

Paul Katsafanas

4 Kant and Nietzsche on Self-Knowledge

A typical view of Kant and Nietzsche on self-knowledge might go something like this: Kant thinks that our mental lives are entirely transparent to us, whereas Nietzsche claims that the mind is utterly opaque. Kant believes that we usually know what we are doing and why, placing great faith in our introspective capacities; Nietzsche believes, to put it in his own words, that “actions are *never* what they appear to be [...] all actions are essentially unknown” (D 116). In short: Kant and Nietzsche are antipodes, the one emphasizing self-knowledge and the other self-ignorance.

A closer examination of Kant and Nietzsche reveals that this picture is misleading on several levels. Kant does not embrace the naïve claim that our minds are transparent to us; he is alive to the possibility of unconscious, inaccessible, hidden, and repressed mental contents. And while Nietzsche firmly maintains that there is no hope of rendering *all* mental content conscious, he does think that there are certain routes that bring us to greater self-knowledge.

In this essay, I pursue a more nuanced account of the differences between Kant and Nietzsche on self-knowledge. I examine the *kinds* of self-knowledge that they recognize, the *extent* to which self-knowledge is possible, and the best *routes* to attaining it. The essay falls into three sections. Section One examines Kant’s claim that there are two capacities for self-knowledge, which he terms “inner sense” [*innerer Sinn*] and “apperception” [*Apperzeption*]. Neither of these capacities is immune to error, and neither provides us with comprehensive knowledge of our mental economies. Section Two argues that Nietzsche agrees with Kant that these two capacities yield some self-knowledge. Yet Nietzsche is far more pessimistic about their prospects; given the complexities of our mental economies, these capacities give us at best a rudimentary knowledge of our mental lives. Nonetheless, the differences between Kant and Nietzsche on these points are a matter of degree rather than of kind.

Section Three introduces what is, in my view, a more important difference between Kant and Nietzsche. Nietzsche argues that we can attain self-knowledge by—as he puts it—looking away from ourselves. I argue that this is the real point of departure between the two thinkers; it marks Nietzsche’s most profound break with Kant on the topic of self-knowledge. In addition, I contend that in examining Kant and Nietzsche’s views on self-knowledge, we can see a profound shift in ways of regarding our mental lives. Although Kant recognizes the possibility of self-deception and acknowledges that the mind is in certain respects opaque, he clings to the premodern idea that opacity is—or at least

should be—correctible. Nietzsche, on the other hand, treats opacity as ineliminable and sees the traditional routes to self-knowledge as misdirected.

1 Kant on self-knowledge

Kant famously writes,

The **I think** must be **able** to accompany all my representations: for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing to me. (CPR B 131–132)¹

The claim that the “I think” must be capable of accompanying all representations might suggest that all representations are self-conscious or at least self-consciously accessible. Moreover, Kant claims that “maxims”—specifications of what we are doing and why—are the object of moral assessment, which might suggest that he thinks we can have certain knowledge of our motives for acting. And he gives self-consciousness pride of place, writing that

The fact that the human being can have the “I” in his representations [i.e., is self-conscious] raises him infinitely above all other living beings on earth. Because of this he is a *person* [...] i.e., through rank and dignity an entirely different being from *things*, such as irrational animals, with which one can do as one likes. (*Anthropology*, AA 7: 127)

Given the prominence and preeminence of self-consciousness in Kant’s writings, it is tempting to assume that he takes our mental lives to be transparent to us.

However, this is a mistake. Kant quite bluntly asserts that “the depths of the human heart are unfathomable” (*Metaphysics of Morals*, AA 6: 447).² He writes that “it is absolutely impossible by means of experience to make out with complete certainty a single case in which the aim of an action otherwise in conformity with duty rested simply in moral grounds” (*Groundwork*, AA 4: 407). In other words, it is absolutely impossible to know, with complete certainty, whether I have acted on the motive that I take myself to have acted upon. Indeed, when I

1 I cite Kant’s works by the volume and page number of the Berlin Akademie Ausgabe, *Kants gesammelte Schriften*, hereafter abbreviated AA, except for the *Critique of Pure Reason* (hereafter CPR), which is cited by A/B version page numbers.

2 He continues, “Who knows himself well enough to say, when he feels the incentive to fulfill his duty, whether it proceeds entirely from the representation of the law or whether there are not many other sensible impulses contributing to it that look to one’s advantage (or to avoiding what is detrimental) and that, in other circumstances, could just as well serve vice?” (*Metaphysics of Morals*, AA 6: 447).

think I have discovered the motive for my action, this “discovery” is often an invention: in a line that sounds decidedly Nietzschean, Kant writes, “without noticing it, we suppose we are discovering within ourselves what we ourselves have put there” (*Anthropology*, AA 7: 133). Our alleged self-knowledge often turns up only distortions, confabulations, and errors. Thus, Kant writes

the field of sensuous intuitions and sensations of which we are not conscious, even though we can undoubtedly conclude that we have them, that is *obscure* representations in the human being (and also in animals) is immense. Clear representations, on the other hand, contain only infinitely few points of this field that lie open to consciousness; so that as it were only a few places on the vast *map* of our mind are *illuminated*. (*Anthropology*, AA 7: 135)

So Kant is at once deeply committed to the importance of self-consciousness and skeptical about its extent. To see why this is so, we need to investigate the way in which Kant thinks we achieve self-knowledge. There are two paths to self-knowledge, each with its own peculiar limitations.

1.1 Two kinds of self-knowledge

Kant claims that there are two sources of self-knowledge: inner sense [*innerer Sinn*] and apperception [*Apperzeption*].³ Inner sense is what we would today call introspection: it is “consciousness of what we undergo as we are affected by the play of our own thoughts” (*Anthropology*, AA 7: 161). It provides us with knowledge of our own sensations. Kant believes that it is a form of perception: he writes that inner sense is “a mere faculty of perception” (*Anthropology*, AA 7: 45).

