

## INTRODUCTION

---

### From Inner Experience to the Self-Formation of Psychological Persons

#### i.1 Two Theses

As the pre-eminent Enlightenment philosopher, Immanuel Kant is famous for emphasizing that each and every one of us is called to “make use of one’s own understanding without direction from another” (*Enlightenment* 8:35). We are all called to make up our own minds, independently from the external constraints imposed on us by others. In the face of this Enlightenment calling, much of Kant’s philosophy, then, reads as a manual for how to employ one’s mental faculties in the proper way – faculties that are supposed to be universally realized by all human beings. Given his focus on a universal conception of the human mind, Kant tells us surprisingly little about what makes us the unique individual persons we are and how we come to know ourselves as such.

This book explores Kant’s distinctive account of *psychological personhood* by unfolding, in accordance with the tenets of his Critical Philosophy, his account of *empirical self-knowledge* as the knowledge that one has of oneself as a unique psychological person. A central role is played by the capacity to judge one’s own psychological features, that is, the capacity for what Kant calls *inner experience*. Primarily, inner experience concerns a person’s conscious mental states, such as occurrent sensory perceptions, thoughts, memories, imaginations, feelings, and desires. Moreover, inner experience also concerns general psychological properties such as personality traits and character dispositions, standing attitudes, commitments, and values. Although inner experience has been neglected in the contemporary literature on Kant, I argue that, for Kant, it is a primary means by which persons not only gain knowledge of a range of psychological phenomena that make up their mental lives, but also determine who they are. So in this book I defend two central theses.

First, for Kant, *inner experience is empirical cognition of oneself, not as a mere object, but as a psychological person.*<sup>1</sup> On my reading, Kant conceives of inner

<sup>1</sup> Throughout this work, I employ the term *experience* in this Kantian sense, which I introduce in i.2. Moreover, I use the terms *capacity* and *faculty* interchangeably, if not stated otherwise.

experience as analogous to the experience of mind-independent objects in some respects, yet as fundamentally different from it in other crucial respects. On the one hand, a person should be construed by analogy with external objects of experience insofar as the person's specific psychological features are embedded in the spatio-temporal, causally structured world and therefore give rise to a particular kind of experience, namely inner experience. On the other hand, a person fundamentally differs from *mere* objects of experience in that a person must also be construed as a mental being endowed with particular faculties for representation and a distinctive representational perspective, as well as with the ability for self-determination.

My second thesis is, then, that *psychological persons form themselves in the course of realizing their mental capacities under the guidance of a unifying idea, the idea of the soul*. So this book defends what I call the *self-formation view* of the psychological person. On this view, a psychological person is understood not as a self-contained entity that exists prior to the particular happenings of one's mental life, but rather as an entity that first emerges through self-formation in the course of mental activity. An individual person is precisely the unique mental whole that progressively evolves through exercising mental capacities under the normative guidance of a unifying idea, viz. the idea of the soul. The central task of such an interpretation is to discern the conditions that make the formation of oneself as a psychological person possible.

I focus primarily on the experiential, rather than the agential, side of personhood. That is, I explore the nature of psychological persons insofar as they can know themselves through inner experience, rather than insofar as they act as rational agents in the world. Therefore, I examine the conditions of self-formation with regard to the conditions of inner experience, rather than with regard to the conditions of agency. For the former, I draw mainly on the resources of Kant's theoretical philosophy, whereas the latter would involve a close examination of Kant's practical philosophy. Despite confining myself primarily to the experiential side, I firmly believe that the conception of psychological personhood I am offering here is compatible with Kant's theory of agency and can be expanded in this direction in the future.

In sum, my interpretation has the following three characteristics:

- (1) It remains agnostic regarding the intrinsic nature of that which appears in inner experience, such as an underlying pre-existing soul or non-material substance.
- (2) It takes the idea of the soul – as the concept of a unified mental whole – to be a regulative guideline for determining one's psychological features in time.
- (3) It takes the idea of the soul to define normative demands both for acquiring self-knowledge and for realizing oneself as a unified person.

