Inner opacity. Nietzsche on Introspection and Agency

(Mattia Riccardi, MLAG/IF/FLUP, Porto)

1) The Newton of Psychology[1]

In one of his notes, Georg Christoph Lichtenberg once wrote: "Newton was able to separate the colors. Which name will the psychologist have who tells us what the causes of our actions are composed of? Most things, when they become noticeable to us, are already too big" (Lichtenberg 1867: 51; see 1973: *Heft* C 303, 213). It should come as no surprise that Nietzsche found these words appealing. We can easily imagine him seeing himself as the one saddled with the task of becoming the Newton of psychology envisaged by Lichtenberg.[2] But what does it actually take to accomplish such an ambitious project? A first point is explicitly made by Lichtenberg. As Newton showed that white light decomposes into different pure colors, his counterpart *in psychologicis* is expected to show how to break down the apparent unity of an action's cause into its relevant components. A second point, though not surfacing directly in Lichtenberg's note, can be easily appreciated by spelling out the suggested analogy with the project of modern physics of which Newton's *Principia* is the crowning work. What I have in mind is the kind of causal explanation employed by modern science in contrast to the teleological one common in the Aristotelian tradition. The naïve understanding we have of ourselves as agents is to a great extent teleological, as we usually explain our actions by naming their strived-for goals. Thus, it seems fair to assume that instead of teleological explanations of this sort, a would-be Newton of psychology should analogously provide an account of action couched in causal terms. In fact, Nietzsche offers a theory of action which aims at satisfying the two desiderata just outlined. On the one hand, he argues that the drives are the relevant components into which a certain

---

[1] Many thanks to André Itaparica and Brian Leiter for their helpful comments on a previous draft of this paper.

[2] Nietzsche partially underscored this passage and marked it with vertical lines on the left margin. His own copy of the book is now at the *Herzogin Anna Amalia Bibliothek*, which is part of the *Klassik Stiftung* in Weimar. See also Campioni *et alia* (2003): 354.

action ultimately breaks down. As Richardson (2004: 5, see also 35-6) puts it, drives are the "principal explanatory tokens" which figure in Nietzsche's view of agency.[3] On the other hand, the kind of explanation Nietzsche offers by appealing to the working of our drives is causal, rather than teleological.[4]

The point I would like to focus on in the present paper, however, emerges only in the last part of Lichtenberg's quoted note. There, he writes that "most things, when they become noticeable to us, are already too big". What exactly does he mean by this? Recall that Lichtenberg has just said that the main task of psychology is to individuate the components out of which actions are ultimately explained. Now, this indicates that he takes such components to be something which are hard to discover, for such a task requires nothing short of a Newton of psychology. It also suggests that the "things" Lichtenberg describes as being "already too big" when they "become noticeable to us" are the *causes of our own actions*. The sense in which the causes of our own actions, once we become aware of them, are "already too big" seems to be that they appear to us as some kind of *simple* events, or states, whereas they in fact result from the interplay among different causal components—those that psychological inquiry is supposed to uncover. Events or states of this kind constitute episodes of willing—what philosophers usually call volitions. Thus, what Lichtenberg is saying is that the experience we have of episodes of this sort is such that it does not reveal to us how our own actions are actually produced.

Nietzsche agrees with Lichtenberg also on this third skeptical point about our self-experience as agents. In aphorism 116 from *Daybreak* he writes that the belief "that one knows, and knows quite precisely in every case, *how human action is brought about*" is but a "primeval delusion" (D 116). Importantly, this claim is motivated by appeal to a much broader skepticism which targets introspection as such: In my view, Nietzsche's endorsement of it derives from his commitment to a much broader skepticism which targets introspection as such—a position he expresses in a note from 1888 as follows:

---

[3] The same claim is also put forward by Katsafanas (forthcoming).
[4] That Nietzsche's idiom should be understood as causal is argued by Leiter (2002) and Risse (2008). For disagreement, see Clark, Dudrick (2012). For a general defense of a Nietzschean approach to moral psychology, see Knobe, Leiter (2008).