But we also have a second source of self-knowledge:

Man ... who knows the rest of nature solely through the senses, knows himself also through pure apperception; and this, indeed, in acts and inner determinations which he cannot regard as impressions of the senses. (CPR A 546/B 574)

Apperception is knowledge of our own activity. Kant’s idea, here, is that we do not attain knowledge of our own judgments and actions through inner sense. Take a typical case: I decide to walk to my office to do some work. I set out; the action is underway. I needn’t observe myself or attempt to introspect my motives in order to determine what I am doing; I know, without needing to resort to the observation of sensory states, that I am walking to my office in order to do some work. It seems natural to say: I know what I am doing because my action is a

³ See CPR A 107, B 132, B 152–159, and *Anthropology*, AA 7, Section 24.

product of my *decision*, my “inner determination”, to walk to my office. In short, it seems that I sometimes know that I am A-ing precisely because I have made a decision to A. It is in this way that apperception is non-sensory.

Kant makes it clear that these two forms of self-knowledge are distinct; neither can be reduced to the other. He complains that “the terms *inner sense* and *apperception* are normally taken by psychologists to be synonymous, despite the fact that the first alone should indicate a psychological (applied) consciousness, and the second merely a logical (pure) consciousness” (*Anthropology*, AA 7: 142). In the following passage, he explains why the conflation of apperception and inner sense is an error:

Inner sense is not pure apperception, consciousness of what the human being *does*, since, this belongs to the faculty of thinking. Rather, it is consciousness of what he *undergoes*, insofar as he is affected by the play of his own thoughts. It rests on inner intuition ... There is then only *one* inner sense, because the human being does not have different organs for sensing himself inwardly ... (*Anthropology*, AA 7: 161)

So Kant draws a sharp distinction between the way in which we know our own judgments and their products, on the one hand, and the way we know our sensations and appetites, on the other. Apperception is a power of thinking; inner sense is a mode of sense-perception. Moreover, these passages emphasize that inner sense pertains to the passive—to what we *undergo*. Apperception pertains to the active, to what we *do*.

These distinctions between activity and passivity play an important role not only in Kant’s moral theory, but in his analysis of the mind:

in regard to the state of its representations, my mind is either *active* and exhibits a faculty (*facultas*) or it is *passive* and consists in *receptivity* (*receptivitas*). A cognition joins both together ... Representations in regard to which the mind behaves passively, and by means of which the subject is therefore *affected* ... belong to the *sensuous* cognitive faculty. But ideas that comprise a sheer *activity* (thinking) belong to the *intellectual* cognitive faculty. (*Anthropology*, AA 7: 140)

Apperception and inner sense fall on opposite sides of this divide.

In sum, then, inner sense is a type of perception directed at our sensations. Apperception is a consciousness of our own activity, and is non-sensory.

1.2 The limitations and failures of self-knowledge

Above, we noted that Kant believes that self-knowledge often fails: “it is absolutely impossible by means of experience to make out with complete certainty a

single case in which the aim of an action otherwise in conformity with duty rested simply in moral grounds” (*Groundwork*, AA 4: 407). He reiterates this point elsewhere, writing:

even a human being’s inner experience of himself does not allow him so to fathom the depths of his heart as to be able to attain, through self-observation, an entirely reliable cognition of the basis of the maxims which he professes, and of their purity and stability. (*Religion*, AA 6: 63)

Notice that Kant does not limit failures of self-knowledge to sensations, the province of inner sense; these passages emphasize that knowledge of our own actions—the realm of apperception—can also fail. Let’s start, though, with the problems of inner sense.

Inner sense can fail due to the intrinsic limitations of perceptual capacities. It faces all the vulnerabilities of more familiar forms of perception, such as vision. In particular, for Kant perception provides us with access to appearances rather than things-in-themselves. He writes,

experience (empirical cognition), inner no less than outer, is only the cognition of objects as the *appear* to us, not as they *are* (considered in themselves alone). For what kind of sensible intuition there will be depends not merely on the constitution of the object of the representation, but also on the constitution of the subject and its receptivity. (*Anthropology*, AA 7: 141)

The way in which an object appears to me depends not only on the nature of the object but also on the nature of my cognitive capacities. I see the object before me as a brown oblong table; but these colors, shapes, and classifications need not be attributes of the object as it is independently of perception. Just so with my inner states: I perceive myself as resentful of Fred’s success, but this may diverge from my actual motives. As Kant puts it, inner sense “represents to consciousness even our own selves only as we appear to ourselves, not as we are in ourselves. For we intuit ourselves only as we are inwardly *affected*” (CPR B 153). So inner sense is susceptible to all the problems that plague outward sense perception.

A second mode of failure for inner sense arises not from the intrinsic limitations of the faculty, but from an extrinsic factor: pressures toward self-deception. Kant writes that “one is never more easily deceived than in what promotes a good opinion of oneself” (*Religion*, AA 6: 68). Suppose I engage in behavior that might be appropriately described either as impressing Robert or belittling Claire; I am more likely to interpret myself as acting on the former, praiseworthy motive rather than the latter, disagreeable motive. After all, the former interpretation enables me to preserve a positive image of myself. As Kant puts it, “We like to flatter ourselves by falsely attributing to ourselves a nobler motive, whereas in fact we can never, even by the most

strenuous self-examination, get entirely behind our covert incentives” (*Groundwork*, AA 4: 407). Put simply, “the human being has from nature a propensity to dissemble” (AA 25: 1197), and this shows up in our interpretations of ourselves.