The starting point from which I develop this interpretation is the puzzle of self-reference, to which I now turn.

### i.2 The Puzzle of Self-Reference: Parity or Disparity?

Empirical self-knowledge raises an intricate puzzle – a puzzle that is indeed a problem for any philosophical or scientific theory addressing it. On the one hand, self-knowledge is *reflexive* in that it points back to the representing subject who has such knowledge. On the other hand, self-knowledge *refers* to a particular individual “object”, namely oneself, with specific psychological features. The puzzle thus concerns the issue of how actively thinking subjects can represent themselves as passively given objects without distorting themselves or becoming estranged from themselves. That is, how can self-knowledge be *self-referential* at all?

Let me expand on this thought. Self-knowledge seems to involve two ways of representing oneself: representing oneself *as subject* and representing oneself *as object*. In contemporary philosophy of mind, this issue has been reflected in the distinction between two kinds of self-consciousness, between “consciousness of oneself *as subject*” and “consciousness of oneself *as object*”.<sup>2</sup> Contemporary philosophers of language often appeal to Wittgenstein’s famous distinction between two uses of the first-person pronoun “I”: the use of “I” as subject and the use of “I” as object.<sup>3</sup> While this terminology is certainly helpful, it leaves the following two questions unanswered. Firstly, if I were to become the object of my own experience, what *kind of object* would I be for myself? Would I be a mind endowed with mental capacities, an embodied human being, a collection of mental states, or rather something else? Secondly, if my self-knowledge is primarily concerned with the subjective contents passing through my consciousness, can these contents ever become items of knowledge meeting the standards of *objective validity*? That is, can I objectively know myself?

This study is driven by the belief that Kant’s Critical Philosophy – as an enquiry into the necessary conditions of the possibility of experience – provides an exceptionally productive framework for examining these questions. In his *Critique of Pure Reason* (1781/1787; henceforth, first *Critique*), Kant assigns a central role to the thinking subject in the constitution of experience of objects and thereby conceives of experience as *empirical theoretical cognition of objects*, rather than as merely subjective sensation.<sup>4</sup> A detailed analysis of Kant’s account of mental faculties offers, I shall

<sup>2</sup> For example, Cassam (1997), Carl (2014), Frank (2007), Longuenesse (2017).

<sup>3</sup> Wittgenstein (1958); see also, for example, Shoemaker (1968), Perry (1979), Lewis (1979), and, more recently, García-Carpintero and Torre (2016).

<sup>4</sup> In this study, I confine myself to empirical theoretical cognition of objects, which I will just call *cognition*, if not stated otherwise.

argue, crucial resources for resolving the puzzle of self-reference. Such an analysis will reveal the distinctive ways in which we relate to ourselves as objects, while at the same time acknowledging our nature as thinking subjects.

According to Kant's transcendental philosophy, *experience* (*Erfahrung*) is a kind of *empirical cognition* (*empirische Erkenntnis*), that is, experience consists in sensation-based judgements about an object or an objective reality. Experience results from the mind's activity of bringing a multitude of sensations under empirical concepts and of combining those concepts into judgements. The two main faculties involved in experience are the faculty of sensibility and the understanding as the faculty to judge. *Sensibility* (*Sinnlichkeit*) immediately relates to objects and yields sensory *intuitions* (*Anschauungen*) of them according to the forms of time and space. The *understanding* (*Verstand*) applies general *concepts* (*Begriffe*) and yields *judgements* (*Urteile*) about objects of experience in accordance with its basic forms, the so-called categories such as unity, substance, and causality. Both the forms of sensibility and the forms of the understanding are conditions of the possibility of experience that are universally shared by all humans.