> "Actions are *never* what they appear to us to be! We have expended so much labour on learning that external things are not as they appear to us to be—very well! the case is the same with the inner world!" (*ibid.*)

In a note from 1888, this position is described as a version of "phenomenalism" about inner experience:

> "We don't have to look for phenomenalism in the wrong place. Nothing is more phenomenal, (or more clearly) nothing is more of an illusion than this inner world we observe with the famous 'inner sense'". (KSA 13: 14[152])[5]

Phenomenalism is usually a theory about the external world. It says that epistemic warrant is confined to what philosophers sometimes call *seemings*. For instance, according to this position, I can be certain that it *seems* to me that there is an apple there on the table, not however that such a fact actually obtains out in the world. Standard phenomenalism, however, grants—in fact, presupposes—that we are, in some way or another, directly aware of (at least some of) our mental states. This, however, is the very assumption Nietzsche rejects in the quoted note. To have a handy label, I shall refer to the view he thereby defends as *Inner Opacity* (IO). The remainder of the paper is dedicated to the spelling-out of this position by focusing, in particular, on how it bears on Nietzsche's view of agency.

In section 2, I shall start by briefly introducing Nietzsche's notion of consciousness. Given that we engage in introspection consciously, this first move will help us set the stage. In sections 3 and 4, I shall spell out in more detail his general view on introspection. The main idea will be that (human) self-knowledge has evolved together with (human) mind-reading capacities, that is, our capacity to tell what somebody else is thinking (feeling, believing etc.). In short, Nietzsche does not think that self-knowledge is in some sense *primitive* or even more *immediate vis-à-vis* other-knowledge. Though this view may at first seem startling, it just takes a little reflection to see that something along these lines is in fact entailed by (IO). (Of course, this hardly suffices to make the claim more palatable.) Finally, I shall turn to Nietzsche's theory of

---

[5] See also: "Critique of modern philosophy: erroneous starting point, as if there were 'facts of consciousness' ('*Thatsachen des Bewusstseins*')—and no phenomenalism in self-observation", N 1887, 2[204]. (This passage appears in a plan for the fifth book of *Gay Science*. As we shall see in the next section, the position is indeed put forward in aphorism 354 of this work.) Without explicit use of the label "phenomenalism" the same thesis is put forward also in N 1888, 15[90].

agency. I shall start by briefly illustrating his notion of drive and the role it plays in the account he offers of psychological states (section 5). By building on the theses defended in the first part of the paper, I shall argue that Nietzsche sees a fundamental discrepancy between the linguistically-articulated conscious attitudes to which we have introspective access, on the one hand, and the sub-personal processes and states occurring at the unconscious level of the drives, on the other. With this general picture set, I shall consider, in particular, Nietzsche's view of the drives as the constituents of those volitional states which typically issue in actions (section 6). By way of conclusion, I shall briefly assess the consequences that Nietzsche's views have for the function of folk-psychology and the status of conscious attitudes (section 7).

2) Introspection, Consciousness and Language

Introspection is a capacity which is consciously exercised. It is therefore helpful to start our investigation of (IO) by taking a look at Nietzsche's view of consciousness. Interestingly, the only published passage where he explicitly uses the label "phenomenalism" in talking about inner experience is in a long aphorism where he provides the most detailed picture of consciousness to be found in his work:

> "At bottom, all our actions are incomparably and utterly personal, unique, and boundlessly individual, there is no doubt; but as soon as we translate them into consciousness, *they no longer seem to be*… This is what *I* consider to be true phenomenalism and perspectivism: that due to the nature of *animal consciousness*, the world of which we can become conscious is merely a surface- and sign-world, a world turned into generalities and thereby debased to its lowest common denominator, […] that all becoming conscious involves a vast and thorough corruption, falsification, superficialization and generalization […]."
> (GS 354)

According to the view Nietzsche introduces here as the "true phenomenalism" (and "perspectivism"), the "world" to which we have conscious access is, in some sense, deeply falsified. Thus, the kind of "phenomenalism" in question turns out to be primarily motivated by Nietzsche's further views on consciousness—which is the main

4

concern of the long aphorism from which the quoted passage stems. Which features of consciousness then make (IO) true in Nietzsche eyes?[6]