So failures of inner sense can be traced both to the intrinsic limitations of perception and to extrinsic interferences. Consider how this might apply to apperception. Apperception is non-sensory, so it cannot be directly affected by the ordinary limitations of perceptual systems. However, it can be influenced by them at one remove. Take an external analogue: if poor vision leads me to perceive the shadow as a threatening person, I may judge that I ought to run away in order to escape from danger. I will have apperceptive knowledge of this judgment and the resultant action. In one way, no error infects this judgment: I know that I am running away in order to escape danger. In another way, there is an error: there is no danger from which to escape. Just so with self-knowledge. I have inner awareness of my motive for impressing Robert but not for belittling Claire. As a result, I decide to act in a way that will impress Robert, but fail to notice that it also comes at Claire’s expense. The error begins with inner sense but resounds into the domain of apperception.

Additionally, the pressure toward self-deception can infect apperception as easily as inner sense, leading me to misrepresent my own grounds for action. Suppose we alter the above example slightly: I have inner awareness of both motives, but decide to act on the praiseworthy one. Out of a desire to think well of myself, I ignore the effects of the disagreeable motive and take myself to be acting only on the praiseworthy ones. Nonetheless, external observers would describe me as influenced by the disagreeable motive. Here, my decision is not fully transparent to me; self-deception has led me astray.

Thus, apperception can be misled by failures of inner sense and pressures toward self-deception. Might apperception also suffer from intrinsic limitations? Kant isn’t explicit about this, but we can envision one way in which it might. Suppose our apperceptive knowledge results from the fact that our decisions determine our actions. For example, I have apperceptive knowledge that I am walking to my office because I have decided to perform this action and the decision has caused the action. If our decisions play a less important role than Kant thinks—if our decisions *contribute to* but do not *fully determine* the nature of our actions—then apperceptive knowledge might be limited. I will argue, below, that this is exactly the possibility that Nietzsche explores.

2 Nietzsche on self-knowledge

In this section, I will argue that while Nietzsche acknowledges that both introspection and apperception can provide us with some degree of self-knowledge,

he is less impressed than Kant with their powers.⁴ Nietzsche argues that the limitations of these capacities are far more complex and pervasive than Kant acknowledges. It's not just that we miss hidden motives and are misled by self-deception; the problems are much deeper, to the extent that we often completely misunderstand our actions. The same point applies to apperception: while Kant acknowledges that our choices sometimes have facets of which we are unaware, Nietzsche adds that our choices play a much less important role in the production of our actions than we suppose, and accordingly provide us with less extensive knowledge of what we are doing. Thus, the acts issuing from our decisions can be as obscure as the acts of others.

2.1 Nietzsche on introspection or inner sense

As we saw above, Kant believes that the main limitations on inner sense arise through the pressures of self-deception. Nietzsche certainly agrees that self-serving interpretations of attitudes are ubiquitous and constitute a danger for those seeking self-knowledge. For example, he writes,

Self-observation. – Man is very well defended against himself, against being reconnoitered and besieged by himself, he is usually able to perceive of himself only his outer walls. The actual fortress is inaccessible, even invisible to him, unless his friends and enemies play the traitor and conduct him in by a secret path. (HH I 491)

However, there is an important difference between Nietzsche and Kant on this point. Kant's emphasis on self-deception suggests that, were self-deception absent, there would be no great obstacles to self-knowledge; inner sense, cleared of the distortions of self-deception, would provide us with generally accurate and reasonably comprehensive knowledge of ourselves, just as perception provides us with generally accurate and reasonably comprehensive knowledge of the external world.

Nietzsche rejects this assumption. The idea that we *deceive ourselves* about our motives implies that we have some dim cognizance of our actual motives, but hide them from ourselves. Take the case that concerns Kant: although I present myself as acting for beneficent reasons, I am actually motivated by self-interest. If this is to count as a case of self-deception, I must have at least some

⁴ Of course, Nietzsche doesn't use the Kantian terms "inner sense" and "apperception". However, he does discuss the capacities picked out by these Kantian terms: introspection and the knowledge resulting from judgment.

inking that I am actually motivated by self-interest. Otherwise it would be a mere error; no *deception* would be involved.

Nietzsche thinks that in many cases we lack even this dim awareness of our hidden motives. He claims that it is a “universal madness” to think that we generally know what we are doing. For “the opposite is precisely the naked reality demonstrated daily and hourly from time immemorial! [...] actions are *never* what they appear to be [...] all actions are essentially unknown” (D 116). For example, consider the scientists and philosophers discussed in the *Genealogy*, who Nietzsche presents as motivated by the ascetic ideal. His claim is *not* that these individuals secretly know that they are motivated by asceticism and deceive themselves about this; instead, his claim is that although scientists and philosophers may never have given a thought to asceticism, they are nonetheless motivated by it. There is no *deception* involved here; there is just an error. (The ascetic priests are different—they, Nietzsche emphasizes, are not just in error but self-deceived.)

So the problem for Kant is self-deception. The problem for Nietzsche is different and more complex. There are a number of factors rendering self-knowledge difficult. First, the influences of motives on thought are so pervasive and so complex that even honest attempts to know oneself will often fail (D 116, GM Preface 1). If each action were motivated by one or several desires and affects, then it would be easy enough to identify them. However, Nietzsche believes that the etiology of our actions is far more complex than this.

For one thing, we tend to attribute our actions to coarse, forceful motives such as lust, anger, love, pity, and so on. Yet Nietzsche believes that these are not the factors that have the most decisive influence on us. He writes that “the milder, middle degrees [of our affects], not to speak of the lower degrees which are continually in play, elude us, and yet it is they which weave the web of our character and destiny” (D 115). Some of these mild motives will be minute affective reactions; others will involve external environmental factors. For example, in *Ecce Homo*, Nietzsche claims that the most important factors in his own development were ones that might wholly escape notice:

these small things—nutrition, place, climate, recreation, the whole casuistry of selfishness—are inconceivably more important than everything one has taken to be important so far. (EH Clever 10)

So Nietzsche’s point is that we tend to overlook the effects of apparently inconsequential factors such as minute, mild affects, climate and diet; yet it is just these states that have a decisive influence on our actions.