The paradigmatic case of experience that Kant considers throughout the first *Critique* is the cognition of material objects in space. The notion of inner experience is nonetheless a ubiquitous theme in the first *Critique*, and in other works from the same period.<sup>5</sup> Kant construes inner experience – by and large – by analogy with outer experience. By “observing” ourselves through inner sense, as opposed to observing outer objects through outer sense, we are able to become aware of our mental states, such as perceptions, feelings, and passing thoughts. On the basis of this awareness, we are then able to make judgements about these states and their temporal relations, and arguably about more general psychological properties, such as character traits, moods, passions, and standing attitudes. Inner experience consists in such empirical judgements about one's psychological features.<sup>6</sup>

This conception of inner experience raises precisely the puzzle that I have stated above. In light of Kant's account of experience in the *Critique*, we can now ask: firstly, if inner experience concerns the subjective states of consciousness, can it then be understood as the representation of an object and, if so, of what kind of object? Secondly, can inner experience fulfil the conditions of objective validity necessary for empirical cognition of myself? Kant is aware of the intricate puzzle that is raised by the fact that “the I that I think is to differ from the I that intuits itself . . . and yet be identical with the latter as the same

<sup>5</sup> For example, B277–9, A672/B700; *Anth* 7:141–2, 7:161–2.

<sup>6</sup> He also uses other phrases, each emphasizing a different aspect, such as *determining my existence in time* (see Bxl, A35/B53, B157n, B430–1), *cognizing myself as I appear to myself* (see B68, B155, B157–8, A337/B394), and *connecting my inner appearances and actions* (see A672/B700, A683/B711).

subject” (B155–6). But unfortunately Kant does not offer a clear account of his solution to the puzzle.

Despite its frequent recurrence in the *Critique* and other texts, the theme of inner experience has mostly aroused suspicion and perplexity among commentators regarding its nature and epistemic status, and only little interpretive consensus about it has been achieved. Most notably, there is a striking lack of discussion of it in the Anglophone literature, and also a surprisingly small amount of discussion in German or in other languages.<sup>7</sup> When it is discussed, various commentators are sceptical about whether inner experience is able to play a significant role in Kant’s Critical Philosophy, and some even deny that inner experience can be theoretical cognition of an object at all.<sup>8</sup>

There are two main rival lines of interpretation of inner experience within Kant scholarship and beyond. One line of interpretation construes inner experience as cognition of a mental object on a par with the construal of outer experience as cognition of physical objects. Another denies that such an objectual or objectified grasp of oneself is in any way possible and claims that self-knowledge, if one can speak of it at all, must be knowledge of an entirely different kind. In the history of the reception of Kant’s thought, both these approaches have had persuasive defenders and both are still alive today. The tendency towards the objectification of the mental has been pursued predominantly by neo-Kantian and naturalist interpretations of Kant’s philosophy; the subjectivist line of interpretation has been defended mainly by German idealists, phenomenologists, and existentialists.<sup>9</sup> This leads me to distinguish, in a more systematic way, two interpretive approaches to Kant’s conception of inner experience, both of which are problematic: the *parity view* and the *disparity view*.

According to the *parity view*, inner experience is construed in the same way as outer experience and so, like the latter, as empirical cognition of an object. This view appeals to an alleged structural parallel between inner and outer sense and holds that inner and outer experience are determined through the same set of formal conditions, despite some specific differences, such as the non-spatiality of mental states and the lack of the attractive and repulsive forces of matter in the case of inner experience.<sup>10</sup>

<sup>7</sup> The notion of inner experience is rarely discussed or altogether absent from studies of Kant’s accounts of the mind, the self, and empirical psychology, such as Ameriks (2000), Kitcher (1990, 2011), Wuerth (2014), and Longuenesse (2017). Exceptions include some neo-Kantians, such as Meyer (1870), and, more recently, Mohr (1991), Emundts (2007), Frierson (2014), and Dyck (2014).

<sup>8</sup> For references, see my discussion of the disparity view later in this section.

<sup>9</sup> For the former, see Wundt (1888, 1902), Cohen (1885), Natorp (1912); for the latter, see Fichte (1794/1997), Husserl (1900/1975, 1931/1988), Heidegger (1929/1990).

<sup>10</sup> A strong tendency towards the parity view can be found in Vogel (1993), Frierson (2014), Chignell (2017); for discussion, see Kraus (2019a).