To start answering this question, we need to first consider which notion of consciousness Nietzsche is concerned with. Previously in the same aphorism, he claims that "*consciousness in general has developed only under the pressure of the need to communicate*", so that "the development of language and the development of consciousness […] go hand in hand" (*ibid.*). At first sight, this claim looks quite puzzling, for most of the senses in which we may read "consciousness" would just render it utterly false. For instance, it surely seems wrong to claim that phenomenal consciousness is in any sense language-dependent: to experience redness one is not required to be the speaker of a language having an equivalent of the English word "red". Similarly, it seems wrong to say that conscious perception presupposes linguistic ability, as we can say of a cat that it sees—or, perhaps more plausibly, smells–the sausage as Maria takes it out of the fridge. To my lights, the only plausible candidate here is self-consciousness, or something very close to it. For, contrary to the other meanings one may attach to the term, self-consciousness can be reasonably said to be language-dependent in that it requires the capacity to self-refer—a capacity we acquire by learning how to use the first-person pronoun.[7]

More importantly for our present purpose, in saying that consciousness is intimately related to language and to our communicative practices, Nietzsche arguably suggests that *un*conscious cognition is non-linguistic in nature. More precisely, whereas conscious mental states are linguistically (propositionally) articulated, the kind of content which is fed to sub-personal processes occurring at the unconscious level is not linguistic (propositional).[8] Thus, according to Nietzsche, the mental content of a certain state gets articulated linguistically (propositionally) only as it becomes conscious. Crucially, here is where (IO) sets in, for Nietzsche takes this process of conscious

---

[6] In what follows, I rely on a reading of Nietzsche's view on consciousness I defend in more detail in my (forthcoming-a).

[7] The fact that Nietzsche qualifies the kind of consciousness he is concerned with also as "*animal*" may seem to speak against my reading of this notion in terms of self-consciousness. However, the real puzzle is how to make sense of the apparently contradictory claim that the relevant kind of consciousness, irrespective of how we are to understand it, is qualified as animal and linguistic alike. I shall come back to this puzzle later in the paper.

[8] Katsafanas (2005) proposes a similar view. There is, however, an important disagreement between Katsafanas and me regarding the possibility of unconscious conceptual content, which he rejects and I allow for. For a detailed discussion, see my (forthcoming-a) as well as Katsafanas' response in his contribution to this *Special Issue*.

articulation to be responsible for our inner world's being "merely a surface- and sign-world", as he puts it. However, several points still need to be addressed in order to appreciate why he so believes.

To bring to light some of these points, it will be helpful to take a closer look at the story Nietzsche tells about how consciousness evolved out of the "*need to communicate*". He writes that, "as the most endangered animal, he [the man] *needed* help and protection, he needed his equals; he had to express his neediness and be able to make himself understood—and to do so, he first needed 'consciousness', i.e. even to 'know' what distressed him, to 'know' how he felt, to know what he thought" (GS 354). Suppose we were living in a primitive society, exposed to threats of all sorts and with utterly sparse resources and fragile skills to face them. The advantages, which in such a situation derive from one's belonging to a social group, are based on the capacity for mutual communication. In particular, this requires both that one be able to tell others what the content of one's mental states is and that others be able to understand what one is thereby saying. The first requirement can be met only if one is in a position, as Nietzsche has it, to "'know' what distressed him, to 'know' how he felt, to know what he thought", i.e. if one has some kind of epistemic access to the contents of one's mind. But this, Nietzsche argues, is precisely what it means to be conscious of them. The second one is satisfied only if the parties involved in communication make use, *inter alia*, of a shared set of mental terms.[9] Crucially, Nietzsche holds that the relevant psychological categories govern not only our practice of ascribing mental states to others, but also the way in which we conceptualize our own inner states.

From this a crucial point follows about the relation between the mental vocabulary we acquire by means of social interaction and the epistemic access we have to our own mind. For the latter is, according to Nietzsche, shaped by the former. This means that the access we have to ourselves is not *direct*, but rather *mediated* by whatever folk-psychological framework we learn from the surrounding environment. Given his endorsement of (IO), it should not surprise us that Nietzsche holds the access we have to our own minds to be *indirect*. However, to spell out the details of his view on introspection, it will be important to understand better how a socially acquired folk-

---

[9] As we shall see, this is not entirely correct, for Nietzsche argues that there are also non-verbal means by which one can communicate one's own states, like expressions or gestures. This also motivates the "*inter alia*" qualification.

psychological framework is supposed to impact on self-knowledge. As it will turn out, Nietzsche thinks that the same set of cognitive capacities enables both introspective self-knowledge and outward-directed mind-reading. In the following two sections, I shall investigate how he substantiates this general claim. I shall start by considering how minimal forms of self-knowledge and mind-reading co-emerge in the animal realm. Then, I shall address the role Nietzsche ascribes to language in order to explain the peculiar human form of these capacities. In both cases, I shall focus in particular, though not exclusively, on some aphorisms from *Daybreak*, where, as it seems to me, some of the views expounded in GS 354 are anticipated.