More generally, for Nietzsche there is no definite set of motives prompting action: no matter how deeply we investigate, no matter how many motives we

uncover for an action, we can always find a deeper layer. This is part of what Nietzsche means when he writes,

Cause and effect: there is probably never such a duality; in truth a *continuum* faces us, from which we isolate a few pieces, just as we always perceive a movement only as isolated points, i.e. do not really see, but infer ... (GS 112, emphasis added)

Our attempts to ascertain the motives for our actions amount to nothing more than isolations of a few points on a continuum. We may find one, two, or thirty motives for an action, but there will always be more; there will always be different ways of understanding the causes, finer discriminations among the motives, different classifications or descriptions of them, and different prioritizations of them. Thus, Nietzsche writes,

Everything which enters consciousness is the last link in a chain, a closure. It is just an illusion that one thought is the immediate cause of another thought. The events which are actually connected are played out below our consciousness: the series and sequences of feelings, thoughts, etc., that appear are symptoms of what actually happens! – Below every thought lies an affect. *Every thought, every feeling, every will is not born of one particular drive but is a total state*, a whole surface of the whole consciousness, and results from how the power of *all the drives that constitute us is fixed at that moment* – thus, the power of the drive that dominates just now as well as of the drives obeying or resisting. The next thought is a sign of how the total power situation has now shifted again. (KSA 12, 1[61] = WLN, 60)

Our actions are products of our “total state” (*Gesamtzustand*), yet this total state cannot be adequately captured by talk of discrete motives, or discrete causes and effects.⁵ Nietzsche’s claim, here, is that trying to find a discrete cause for an action is analogous to trying to find one for a historical event. Take, for example, the First World War. We can isolate certain events – an assassination, a treaty, an ultimatum – but each of these is only one point in an enormously complex series of events, and each acquires its importance only as part of a much broader context. Moreover, there are very different ways of framing the causal history of this event. One explanation of the First World War might focus on the actions of a few great individuals, another on economic forces, another on spiritual malaise, another on a desire for meaningful direction, another on political machinations, another on imperialism, another on nationalism, another on delays in diplomatic communications. Each of these stories might be extremely informative; thinking that one of them must be uniquely correct (or maximally informative) would be decidedly

⁵ Constâncio (2011) provides an especially lucid analysis of this point in his discussion of the “continuum model”.

odd. Nietzsche's point is that just the same reasoning applies to actions. Different specifications of motives, different description of causal histories will highlight different features of the agent's mental economy, and each will be incomplete. Other stories highlighting different features will always be possible.

This brings us to a closely related point involving language and concepts. I have argued elsewhere that Nietzsche understands conscious states as linguistically or conceptually articulated (Katsafanas 2005; see also Constâncio 2011). Whereas unconscious states and processes have definite, structured content, this content is articulated in a nonconceptual form. States become conscious when their content is translated into a conceptual form. If this is correct, then the very way in which we become conscious of mental activities is constrained by our linguistic and conceptual resources. If I lack the concept *ressentiment*, for example, I cannot be conscious of my *ressentiment* as *ressentiment* (I may be conscious of it as a vague ache, or as jealousy, or as envy, but not as *ressentiment*). Nietzsche claims that we lack concepts for most of the mild motives of the sort mentioned in D 115 (quoted above). It follows that our ability to bring these motives to consciousness is severely limited. I will return to this point below in Section Three, arguing that it generates a deep obstacle to self-knowledge.

For all of these reasons, Nietzsche thinks that the obstacles to self-knowledge extend far beyond self-deception. Indeed, he doubts that we can achieve self-knowledge at all:

We are necessarily strangers to ourselves, we do not comprehend ourselves, we *have* to misunderstand ourselves, for us the law 'each is furthest from himself' applies to all eternity. (GM Preface 1)

We are *necessarily* [*nothwendig*] strangers to ourselves—this is not a contingent state that might be overcome by further acts of introspection. For introspection provides only a modicum of self-knowledge, and misses much of the complexity of our mental economies.

2.2 Nietzsche on apperception

What about the second form of self-knowledge, apperception? Again, Nietzsche is less impressed than Kant with its prospects. It's true that I can attain *some* knowledge of what I am doing by deciding, judging, and so forth. But Nietzsche emphasizes that the *content* of these judgments and decisions is not as transparent to us as Kant believes.

Consider a representative passage. BGE 32 remarks that moral theories that emphasize conscious intentions—and this would include Kant's theory—involve

“the first attempt at self-knowledge”. However, these theories err by positing “the intention as the whole origin and prehistory of an action”. Today, Nietzsche suggests, we should move past this, noting that “everything about [the action] that is intentional, everything about it that can be seen, known, ‘conscious’, still belongs to its surface and skin—which, like every skin, betrays something but conceals even more. In short, we believe that the intention is merely a sign and symptom that still requires interpretation” (BGE 32).

In short, conscious decisions and intentions are superficial. But why? The problem is that background drives and affects shape the course of conscious deliberation in ways that the agent typically fails to recognize. Nietzsche believes conscious thought is “secretly guided and channeled” by the agent’s drives (BGE 3); when an agent steps back from and reflects upon a drive, the agent’s “intellect is only the blind instrument of *another drive*” (D 109). I have argued elsewhere that drives affect perceptual saliences, the ways in which we classify or categorize experiences, the manner in which we frame decisions, and the ways in which the agent’s decision progresses (see Katsafanas 2013). For example, suppose I decide to help Megan and take myself to be doing so out of compassion. Deliberating on the case, I might have the following conscious thoughts: I will offer Megan assistance, because I pity her unfortunate state. For Nietzsche, there is much more going on in the background.

First, I may be ignoring or failing to detect the presence of other motives, such as a desire to ingratiate myself with Megan or to impress a bystander (these are the sorts of concerns that occupy Kant, as we saw above). In this vein, Nietzsche remarks that “the will to overcome an affect is ultimately only the will of another, or several other, affects” (BGE 117). My reflective thought may be driven by background motives.