In contrast, another diverse set of interpretations emphasizes the disparity between inner and outer experience and therefore rejects the identification of inner experience with the empirical cognition of an object in the proper sense. Although these interpretations come in a wide variety of flavours, I subsume them under the common label *disparity view*. Some defenders of the disparity view argue that inner experience amounts not to cognition of a mental object but to some sort of empirical awareness of one's own mental activity, which is conditioned only by the features of reflexive thought.<sup>11</sup> Others detect an apparently insurmountable disparity between inner and outer experience on the grounds that inner experience allegedly lacks a referent, since no object is given to inner sense that could be determined by the category of substance.<sup>12</sup> For a third group, Kant's notion of inner experience can be meaningfully discussed only as an aspect of practical deliberations in the context of agency.<sup>13</sup> Furthermore, Kant's apparent denial of proper scientific status to empirical psychology in the *Metaphysical Foundations of Natural Science* (1786; henceforth *Foundations*) has led many commentators to think that Kant must reject the very possibility of any theoretical knowledge of psychological phenomena and hence of inner experience as empirical cognition.<sup>14</sup>

The novel account that I develop in this study proposes a more nuanced understanding of inner experience as empirical cognition, which is able to answer the worries of the disparity theorists. In line with the parity view, I shall argue that inner experience is empirical cognition of a person's psychological features. Yet the special nature of persons, which gives rise to some of the disparities pointed out above, will warrant a crucial qualification: inner experience is only *analogous* to experience of spatio-material objects. The most important difference is that inner experience is not cognition of a persistent substance given to inner sense; rather, the object of inner experience is first formed in the course of one's mental activity under the guidance of the unifying idea of the soul. Hence, the account of inner experience offered here – in combination with the self-formation view of psychological personhood – preserves conceptual coherence with Kant's notion of outer experience and yet remains sensitive to the distinctive systematic concerns that arise from the puzzle of self-reference.

<sup>11</sup> For example, Henrich (1994) and Keller (1998); see also Kitcher (2011), Brook (1994), Hatfield (2006), Schmidt (2008), Longuenesse (2017).

<sup>12</sup> For example, Strawson (1966), Mischel (1967), Gouaux (1972), Washburn (1976), Nayak and Sotnak (1995), Westphal (2004), Friedman (2013).

<sup>13</sup> For example, Sturm (2001, 2009:205–60) and Emundts (2017); see also Makkreel (2001) and Cohen (2009).

<sup>14</sup> See *MFNS* 4:471. For example, Mischel (1967), Gouaux (1972), Washburn (1976), Leary (1978), Schönrich (1991), Friedman (2013), Hudson (1994), Klemme (1996), Makkreel (2003). For discussion, see Kraus (2018).

### i.3 The Argument of the Book: Varieties of Objects and Varieties of Self-Consciousness

What is at stake in the dispute between parity and disparity views is whether we can discern an “object” of which inner experience could achieve empirical cognition. The central question is then: what, if anything, do we cognize in inner experience? Reading Kant’s texts, the following candidates may recommend themselves: the thinking subject (*denkendes Subjekt*), the self (*Selbst*), the mind (*Gemüth*), the soul (*Seele*), a mere collection of inner appearances (*innere Erscheinungen*), or the whole embodied human being (*Mensch*). A central element of inner experience is the term “I” (and its cognates), as Kant himself states in a note: “All my inner experience is a judgement in which the predicate is empirical and the subject is I” (*Refl* 5453, 18:186, my translation). Accordingly, inner experience typically consists in what one may call *I-judgements*, that is, judgements that can typically be expressed by statements of the basic form “I  $\varphi$ ” (and derivations thereof), where the subject term is always the first-person pronoun “I”, and “ $\varphi$ ” a mental predicate. Examples include the self-ascription of occurrent mental states, such as “I see a red rose”, “I believe that the sun is shining”, or “I feel joy”, as well as the self-ascription of more general, temporally extended properties such as “I am a generous person” or “I aim to become a successful piano player”. What, if anything, does the term “I” refer to in such I-judgements?