3) Animal Mirrorings

Aphorism 26 from *Daybreak*, whose title is "Animals and Morality", starts by noting that "[t]he practices demanded in polite society: careful avoidance of the ridiculous, the offensive, the presumptuous, the suppression of one's virtues as well as of one's strongest inclinations, self-adaptation, self-deprecation, submission to orders of rank— all this is to be found as social morality" (D 26). For instance, "the animals learn to master themselves and alter their form, so that many, for example, adapt to the coloring of their surroundings […], pretend to be dead or assume the forms and colors of another animal or of sand, leaves, lichen, fungus (what English researchers designate 'mimicry')" (*ibid.*). The point of comparing phenomena of chromatic mimicry with the social practices described at the very beginning of the aphorism is that of highlighting how both constitute cases of adaptation to the environment. However, this way of putting things may easily obscure one of Nietzsche's central claims. What I have in mind is the thesis—partially revealed by his mentioning the animals' need to "master themselves"—that some of those cognitive abilities which co-evolved together with culture are already prefigured in the animal realm. In particular, and crucially, Nietzsche refers to the capacity of self-knowledge. As he writes, the animal "assesses the effect it produces upon the perceptions of other animals and from this learns to look back upon itself, to take itself 'objectively', it too has its degree of self-knowledge" (*ibid.*). At the same time, the animal confronting other animals displays some rudimentary form of mind-reading. For it not only "assesses the movement of its friends and foes", "learns

their peculiarities by heart", but "can likewise divine from the way their approach that certain kind of animals have peaceful and conciliatory intentions" (*ibid.*).

First, note that this helps us solving the puzzle raised by Nietzsche's qualification of consciousness as "*animal*" in GS 354.[10] As you might recall, many of the features he ascribes to consciousness in that aphorism indicate that the notion he has in mind is that of self-consciousness. In particular, Nietzsche claims that consciousness bears some intimate relation to language. Why, then, does he refer to this kind of consciousness as "*animal*"?

The picture sketched in D 26 suggests that elementary forms of self-knowledge—perhaps, one should better say, self-awareness—are already present in the animal realm in consequence of a certain animal being confronted with other animals—either conspecific or not. Arguably, such effects not only are stronger in species whose individual members live in groups, but also grow together with the increasing complexity of the social interactions within such groups. Now, in GS 354, Nietzsche seems to argue that this last process depends on, or at least is immensely facilitated by, practices of non-verbal communication. As he notes, "not only language serves as a bridge between persons, but also look, touch and gesture" (GS 354).[11] Thus, he seems to suggest that the development of proprietary human, language-dependent, forms of self-consciousness builds on communication skills already to be observed among animals: "only as a social animal did man learn to become conscious of himself" (*ibid.*). Therefore, in qualifying *our* consciousness as animal Nietzsche highlights its ancestral provenance. Of course, this is not to deny that the emergence of language determines quite an astounding qualitative gap between the way in which we are conscious of ourselves and the way in which animals are. Indeed, Nietzsche holds that what most characterizes our way of being (self-)conscious is precisely its linguistic nature.

More relevant is how the picture Nietzsche proposes in D 26 suggests that mind-reading and self-knowledge relate. There he argues that the individual animal *first* learns to attribute mental states—perceptions, in the example—to other individuals and *then* "from this learns to look back upon itself". Thus, self-knowledge is not prior to other-knowledge. Rather, it is arrived at only after one has learned to turn on oneself

---

[10] See also footnote 7 on this.
[11] On the relation between gesture and language, see already HUH 216.

whatever mind-reading abilities one usually employs to interpret the behavior of other individuals.[12] But how does this happen?