Second, and more importantly, I may be ignorant of the way in which my motives are affecting my perception of the situation. Nietzsche claims that a drive will “emphasize certain features and lines in what is foreign, in every piece of the ‘external world’, retouching and falsifying the whole to suit itself” (BGE 230). This point is repeated in other passages:

From each of our basic drives there is a different perspectival valuation [*perspektivische Abschätzung*] of all events and experiences. (KSA 12, 1[58])

There is no doubt that all sense perceptions are wholly permeated with *value-judgments* ... [*gänzlich durchsetzt sind mit Wethurtheilen* ...] (KSA 12, 2[95])

If my view of the world is structured by motives that I fail to detect, then my reflective deliberations—which, after all, take place in the context of this structured view—will have presuppositions, biases, and distortions that escape my

notice. To give a simple example: I may perceive Megan as in an unfortunate state precisely because I actually resent her success; I may see her as in need of pity precisely because I secretly envy her. An observer with different motives would see the situation quite differently.⁶

Third, I may misunderstand the nature of the emotion that I am expressing. I take pity to be a state aimed at helping those in need, whereas Nietzsche suggests that it can constitute a covert attempt to extend my power over others (see D 132–138, GS 338, GM III 14, BGE 260). Thus, even if I manage to *identify* one of my motives, I may not *understand* it.

Fourth, all of this complexity only scratches the surface. For surely it is a mistake to claim, with Kant, that there is a *unique* or *single* reason for my action. As discussed in the prior section, Nietzsche believes that rather than attributing actions to discrete causes, we should see them as emanations from the “total state” of our mental economy. Conscious decisions, thoughts, and motives are one part of this total state, but only a small part. For example, once I have recognized that my pity for Megan constitutes an attempt to express my power over her, I can cut still deeper, asking why I am motivated to express my power in just this way, why pity in particular seems fitted here. Doing so would reveal even more of the “total state” of my mental economy.

This is just a quick and simplified example, but we can already see what bothers Nietzsche about the Kantian analysis of apperceptive knowledge. Kant is perfectly right that our decisions and judgments provide us with some self-knowledge. But Nietzsche draws attention to the fact that we are often ignorant of the true nature of our decisions and judgments. My decisions, influenced by the factors discussed above, have many layers of complexity that elude ordinary awareness. So they, too, provide no sure route to self-knowledge.

In addition, these reflections entail that there is no clean distinction between apperception and introspection. I may have apperceptive knowledge that I am acting so as to help Claire, but understanding what this actually means requires extended introspection, self-observation, and analysis. So the claim that we can neatly partition “doings” and “undergoings” comes to seem too simple. As Nietzsche puts it,

“I have no idea what I am *doing*! I have no idea what I *ought to do*!” – you are right, but be sure of this: *you are being done!* [du wirst gethan!] at every moment! Mankind has in all ages confused the active and the passive: it is their everlasting grammatical blunder. (D 120)

⁶ I have analyzed this point at length in Katsafanas (2013) and Katsafanas (2015).

When we think we are doing, we are also being done. Each doing is also an undergoing, and Kant's clean divisions of our mental lives are, in Nietzsche's view, too superficial.

3 The paths to self-knowledge

We have seen that Kant and Nietzsche agree that there are two routes to self-knowledge: introspection and judgment (or inner sense and apperception). Kant acknowledges that both kinds of self-knowledge can be compromised by errors and pressures toward self-deception. Nietzsche agrees, but is far more skeptical: he sees introspection and apperception as intrinsically flawed, capable of providing us only with a selective and partial picture of our inner lives. Moreover, he gives us reasons for doubting that these capacities can be isolated from one another.

At this point, then, it might seem that Nietzsche's main advance over Kant is simply that he advocates a more realistic account of the limitations of human knowledge. However, a further difference arises when we consider the ways in which Kant and Nietzsche think we can correct the deficiencies of introspection and apperception. Given the different obstacles that the two thinkers acknowledge for self-knowledge, it stands to reason that they will identify different routes to its attainment. Below, I treat their accounts in turn.

3.1 Kant: Introspection and conscience

Kant claims that “the **First Command** of all Duties to Oneself” is to know oneself (*Metaphysics* of Morals, AA 6: 441). As he puts it,

This command is “*know* (scrutinize, fathom) *yourself*,” not in terms of your natural perfection (your fitness or unfitness for all sorts of discretionary or even commanded ends) but rather in terms of your moral perfection in relation to your duty. That is, know your heart—whether it is good or evil, whether the source of your actions is pure or impure, and what can be imputed to you as belonging originally to the *substance* of a human being or as derived (acquired or developed) and belonging to your moral *condition*. (*Metaphysics* of Morals, AA 6: 441)

In other words, the first duty to myself is to discover the propensities, dispositions, motives, and desires that bear on the moral standing of my actions. It's easy to see why Kant thinks that this is an important duty: in order to fulfill the

demands of morality, I must act on the appropriate maxims.⁷ But, as we saw in Section One, I cannot know what maxims I am acting upon without knowing my own propensities, dispositions, motives, and desires. Striving to fulfill the demands of morality therefore requires striving to know myself.⁸

But how, given the impediments to self-knowledge discussed above, am I to do so? Kant's proposal is straightforward: "only the descent into the hell of self-cognition can pave the way to godliness" (*Metaphysics of Morals*, AA 6: 441). That is, the best hope for self-knowledge is engaging in extensive acts of introspection, attempting to discover whether any covert or hidden motives might be influencing one's judgments.

Of course, there is a potential problem: we saw, above, that self-deception can lead introspection astray. How is further introspection to help, then? Won't the further acts of introspection be misled by the very same motives? If "one is never more easily deceived than in what promotes a good opinion of oneself" (*Religion*, AA 6: 68), then why would extensive introspection be likely to yield accurate verdicts? Suppose, for example, that I want to determine whether my dealings with Megan have been influenced by selfish motives. Regardless of how much I introspect, it seems that I will be influenced by the desire to acquit myself of selfishness.

