inner cannot themselves be something inner, but must rather be understood as *public representations of the inner*. Moreover, as the phrase "unser eigenes Subjekt nur als Erscheinung" itself naturally suggests, these public representations must above all comprise *the bodily expressions of our inner states*. Thus, the train of thoughts in the passage supports the externalist interpretation, according to which an appearance is always an external object with a representing function.<sup>43</sup>

with

We can illustrate the doctrine with the feeling of pain. Pains endure in time and typically increase to a peak of intensity before they fade away. Since we represent time by means of space, we could not have been aware of the temporal properties of pain unless they too were represented in space, by means of external objects. Moreover, these external objects must involve human bodies, because pain-behavior is the *appearance* of the subject's feeling of pain.

Admittedly, Kant never explicitly says that the inner intuitions corresponding to psychological concepts are behavioral. Perhaps the reason is that the exact nature of inner intuitions is an empirical-anthropological matter that lies outside the reach of transcendental philosophy, which is concerned merely with the *a priori* form of experience. Unfortunately, however, there is no clarification of the matter in Kant's treatment of inner sense in the *Anthropologie*, albeit some comments about physiognomy in the second part, "von der Art, das Innere des Menschen aus dem Äusseren zu erkennen," for instance about "der Mienen, welche unvorsetzlich das Innere verraten," may suggest that Kant was aware of the internal relation between inner states and their external expressions.<sup>44</sup>

<sup>41</sup> Paul Churchland, *Matter and Consciousness*, second edition (Cambridge, Massachusetts: The MIT Press, 1988), p. 74.

<sup>42</sup> Henry Allison, *Kant's Transcendental Idealism. An Interpretation and Defense* (New Haven: Yale University Press, 1983), p. 258.

<sup>43</sup> Saugstad, "Kant on Action and Knowledge," p. 398; "Sensibility, Space and Public Display," p. 132.

<sup>44</sup> Kant, *Anthropologie*, B 278/301.

Nor is there any trace in Kant of Wittgenstein's suggestion that the verbal expressions of sensations, such as pain, replace the natural expressions. Though this could explain why Kant did not explicitly develop the behavioral part of his theory-sketch, a Kantian does not strictly need Wittgenstein's suggestion to do so. Wittgenstein's suggestion is intended to free us from the picture of private naming that holds us captive, but Kant's theory of pure apperception already contains the requisite insights to avoid this trap.

I suggested (in section 3) that the point of the necessary accompanyability of the "I think ..." is to show that all concepts must be represented publicly. Concepts serve as rules for the spontaneous actions of the understanding, which include the actions through which we produce sensible intuitions (the figurative syntheses of productive imagination). In geometry, productive imagination provides concepts with meaning "durch die Konstruktion der Gestalt, welche eine den Sinnen gegenwärtige (obzwar a priori zustande gebrachte) Erscheinung ist," in arithmetic by means of "den Fingern, den Korallen des Rechenbretts, oder den Strichen und Punkten, die vor Augen gestellt werden" (B 299). Psychological concepts about inner states are not *a priori*, because the inner states themselves are given *a posteriori*, but the necessary accompanyability of the "I think ..." entails that productive imagination must have a similar role also in their case. But what could the spontaneous actions of productive imagination possibly amount to in the case of psychological concepts about inner states? Pain-behavior and the other natural expressions of inner states are *not* spontaneous in Kant's transcendental sense.

I think there is an obvious reply. Inspired by Wittgenstein's question about the dog,<sup>45</sup> I want to suggest that the figurative syntheses through which we represent psychological concepts about the inner are those free, intentional actions by which we *simulate* the natural expressions of inner states, e.g. the simulation of pain-behavior. Of course, psychological concepts, being empirical, cannot be represented in sensible intuition "ohne das Muster dazu aus irgendeiner Erfahrung geborgt zu haben" (B 741), as indeed mathematical concepts are. Thus, Kant's general thesis that we display concepts in their corresponding sensible intuition through the figurative syntheses of productive imagination logically implies that we display psychological concepts about inner states through *theatrical acts*. On the proposed logical development of Kant's doctrine of inner sense, then, the ability to simulate the natural behavioral expressions of inner states is required for having the corresponding psychological concepts. Dogs, for instance, do not have the concept of pain because they cannot simulate pain (not the other way around).