Once again, recall that according to Nietzsche the process by which consciousness evolved is tied to the "*need to communicate*". This suggests that more primitive forms of consciousness correlate with more primitive forms of communication. As we have just seen, gestures—as well as facial expressions—may well play the relevant role here. However, to be effective, communication also presupposes the possibility of mutual understanding, which in turn requires one to be in a position to correctly "read" the behavior of other individuals. For instance, to make sense of your pointing at that cup of coffee over there on the kitchen table, I need to be able to interpret your movements as realizing your intention of drawing my attention to the relevant object. In his work from the so-called middle period, Nietzsche sometimes appeals to imitation as the mechanism performing this cognitive task. In an aphorism on the topic of "Empathy (*Mitempfindung*)", he writes for instance:

> "To understand another person, that is, to *imitate his feelings in ourselves*, we do indeed often go back to the *reason* for his feeling thus or thus and ask for example: *why* is he troubled? — so as then for the same reason to become troubled to ourselves; but it is much more usual to omit to do this and instead to produce the feeling in ourselves after the *effects* it exerts and displays on the other person by imitating with our own body the expression of his eyes, his voice, his walk, his bearing (or even their reflection in word, picture, music)." (D 142)

---

[12] In his comments on a previous draft of this paper, André Itaparica questions my reading of this passage from D 26. As he rightly stresses, the aphorism is not primarily concerned with self-knowledge, but rather with other-knowledge. Given this broad context, Nietzsche's remark on self-knowledge could be interpreted as the claim that *only a certain kind* of self-knowledge depends on outer-directed mind-reading capacities. In particular, according to this alternative proposal, Nietzsche would simply be arguing that an animal learns to behave prudentially by taking into account how other animals reacted to its own previous behavior and that this capacity enables it to adopt an "objective stance" toward itself. This, however, is fully compatible with the existence of *other forms* of immediate or, at least, more basic self-knowledge which does not depend on such a capacity. It seems to me, however, that Nietzsche's text favors my reading, as he writes that the animal in question has "its degree of self-knowledge" (D 26) in that it displays precisely the mind-reading skills just described. Thus, he seems to be saying that this, and only this, is the kind of self-knowledge available to it. To this, one could reply that it seems highly implausible that an animal could know that its leg hurts only once it has developed some—however minimal—mind-reading ability. There is something true to this objection, but I think that the issue turns on the notion of "self-knowledge" we are concerned with. One can surely allow that an animal has some kind of nonconceptual self-awareness of bodily states like pain which is independent from outer-directed mind-reading, and still contend, as Nietzsche does, that genuine self-knowledge requires one to engage in communicative practices and social interactions.

Here, Nietzsche distinguishes two different ways in which imitation may occur. The first one looks like an instance of simulation carried out reflexively. The second corresponds to a case of what is nowadays called emotional mirroring or contagion[13],— a phenomenon usually taken to occur unconsciously. Whereas the latter is widespread in the animal realm as well, the former constitutes a human-specific mind-reading capacity.

Nietzsche accounts for "how we become so fluent in the imitation of feelings" by appeal to our "subtle and fragile nature" as well as to our "*timidity*", which constitutes what positively acts as the "instructor in that empathy, that quick understanding of the feelings of another (and of animals)" (D 142). This is the same picture he will later provide in GS 354, where, as we have seen, the human being is depicted as "the most *endangered* animal". Whereas this feature explains why our hominid ancestors joined into primitive societies, the social pressures thereby established determined the development of ever more complex mind-reading skills, like e.g. the kind of reflexive simulation described in the passage just quoted.[14] In Nietzsche's eyes, the parallel emergence of ever more sophisticated forms of self-knowledge is a by-product of this process. Here, however, is where language enters the scene.

4) The Linguistic Shape of the Inner

To fully appreciate the role Nietzsche ascribes to language in shaping the access we have to our own mind, it is again helpful to look at *Daybreak*. There, he argues that "[l]anguage and the prejudices upon which language is based are a manifold hindrance to us when we want to explain inner processes and drives: because of the fact, for example, that words really exist only for *superlatives* degrees of those processes and drives" (M 115). Nietzsche names instances of such "*extreme* states" as "[a]nger, hatred, love, pity, desire, knowledge, joy, pain"; actually, he goes on, "even the most moderate *conscious* pleasure or displeasure, while eating food or hearing a note, is perhaps, rightly understood, an extreme outburst" (*ibid.*). From this, he concludes that

---

[13] The two terms are used, for instance, by Hurley, Chater (2005: 41).
[14] See Carruthers 2012: 211 for a similar story

"[*w*]*e are none of us* that which we appear to be in accordance with the states for which alone we have consciousness and words, and consequently praise and blame" (*ibid.*).[15]