Kant's response to this concern rests on the idea that while I can never be certain that I have overcome all traces of self-deception, I can be certain that I have rigorously attempted to do so. To see this, consider his analysis of conscience. Conscience is the unerring certainty that one has given one's best effort to subject oneself to scrutiny. Kant writes,

an *erring* conscience is an absurdity. For while I can indeed be mistaken at times in my objective judgment as to whether something is a duty or not, *I cannot be mistaken in my subjective judgment as to whether I have submitted it to my practical reason* (here in its role as judge) for such a judgment; for if I could be mistaken in that, I would have made no practical judgment at all, and in that case there would be neither truth nor error. (*Metaphysics of Morals*, AA 6: 401, emphasis added)

Kant emphasizes that I can be mistaken about whether something is my duty. However, I cannot be mistaken about whether I have attempted to determine what my duty is. This is the sense in which conscience is unerring.

⁷ In particular, I must act only on those maxims that pass the Categorical Imperative test.

⁸ For a helpful discussion of this aspect of Kant's view, see Ware (2009). Notice also that there is an interesting difference between Kant and Nietzsche on this point: whereas Kant claims that we have a moral commitment to seeking self-knowledge, Nietzsche warns us that seeking self-knowledge is ruinous for most individuals (EH *Clever* 9).

The same point applies to self-knowledge. I am trying to do my duty, and accordingly descend into the “hell” of introspection, examining my motives. I can be mistaken about what my motives are. But conscience provides me with an unerring knowledge of whether I have done my best to discern what my motives are. While I may not succeed in garnering accurate self-knowledge, I cannot fail to know whether I have genuinely tried to do so.

So Kant’s picture is relatively simple: the problem is self-deception; the solution is trying to detect the self-deception through honest and comprehensive acts of introspection. There are risks here: Kant claims that “nothing is more harmful to a human being than being a precise observer of himself” (AA 25: 252). For “all self-scrutinizers fall into the gloomiest hypochondria” (AA 25: 863).⁹ Confronted with my own selfish, morally deficient motives, I may be unable to maintain a positive image of myself. Nonetheless, this is the risk that must be weighed against the rewards of attaining self-knowledge.

3.2 Nietzsche: Looking away from oneself

Kant claims that we can attain self-knowledge via conscientious introspection because he believes that the chief obstacle to self-knowledge is self-deception; conscience and further introspection, he hopes, can remove that obstacle. Nietzsche, of course, is less sanguine. As we saw above, the obstacles to self-knowledge extend much further than self-deception; introspection has intrinsic limitations. It’s not that we have an adequate capacity that has, unfortunately, gone astray due to a corrupt character. Instead, we have an inadequate capacity, unsuited to its goal. Further exercises of this capacity are unlikely to help.

Additionally, Nietzsche rejects the Kantian claim that deliverances of conscience have a special epistemic standing. The sense that one has given one’s best effort to introspection may be erroneous. Nietzsche’s dismissive attitude toward conscience is perhaps clearest in WS 38, where he writes, “The bite of conscience, like the bite of a dog into stone, is a stupidity.” Elsewhere, he bemoans the fact that most individuals’ “intellectual conscience” – which would be an affect that motivates one to believe only that for which there is good evidence – is severely lacking (GS 2). More generally, Nietzsche explicitly distances conscience from justification: one’s conscience is nothing more than another feeling whose content is shaped by epistemically irrelevant factors (cf. GM II 24). As he puts it in WS 52, “The content of our conscience is everything that was reg-

⁹ These passages are quoted in Wood (2003).

ularly demanded of us without reason during our childhood, by persons whom we respected or feared.”

So neither further acts of introspection nor reliance on conscience will be particularly helpful in attaining comprehensive self-knowledge. How, then, should we proceed? Nietzsche writes that “the psychologist must look away from *himself* in order to see at all” (TI Arrows and Epigrams 35). The same point is echoed in the Preface of the *Genealogy*, where Nietzsche emphasizes that we have never *sought* ourselves—and then presents genealogy as a route to self-knowledge. On the face of it, this is an astonishing claim. I want to attain knowledge of myself, yet I am told to look away from myself in order to do so. How can looking at something *other* than the object I want to understand help? What might Nietzsche have in mind?

If the *Genealogy* is any indication, then looking away from oneself to attain self-knowledge involves looking at genealogies. So let’s consider how genealogy might uncover facts that are relevant for self-knowledge.

We can start with the most obvious possibility. Nietzsche famously remarks,

Direct self-observation [*unmittelbare Selbstbeobachtung*] is not nearly sufficient for us to know ourselves: we require history, for the past continues to flow within us in a hundred waves. (HH II 223)

Commentators often interpret this passage as claiming that motives or forces present in the deep recesses of the past still persist in the present. For example, the ascetic priests embraced the values of humility, inoffensiveness, and compassion out of *ressentiment*; so, somehow, when I embrace those values today, I am still doing so out of *ressentiment*.

If this claim were true, genealogy would be a good path to self-knowledge. Unfortunately, though, the claim is completely implausible. The fact that one type of person in a completely distinct social, cultural, and historical setting embraced a value because they experienced a certain affect does not entail that I, in circumstances that could hardly be more different, must embrace the same value out of the same affect. Consider just how *ahistorical* such a claim would be—it would entail that there are timeless, ineluctable relationships between affects and values. This hardly seems consistent with Nietzsche’s analysis of the way in which affects such as bad conscience, reverence, envy, so on have been tethered to different values over time. So Nietzsche must mean something else when he claims that the past flows through us in a hundred waves.

In fact, I think Nietzsche identifies two ways in which genealogy is relevant for self-knowledge. Let’s start with the more straightforward one. Above, we saw that one of Nietzsche’s concerns about self-knowledge is that mild and unnoticed motives play a decisive role in shaping our lives, yet elude introspection.