What exactly, on Kant’s view, the “I” of inner experience denotes can be understood only within his more comprehensive theory of representation, which spells out the basic relation of a subject towards an object in terms of a variety of mental faculties and their characteristic forms. Based on this account of mental faculties, Kant distinguishes several kinds of objects and, as I shall argue, several corresponding kinds of self-consciousness, some of which involve the use of the term “I”. Given Kant’s account of empirical cognition as involving both sensibility and the understanding, we can distinguish at least three fundamental ways in which a human subject can relate to an object, corresponding to the following three types of representation:

- (R-i) via sensible representations, viz. *intuitions*, through which a particular object is given;
- (R-ii) via intellectual representations, viz. *concepts*, through which an object is thought;
- (R-iii) via *empirical cognition*, through which an object is cognized on the basis of (R-i) and (R-ii).

Corresponding to these three types of representation, we can now distinguish three kinds of object:



- (O-i) the *object of the senses* (commonly called *appearance*, *Erscheinung*);<sup>15</sup>  
 (O-ii) the *object of thought* (sometimes called *thought-entity*, *ens rationis*, *Gedankending*);<sup>16</sup>  
 (O-iii) the *object of experience*, that is, an *object of empirical cognition* (commonly called *object*, *Gegenstand*).

Kant provides complex arguments for how these fundamental ways of relating to objects can be integrated into a single account and in particular for how intuitions are amenable to the conceptual conditions of thought and judgement. Some of these arguments will be relevant to the course of my later analysis. Important at this stage is the insight that taking something as an object does not necessarily mean representing it as an object of experience. The necessary conditions of the possibility of cognition – or what Kant calls its *transcendental conditions* – include time and space as the forms of sensibility and the categories as the forms of the understanding.<sup>17</sup> If a representation satisfies these transcendental conditions, then it is objectively valid, that is, it can be assessed as to its truth regarding the object and independently of the particular representing mind. We can now ask whether inner experience satisfies these transcendental conditions for theoretical cognition of an object. If this were not the case, then the inner domain for Kant would only amount to a merely subjective awareness of mental states in a more or less indeterminate manner, and not to experience in his distinctively demanding sense. Unfortunately, Kant remains ambivalent with regard to the existence and status of the object of inner experience. He appears to subscribe to two seemingly conflicting claims:

- (K1) *No object is given to inner sense that corresponds to an empirical inner intuition and that subsequently can be cognized in inner experience.*<sup>18</sup>  
 (K2) *In inner experience, I cognize myself as an object.*<sup>19</sup>

The former claim appears to support the disparity view, whereas the latter claim seems to support the parity view.

In order to resolve this conflict, the book sets out to examine which types of representations and which kinds of objects are involved in inner experience.

<sup>15</sup> For example, Bxxvi, A20/B34.

<sup>16</sup> A thought-entity, strictly speaking, is an object of *mere* thought to which no intuition corresponds (see A290/B347).

<sup>17</sup> See A51/B75–6 and A92/B125.

<sup>18</sup> For example, A22/B37, A107, B275, A350, A381, B412.

<sup>19</sup> “Cognition of myself” is mentioned, for example, at Bxl, B139, B155, B158, B277–9, B400, B431; *Anth* 7:142, 7:161; *Prol* 4:336. The “object of inner sense” is specifically mentioned at A342/B400, A357, A368, A385, B403, B415, A443/B471, A846/B874; *CpracR* 5:95; *MFNS* 4:467, 4:542; *Refl* 6313, 18:614; *Anth* 7:142.



The main argument of the book thus consists in discerning the conditions of different types of self-consciousness, that is, of types of representations a subject can have of itself. Corresponding to the notions of object defined above, I distinguish the following types of self-consciousness:

- (SC-i) consciousness of one's inner appearances, received by inner sense;
- (SC-ii) consciousness of oneself as an object of thought, referred to by the term "I";
- (SC-iii) consciousness of oneself through inner experience, expressed in empirical I-judgements.

All these types of self-consciousness – as I will show – have to be classified as *consciousness of oneself as object*. They have to be distinguished from another, still more fundamental type, namely:

- (SC-0) consciousness of oneself as thinking subject, via transcendental apperception.