Three points are worth noting. First, we already find here the basic idea of (IO), namely, that one is not the way one "appears" to oneself in introspection. Second, this fact is said to directly depend on the linguistic nature of the conscious access we have to our own mind. As Nietzsche has it, we become aware of those states "for which alone we have consciousness and words". Third, he argues that the states picked out by the mental vocabulary at our disposal are "*extremes* states" or "*superlatives* degrees" of processes which—as implicitly assumed in the quoted passage—remain for the most part unconscious. Clearly, Nietzsche takes this fact to show that our mental vocabulary is, in some deep sense, inadequate. However, though one might well agree with him on this point, the reasoning offered so far falls short of substantiating such a strong claim as (IO). More needs to be said here.

First, some support for Nietzsche's claim comes from reflection on the richness of our mental life. This point is nicely captured by James Sully, an English psychologist who lived in the second half of the 19th century and with whose work Nietzsche was acquainted. In particular, in a passage Nietzsche seems to have read,[16] Sully notes that "a state of consciousness at any moment is an exceedingly complex thing. It is made up of a mass of feelings and active impulses which often combine and blend it on a most inextricable way" (1882: 196). Moreover, "many of these ingredients are exceedingly shadowy, belonging to that obscure region of sub-consciousness which it is so hard to penetrate with the light of discriminative attention" (197).

Sully's picture seems quite right. Now, according to Nietzsche the way in which one's folk-psychological vocabulary constrains one's introspective exercises is, at least in part, responsible for our failing to attend to the richness and complexity of our mental life. As he argues, "where words are lacking, we are accustomed to abandon exact observation because exact thinking there becomes painful" (M 115). Once one has

---

[15] See also: "*Words present in us*.—We always express our thoughts with the words that lie to hand. Or, to express my whole suspicion: we have at any moment only the thought for which we have to hands the words" (D 257).

[16] A copy of the French translation of Sully's *Illusions* is to be found in Nietzsche's library (see Sully 1883). In particular, it seems that he read the chapter entitled "Illusions of Introspection", from which the following quotes are taken. See Campioni *et alia* (2003): 582. Interestingly, Sully's book was anonymously translated by Henri Bergson.

conceptualized the mental state one is in as an episode of "anger", "joy", or whatever, the task of self-knowledge is usually taken to have been accomplished successfully. Furthermore, as argued by Eric Schwitzgebel, we have quite an insecure grip on the psychological notions we employ on day-to-day basis as well as to the kind of phenomenology associated to the relevant states. About the first aspect, he notes:

> "I don't know what emotion is, exactly. Neither do you, I'd guess. Is surprise an emotion? Comfort? Irritability? Is it more of a gut thing, or a cognitive thing? Assuming cognition isn't totally irrelevant, how is it involved? Does cognition relate to emotion merely as cause and effect, or is it somehow, partly, constitutive? I'm not sure there's a single right answer to these questions." (Schwitzgebel 2008: 249)

Things are hardly better with regard to phenomenology:

> "You've had emotional experiences, and you've thought about them, reflected on how they feel as they've been ongoing or in the cooling moments as they fade. If such experiences are introspectible, and if introspection is the diamond clockwork often supposed, then you have some insight. So tell me: Are emotional states like joy, anger, and fear always felt phenomenally—that is, as part of one's stream of conscious experience—or only sometimes? Is their phenomenology, their experiential character, always more or less the same, or does it differ widely from case to case? For example, is joy sometimes in the head, sometimes more visceral, sometimes a thrill, sometimes an expansiveness—or, instead, does joy have a single, consistent core, a distinctive, identifiable, unique experiential character?" (*ibid.*)[17]

Nonetheless, we typically feel the urge neither to pursue laborious phenomenological investigations about our own inner states, nor to critically assess the serviceability of the mental categories we make use of. We just swiftly apply them to whatever state we introspectively attend to. Though considerable practical advantages arguably derive from this, we now start to appreciate why Nietzsche thinks that the peculiar linguistic

---

[17] Schwitzgebel's list of complex questions about emotional phenomenology goes on, but his point should be clear by now.

form under which mental content becomes introspectively accessible belongs to "merely a surface- and sign-world" (GS 354).[18]