We miss the subtle influences upon action, for we cannot detect them in the moment of choice. Genealogy, though, can uncover them. By examining long stretches of human behavior rather than isolate moments of choice, we are able to detect the effects of these subtle, mild influences.

An image may help. Consider the kinds of explanations that arise in geology. I ask why a canyon stretches before me in the alluvial plain. The geologist tells me that the slight yet constant pressure of rain and wind has, over millennia, chiseled the canyon. Rain and wind appear to be very mild and almost unnoticeable features when compared to the resilience of stone. If I were simply to examine the rocky landscape for a moment, I would never suspect that such slight forces would yield such dramatic results. And yet, taking a longer view, I see that they do. The power of this mild cause cannot be understood without studying long stretches of time. Just so, Nietzsche thinks, with our own psychologies. We focus on the crude, momentary eruptions of force: the strong passions, the conscious decisions, the resolutions of will. And we overlook the host of minute forces that shape us in far more profound ways.

These are just the kinds of factors that genealogy and history can reveal. For example, rather than attributing our drive toward knowledge to an independent thirst for truth, we can see how it was shaped by ascetic tendencies. Or rather than seeing a philosophical system as driven by a pure urge to get at the truth, we can see it as springing from factors that the philosopher fails to recognize: “Gradually, it has become clear to me what every great philosophy so far has been: namely, the personal confession of its author and a kind of involuntary and unconscious memoir” (BGE 6). History and genealogy can help to reveal the presence of various drives, in part by showing how these drives motivate patterns of behavior that might be visible only in the long term.

So one way in which I can learn something about myself by looking away from myself is by detecting—either in historical characters, or even in my own history—gradual shapings of behavior that result from apparently minor factors. But this is not the only way in which genealogy is relevant; let’s turn to the second and more complex point.

Put briefly, genealogy helps us discover unnoticed aspects of the conceptual scheme through which we experience and interpret the world. This point is very difficult and requires careful explication; I address it in more detail in Katsafanas (in progress). However, let me sketch the view here.

In the *Genealogy*, Nietzsche is at pains to emphasize the way in which one conception of agency supplanted another. For example, he writes:

Just as the common people separates lightning from its flash and takes the latter to be a *deed*, something performed by a subject, which is called lightning, popular morality sepa-

rates strength from the manifestations of strength, as though there were an indifferent substratum behind the strong person which had the *freedom* to manifest strength or not. But there is no such substratum; there is no “being” behind the deed, its effect and what becomes of it; “the doer” is invented as an afterthought, – the doing is everything. (GM I 13)

Here, Nietzsche draws attention to the way in which our modern conception of agency treats the agent as distinct from what he does. At its most extreme, this is the libertarian conception of free will, which holds that the agent is to be identified with a characterless and utterly undetermined capacity for choice. The ancient conception, Nietzsche suggests, does not recognize this distinction: the agent’s character is given by the nature of the agent’s action, by what the agent actually does. The agent of antiquity does not take himself to be distinct from his past, his social relationships, and his community, nor does he see his actions as things that might or might not express his character; put simply, he *is* his deeds and relationships. (These accounts of agency obviously require far more explication than I can provide here.)¹⁰

Nietzsche’s concern is not whether any particular philosopher (or ordinary agent) explicitly adopts just these thoughts about agency. Rather, he is concerned with the way that an agent who internalizes something like this sense of agency will experience his action. To the extent that I tacitly adopt the modern conception of agency, I will tend to experience my choices as wholly undetermined by and unreflective of my character. I will tend to see punctual moments of conscious choice as of overriding importance. I will tend to see myself as self-defining and isolated from my environment.

In this manner, the way in which I experience my action depends upon the conception of agency that I tacitly adopt. For example, we moderns are perennially tempted to say, with Kant, that our experience of deliberation commits us to taking ourselves to be free: we cannot engage in genuine deliberation without presupposing that we are free to determine our forthcoming action.¹¹ But Nietzsche’s point is that just this sort of experience is historically contingent. It is not a datum, not a starting point, but a link in a complex story. Genealogy helps me to see that my conception of agency is *optional*: it supplanted an early form, and could be supplanted by another. Genealogy thus helps me to see that the conceptual scheme through which I view the world structures my perception of the world.

10 For more extended analyses of these matters, see Taylor (1992) and Williams (1993).

11 Kant claims that human choice “can indeed be *affected* but not *determined* by impulses [...] *Freedom* of choice is this independence from being determined by sensible impulses” (*Metaphysics of Morals*, AA 6: 213–214). Elsewhere, he writes that the will is “a faculty of determining oneself from oneself, independently of necessitation by sensible impulses” (CPR A 534/B 562).

This is part of what Nietzsche draws attention to in BGE 16, when he denies that there are any immediate certainties:

There are still harmless self-observers who believe that there are “immediate certainties”; for example, “I think,” or as the superstition of Schopenhauer put it, “I will”; as though knowledge here got hold of its object purely and nakedly as “the thing in it self” without any falsification on the part of either the subject or the object. But that “immediate certainty,” as well as “absolute knowledge” and the “thing in itself,” involve a *contradictio adjecto*. I shall repeat a hundred times; we really ought to free our selves from the seduction of words! Let the people suppose that knowledge means knowing things entirely; the philosopher must say to himself: When I analyze the process that is expressed in the sentence, “I think,” I find a whole series of daring assertions that would be difficult, perhaps impossible, to prove; for example, that it is I who think, that there must necessarily be something that thinks, that thinking is an activity and operation on the part of a being who is thought of as a cause, that there is an “ego,” and, finally, that it is already determined what is to be designated by thinking—that I know what thinking is. [...] In place of the “immediate certainty” in which the people may believe in the case at hand, the philosopher thus finds a series of metaphysical questions presented to him, truly searching questions of the intellect; to wit: “From where do I get the concept of thing? Why do I believe in cause and effect? What gives me the right to speak of an ego, and even of an ego as cause, and finally ego as the cause of thought?” (BGE 16)

Here, Nietzsche points to the way in which apparently simple inner perceptions—perceptions of the fact that I am thinking, for example—have a host of presuppositions about the nature of agency, thinking, subjectivity, causality, and thinghood. Were my concept of agency different, Nietzsche emphasizes, I would not perceive my own thinking in the same way. My current conceptual repertoire influences even the most basic perceptions. And part of what genealogy does is uncovers these facts about the nature of my conceptual scheme.