We have here a Kantian argument for the thesis that the criteria for psychological concepts about inner states are primarily behavioral: Pain-behavior serves as the inner intuition of the concept 'pain', a smile as an inner intuition of the concept 'joy', and so on. I want to stress, however, that inner intuitions

<sup>45</sup> Wittgenstein, *Philosophische Untersuchungen*, § 250.

are not primarily the displays we produce by means of our simulations, but above all the displays that occur by means of the natural behavioral expressions that we simulate. The point is that the simulation of pain enables natural pain-behavior to represent 'pain' just as the construction of a circle enables a plate to represent the concept 'round' (B 176).

Let me now propose an interpretation of the notoriously enigmatic doctrine of internal self-affection. Some passages seem to suggest that we affect inner sense *at will* by actively producing the content of inner sense through figurative synthesis (B 154) and acts of attention (B 156n), whereas the definition of inner sense in the *Anthropologie* as "was der Mensch ... *leidet*, wiefern er durch sein eigenes Gedankenspiel affiziert wird,"<sup>46</sup> and related passages (cf. B 405) suggest that the mind is rather affected by its own, changing inner states – whether this "play" is actively produced by the subject or not. Adapting Allison's insights to the externalist interpretation, voluntary action is not the *source* of the content of inner sense, but that which allows us to represent inner states publicly in inner intuition.<sup>47</sup>

I believe this interpretation is confirmed by the three words "die Art, wie," which typically occur in Kant's characterizations of sensibility (cf. B 33). In § 8, "Allgemeine Anmerkungen zur transzendentalen Ästhetik," Kant stresses that time, just because it only contains relations (succession, coexistence and endurance), characterizes "die Art, wie das Gemüt durch eigene Tätigkeit, nämlich dieses Setzen ihrer Vorstellung, mithin durch sich selbst affiziert wird, d.i. ein inneren Sinn seiner Form nach" (B 67-68). It is plausible that the "wie" of the "wie wir innerlich von uns selbst affiziert werden" at the end of the passage quoted from § 24 is an elliptic substitute for "die Art, wie." This entails that Kant's reminders about the fact that time must be represented by means of objects in space is supposed to characterize *the way* in which inner states must be represented if they are to be known. And as time, the form of inner sense, must be represented publicly, the same requirement must also apply to the *contents* of inner sense. Moreover, just as *the way* in which external objects are represented in outer sense cannot be exhaustively characterized only by appeal to space as a pure intuition (that is, the display by means of the ostensive actions that accompany deictic spatial terms like "there," "next to" and "behind"),<sup>48</sup> but must also be characterized by the figurative synthesis of productive imagination, *the way* in which inner states are represented in inner sense also includes the figurative synthesis of productive imagination. Hence, while the appeal to time entails that the inner intuitions representing inner states must involve *some* external objects, the figurative synthesis of productive imagination entails that these inner intuitions are first of all comprised of the natural behavioral expressions of inner states.

<sup>46</sup> Kant, *Anthropologie*, BA 57/161.

<sup>47</sup> Allison, *Kant's Transcendental Idealism*, pp. 266-269.

<sup>48</sup> Saugstad, "Sensibility, Space and Public Display," pp. 132ff.

Internal self-affection can now be understood as follows: The mind (*das Gemüt*) is a unity with different capacities. In the capacity for inner states the mind affects *itself* in the capacity for inner intuitions such that there arise inner intuitions in accordance with the way in which (a) time is represented by means of space and (b) the expressive behavior of the inner is simulated by the subject at will in productive imagination.

Bringing the intuitive and conceptual elements together, it seems like a plausible thesis that the psychological judgment that *another* being (person or animal) is in some particular kind of inner state rests upon our ability to compare that being's expressive behavior with the paradigmatic representations that serve as inner intuitions of the relevant psychological concepts. While the evidence for our psychological judgment that another is in pain also includes symptoms,<sup>49</sup> such as stress-reactions involving changes in blood pressure and heart beat, the pain-behavior that can be simulated is the criterion that establishes the conceptual link between the inner state of pain and the outer states of the body, the link that enables us to pass the judgment.