I will call this also *transcendental self-consciousness*, since it defines – as I will argue – the general form of reflexivity that any conscious representation must display, regardless of its specific type. Transcendental self-consciousness is thus a condition of the possibility of *all* empirical consciousness.

Let me add a further note of clarification with respect to Kant's doctrine of transcendental idealism. The notion of object is ambiguous in yet another way in that it can "be taken in a twofold meaning, namely as appearance or as thing in itself" (Bxxvii). Only appearances can be determinable objects for a subject, whereas things-in-themselves transcend the bounds of our senses and cannot be determined through experience. My aim here is neither to defend an interpretation of Kant's transcendental idealism, nor to give an account of inner experience in terms of the thinking thing-in-itself (or noumenon) that may be believed to appear in inner experience. Rather, I focus on those notions of object that are relevant for an analysis of inner experience. In the most general sense, an object is understood as something that can be represented and to some extent determined by a subject – *an object for a subject*.<sup>20</sup>

#### i.4 The Novel View of the Book: Self-Formation under the Idea of the Soul

As a solution to the puzzle of self-reference, this book argues that in inner experience we cognize ourselves *not* as *mere* objects of experience, since we are

<sup>20</sup> Etymologically, "object" (*Objekt*) is derived from *objectus* (Lat.), meaning "that which is thrown or put before or up against" the subject. Similarly, *Gegenstand* literally translates as "that which stands up against".

not given to ourselves as such objects in the first place. Rather, our inner experience is fundamentally shaped by our nature as human subjects who – endowed with mental faculties and the ability for self-determination – have an individual representational perspective on ourselves and the world. As such subjects, we must conceive of ourselves in inner experience, despite the fact that we cannot intuit ourselves as such. As a result, this book not only develops a novel account of the very “object” of inner experience, which accommodates parities as well as disparities between inner and outer experience, but also offers a novel account of psychological personhood per se.

The account of personhood this book defends, which I call the *self-formation view*, is crucially based on the Kantian idea of the soul. As an idea of reason, no corresponding intuition can be given to it. That is, we cannot intuit ourselves as souls, though we may very well conceive of ourselves as such. Reason, for Kant, is the highest intellectual faculty, whose primary function is the capacity for inferring conclusions that go beyond that which is immediately given in experience. Assigning a central role to the idea of the soul within Kant’s Critical theory of experience may sound perplexing to many readers. The Paralogisms in the first *Critique* are commonly read as a devastating criticism of the metaphysical theories of his day, including rationalist theories of the human soul. In consequence, Kant is often taken to deny us any theoretical knowledge of the human soul, including properties such as substantiality, simplicity, and non-materiality. Yet there is ample evidence that Kant never abandons the notion of the soul altogether, nor does he outright reject all meaningful employment of it. Even in his most mature published works, such as the three *Critiques*, and in his late lectures on metaphysics and anthropology, Kant still makes some positive use of this notion in the context of theoretical reason and psychology – a fact that is rarely discussed in contemporary Kant scholarship or appreciated as an interesting and illuminating aspect of his Critical Philosophy.<sup>21</sup> The self-formation view of psychological personhood recovers the idea of the soul by showing it to be not only compatible with Kant’s Critical Philosophy, but also extraordinarily philosophically productive for accounts of personhood and self-knowledge.

The crucial point is that the idea of the soul – as the concept of a mental whole – defines the unity of a person. It defines the context within which we can first make sense of inner appearances in consciousness as the states of our own mind. Moreover, it defines the guiding principle by which we come to unify all aspects of our mental life as belonging to the same person. The idea of the soul is thus the unifying form of a person’s mental life.

This interpretation is inspired by a claim that Kant puts forward in the Appendix to the Dialectic in the first *Critique*. After having diagnosed the

<sup>21</sup> Exceptions include Klemme (1996), Heßbrüggen-Walter (2004), Serck-Hanssen (2011), Goldman (2012), Dyck (2014), Wuerth (2014), and Falduto (2014:24–33).