The parallel, in Kant, would be knowledge of the Categories. The *Critique of Pure Reason* uncovers the conceptual structure of (so Kant claims) all self-conscious cognition. Precisely because this conceptual structure is intrinsic to all self-conscious cognition, it is not self-knowledge in the ordinary sense: uncovering this structure gives me knowledge not of my own idiosyncratic mental economy, but of the shared mental economies of all rational creatures. But suppose Kant is wrong: suppose the conceptual structure of conscious thought varies across different agents. Suppose, for example, that there is nothing inevitable about the experience of choice as undetermined by desire. Then a historical reconstruction of the ways in which certain conceptual schemes have come to seem inevitable will help us to see that these schemes are, in fact, contingent. And this will show us that Kant is more right than he realizes when he writes that “Without noticing what we are doing, we suppose we are discovering within us what we ourselves have put there” (*Anthropology*, AA 7: 133). For here Kant has in mind the instilling of false

motives and desires within ourselves. But for Nietzsche our self-invention goes far beyond this: the very conceptual schemes through which we experience the world, the schemes which structure our most basic understandings of ourselves and our relations to the world, are historically fluid. This is precisely the way in which the past flows through us in a hundred waves. And this is precisely the kind of self-knowledge that can be attained only by looking away from oneself.

4 Conclusion

Kant recognizes two distinct forms of self-knowledge: introspection and apperception. Introspection acquaints us with the sides of ourselves that Kant characterizes as passive: our sensations, affects, and so on. Apperception is knowledge of our own activities. It is a form of knowledge that arises from the fact that we in some sense create that which we know. Both modes of self-knowledge can go astray, and are particularly prone to being distorted by selfish motives; thus, neither is guaranteed to provide us with comprehensive self-knowledge.

I have argued that Nietzsche agrees with some aspects of this model. He, too, sees introspection and judgment as providing us with a limited degree of self-knowledge. However, he sees these capacities as far more limited than Kant acknowledges. In part, this is due to our psychic complexity: Nietzsche believes that it is a mistake to treat our actions as caused by discrete motives, decisions, and judgments. Instead, our actions are emanations from the “total state” of the organism. Introspection and judgment can do no more than scratch the surface of this total state, identifying only the most forceful and obvious motives while neglecting the many subtle, apparently minor influences.

But this disagreement over the scope of self-knowledge is not the deepest difference between Kant and Nietzsche. In Section Three, I argued that whereas Kant recognizes two forms of self-knowledge, Nietzsche seems to posit three. In particular, Nietzsche claims that we can acquire self-knowledge by *looking away* from ourselves. Section Three provided a brief sketch of two ways in which this might be so. In particular, genealogy enables us to identify some of the subtle factors shaping our actions as well as the influence of our current conceptual repertoires on our perceptions and understandings of our actions. For Nietzsche’s point is that the very way in which we conceptualize our deliberations, motives, and actions is influenced by our current conceptual scheme.¹²

¹² I owe great thanks to Maria João Branco, João Constâncio, Marta Faustino, Pietro Gori, Maria Filomena Molder, Paolo Stellino, Benedetta Zavatta, and other participants in the Nietzsche

References

- Constâncio, João (2011) “On Consciousness. Nietzsche’s Departure from Schopenhauer”, in: *Nietzsche-Studien* 40, 1–42.
- Kant, Immanuel (1996) *The Metaphysics of Morals*, ed. Mary Gregor, New York: Cambridge University Press.
- Kant, Immanuel (1998) *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor, New York: Cambridge University Press.
- Kant, Immanuel (1999a) *Critique of Pure Reason*, eds. Paul Guyer and Allen Wood, New York: Cambridge University Press.
- Kant, Immanuel (1999b) *Critique of Practical Reason*, eds. Paul Guyer and Allen Wood, New York: Cambridge University Press.
- Kant, Immanuel (1999c) *Religion within the Boundaries of Mere Reason*, eds. Allen Wood and George di Giovanni, New York: Cambridge University Press.
- Kant, Immanuel (2006) *Anthropology from a Pragmatic Point of View*, ed. Robert Louden, New York: Cambridge University Press.
- Katsafanas, Paul (2005) “Nietzsche’s Theory of Mind: Consciousness and Conceptualization”, in: *European Journal of Philosophy* 13, 1–31.
- Katsafanas, Paul (2013) “Nietzsche’s Philosophical Psychology”, in: K. Gemes and J. Richardson (eds.), *The Oxford Handbook of Nietzsche*, Oxford: Oxford University Press, 727–755.
- Katsafanas, Paul (2015) “Value, Affect, Drive”, in: Peter Kail and Manuel Dries (eds.), *Nietzsche on Mind and Nature*. Oxford: Oxford University Press. 163–188.
- Katsafanas, Paul (in progress) “The Moral Significance of Perceptual Experience”.
- Taylor, Charles (1992) *Sources of the Self: The Making of the Modern Identity*, Cambridge: Harvard University Press.
- Ware, Owen (2009) “The Duty of Self-Knowledge”, in: *Philosophy and Phenomenological Research* 79(3), 671–698.
- Williams, Bernard (1993) *Shame and Necessity*, Berkeley/Los Angeles: University of California Press.
- Wood, Allen (2003) “Kant and the Problem of Human Nature”, in Allen Wood (ed.), *Essays on Kant’s Anthropology*, Cambridge: Cambridge University Press.

International Lab, who offered many helpful comments and critiques during a presentation of this paper at the Universidade Nova de Lisboa.