A more difficult question is what this theory would imply for first-person psychological sentences. Kant was quite pessimistic about the prospects of self-knowledge, especially about the motives and maxims of our actions:

ja selbst in die innere Erfahrung des Menschen an ihm selbst läßt ihn die Tiefen seines Herzens nicht so durchschauen, daß er von dem Grunde seiner Maximen, zu denen er sich bekennt, und von ihrer Lauterkeit und Festigkeit durch Selbstbeobachtung ganz sichere Kenntnis erlangen könnte.<sup>50</sup>

But he never seems to question that psychological sentences in the first-person single present tense can, under favorable conditions, qualify as *judgments* expressing *knowledge*, even though he clearly did not subscribe to the Cartesian view that self-knowledge is immediate. Wittgenstein, as we have seen, believes that there can be no self-knowledge of pain and other sensations (as opposed to emotions and conations) because there is no room here for doubt in the first-person single present tense.

To defend Kant against Wittgenstein in this respect, one could (1) argue that sensations like pain are excluded from the doctrine of inner sense and exempt from self-knowledge, (2) question Wittgenstein's link between knowledge and the possibility of doubt, or (3) maintain that there is room for doubt and error even in the case of first-person utterances about sensations. Alternative (1) seems ruled out. In spite of apparently conflicting statements about the content of inner sense,<sup>51</sup> I think it cannot be maintained that the sensation of pain is excluded from inner sense.<sup>52</sup> All representations belong "als Modifikationen

<sup>49</sup> Cf. Wittgenstein, *The Blue and Brown Books*, pp. 24-24.

<sup>50</sup> Kant, *Die Religion innerhalb der Grenzen der bloßen Vernunft*, B 78/63; cf. *Grundlegung*, BA 26/407.

<sup>51</sup> Cf. e.g. B 66 with the above quotation from *Religion*.

<sup>52</sup> Cf. Allison, *Kant's Transcendental Idealism*, pp. 258-259.

des Gemüts zum inneren Sinn" (A 98), "den Inbegriff aller Vorstellungen" (B 220), and sensations are subjective representations (B 376). Alternative (2) may not be so easily ruled out, but if we assume that knowledge presupposes justification, it seems that it is appropriate to speak of knowledge only where doubt can be removed. This leaves us with alternative (3). Wittgenstein's argument for the rejection of self-knowledge about sensations seems like a remnant of the Cartesian notion of infallible first-person authority – non-epistemically interpreted. However, in the literature there are strong arguments that substantial error cannot be ruled out even in the case of one's own pain.<sup>53</sup> Thus, there does not seem to be such a sharp, qualitative line of demarcation between sensations on the one hand and emotions and motives on the other as Wittgenstein assumes.

It is a commonplace that in the case of emotions such as love we must often scrutinize our own behavior to determine what feeling we really have. Walsh believes that "Kant shows no awareness of the possibility that I might learn something about myself by reflecting on my own behavior."<sup>54</sup> Be this as it may, such reflection must be involved in the non-Cartesian notion of introspection (*Selbstbeobachtung*) to which Kant's doctrine of inner sense logically leads. For since the criteria for the use of psychological concepts are public, knowledge about my own inner states requires the possibility of checking my utterances through the comparison of my own behavior against the paradigmatic patterns of behavior that represent the relevant psychological concepts in inner intuition. But what about sensations like pain? Of course, we do not have to observe our own behavior and bodily reactions – say, by looking in the mirror – in order to say truly "I have pain." Still, the correct use of the concept 'pain' here must be consistent with my behavior.<sup>55</sup> The upshot is that the correct application of psychological concepts – about sensations no less than emotions – to oneself presupposes the possibility of checking one's own behavior against the paradigmatic pattern of behavior that serves as the inner intuition of the relevant concept. Since we can be mistaken, the correct use will normally be indicative of knowledge. Kant's transcendental idealism of inner sense now amounts to the thesis that we do not have immediate epistemic access to our own inner states, but know them only through the mediation of bodily expressions that we have endowed with a representing function in the way described.

Self-knowledge is not a *pure* first-person perspective, but the kind of mixed first-person perspective in which I, *qua* transcendental subject (embodied rational agent), judge about my inner states *qua* empirical subject (embodied

<sup>53</sup> Cf. Churchland, *Matter and Consciousness*, pp. 76-79; Jonathan Dancy, *Introduction to Contemporary Epistemology* (Oxford: Blackwell Publishers, 1985), pp. 58-61.

<sup>54</sup> Walsh, *Kant's Criticism of metaphysics*, p. 185.

<sup>55</sup> Malcolm Budd, *Wittgenstein's Philosophy of Psychology* (London: Routledge, 1989), p. 61.



possessor of inner states). Such self-knowledge is always *mediated* by the possible comparison of physical objects. Just as the knowledge that my desk is 2 meters long depends on my ability to compare the table (the outer intuition corresponding to the concept 'table') with a meter stick (the outer intuition corresponding to '... meters long'),<sup>56</sup> the psychological knowledge that I have pain depends on my ability to compare my body (the empirical intuition providing the permanent (*das Beharrliche*) for the subject-term "I" in "I have pain" – cf. A 349-350) with paradigmatic pain-behavior (the inner intuition serving as criterion for the application of the predicate-term "pain" and representing the non-spatial, inner state of pain in public).

To sum up, self-knowledge presupposes the ability to relate to oneself quasi in the third-person. While this thesis has some resemblance with Peter Strawson's theory "that it is a necessary condition of one's ascribing states of consciousness ... to oneself ... that one should also ascribe them, or be prepared to ascribe them, to others who are not oneself,"<sup>57</sup> the Kantian theory is rather that self-knowledge is possible because the use of psychological concepts in the first-person is mediated by the possibility of taking an objective stance towards *oneself*.

## 6. The metaphysical nature of the self

Construing "I think ..." as a proto-practical judgment has important consequences for ethics, metaphysics and anthropology. Whereas Wittgenstein *may* have believed that his critique of the philosophical misunderstandings of the grammar of the word "I" undermines traditional occidental metaphysical conceptions of the person as more than an biological being, Kant argues that there is evidence *in our practical use of reason* for a distinctly metaphysical conception of the person as a being endowed with a transcendently free will. I believe this conception of the person partly follows from the "grammar" of the use of "I" in the "I think ...." For if responsibility cannot be given a naturalistic compatibilist interpretation, according to which actions are mere events caused by the mental states of the agent, but rather implies, as Kant argues, that we have a capacity for absolutely spontaneous action, then we can infer from the practical interpretation of the "I think ..." that the thinking subject is intelligible in Kant's sense of the term (B 574).

Hacker is right that Kant's metaphysical view of the self depends on the thesis that "the moral 'self' must be free, subject to the laws of practical reason." But he is wrong that this moral 'self' is an underlying, immaterial entity and also wrong that Kant's view ultimately stems from a mistaken view of the first-person pronoun. Regarding the first error, the person as a noumenon (which in epistemic contexts is called the transcendental subject)

<sup>56</sup> Cf. Saugstad, "Kant on Action and Knowledge," p. 388.

<sup>57</sup> Cf. Peter Strawson, *Individuals. An Essay in Descriptive Metaphysics* (London: Routledge, 1959), p. 99.

is, as we have seen, not immaterial for Kant, but the corporeal human being as endowed with the capacity for absolutely spontaneous but nevertheless manifest actions, such as arising from the chair (B 478). Indeed, it seems to me that only the externalist interpretation of Kant's doctrines of pure and empirical self-consciousness makes sense of his claim "das Ich, der ich denke, von dem Ich, das sich selbst anschaut, unterschieden ... und doch mit diesem letzteren als dasselbe Subjekt einerlei sei" (B 155). For as I hope now has become clear, these "selves" are aspects of one and the same corporeal human being – in the first respect with regard to "was der Mensch *tut*," in the second with regard to "was er *leidet*" (cf. B 153). With regard to the second error, although Kant maintained that "I" represents the agent, and thus that it is a referring term of a practical sort, it is not this view on the first-person pronoun, but rather his position on moral responsibility which is responsible for his metaphysical view of the person as standing partly apart from nature in virtue of the transcendently free will. Hacker would perhaps attempt to dismantle this position through some kind of Wittgensteinian therapy. But although it seems that Wittgenstein favored *some* sort of compatibilism, it is far from clear that his views in "Lectures on Freedom of the Will" are incompatible with Kant's non-naturalistic, dualistic kind of compatibilism.<sup>58</sup> Indeed, it is hardly the case that a naturalistic account of freedom and moral responsibility states only "was Jeder ihr zugibt."<sup>59</sup>

<sup>58</sup> Wittgenstein, *Philosophical Occasions 1912-1951*, J. Klagge and A. Nordmann (eds.) (Indianapolis: Hackett Publishing Company, 1993), pp. 429-444.

<sup>59</sup> Wittgenstein, *Philosophische Untersuchungen*, § 599.