Second, a more far-ranging consequence arises from the communicative purposes underlying the emergence of the shared mental vocabulary we adopt to conceptualize our own inner experience. What Nietzsche has in mind we might call *inner one-sidedness*. As he notes in an aphorism which is closely related to GS 354, "[u]sing the same words is not enough to get people to understand each other: they have to use the same words for the same species of inner experiences too; ultimately, people have to have the same experience *base*" (BGE 268). Here, the problem addressed is similar to that usually discussed in contemporary philosophy under the label of "private language". In order for communication to be possible, the language a speaker utilizes cannot be "private". Rather, the meaning of the words uttered by a speaker has to be publicly graspable. But how is that possible in those cases where the word purports to refer to an inner state? Nietzsche's answer to this problem is that people belonging to a certain social group can successfully communicate their own inner states only in virtue of their coming to have states of the same type.[19] In turn, this kind of psychological attunement among the members of a given community is realized within the very communicative practices they undertake. In particular, as cognitive anthropologist Michael Tomasello has recently argued, it seems that—among the "basic motives" to be individuated behind the evolution of human communication—"sharing attitudes and emotions" is the one playing the relevant "group-identity function" (Tomasello 2008: 210-11). As he illustrates the point, "my goal in expressing my enthusiasm is not, as in the case of informatives, to provide you with information that you want or need, but rather to elicit from you an expression of attitude that aligns with mine. When we feel the same about some common experience, this makes us feel psychologically closer" (211).

---

[18] Schwitzgebel claims that the questions he raises "challenge us not simply because we struggle for the *words* that best attach to a patently obvious phenomenology" (2008: 250). The point is, rather, that "in the case of emotion the *very phenomenology itself* […] is not entirely evident" (*ibid.*). Therefore, "[i]t's not just *language* that fails us […], but introspection itself" (*ibid.*). Though Nietzsche seems to see a closer relation between the poverty of our psychological vocabulary and phenomenological inaccuracy, he too thinks—as we shall see—that there is more to introspection's failure than just this.

[19] That the problem Nietzsche is concerned with here is that of "private language" has already been noted by Danto (2005: 102-5). However, I do not agree with Danto that Nietzsche's solution to this problem is, like that of mid-20th century analytical philosophers, to merely endorse a version of semantic externalism.

This process has, of course, a clear evolutionary advantage, as it enormously facilitates the mutual understanding between individuals belonging to the same group. According to Nietzsche, however, it also has a significant drawback on self-knowledge. Recall that according to his view self-knowledge exploits the same kind of folk-psychological framework which guides our mind-reading capacity. This means that the mental categories underlying the way in which we conceptualize our own inner states are precisely the same as those governing our other-ascriptions of mental states. Thus, given that the relevant framework is shared by the members of a certain community, their mental life will tend to become uniform and consistent with one another. As Nietzsche puts it, we are "continually as it were *outvoted* and translated back into the herd perspective" (GSA 354).

So far, Nietzsche has provided two main points in support of (IO) which are both related to the fact that we conceptualize our own mental states according to a given folk-psychological vocabulary. First, such a vocabulary is highly inaccurate as it does not reflect the phenomenological richness and complexity of our inner life. Second, the way in which each member of a certain community conceptualizes her own states tends to conform to the shared folk-psychological vocabulary. Thus, the space left for something like genuine individuality is minimal. The relevant question, however, is whether these two points suffice to substantiate (IO). To my lights, it is fair to be in doubt about this. Nietzsche's case becomes more convincing only once we factor in the explanatory role he reserves to the drives. This is the topic I shall now turn to.

5) Introducing the Drives

According to Lichtenberg's note quoted at the very beginning of this paper, the main task for the future Newton of psychology is that of individuating "what the causes of our actions are composed of". As I anticipated in section 1, Nietzsche's attempt to satisfy this desideratum consists in positing the drives as the basic explanatory tokens to figure in a genuine account of how actions are produced. But what is exactly a drive?

Nietzsche's notion of drive has been amply discussed in recent literature. Clark and Dudrick (2012) have proposed a homuncularist reading inspired by the work of Daniel Dennett. Richardson (2004) and Katsafanas (forthcoming) argue that

Nietzschean drives are best understood in dispositionalist terms. As this is not the proper place to go into the details of the debate, I shall simply assume that Nietzsche holds two quite general theses about the drives.[20]

First, I assume that drives can be suitably described in broadly *functional* terms. Two main reasons speak in favor of this claim. To begin with, it is suggested by some of the scanty descriptions Nietzsche offers of the drives. Furthermore, as it seems to me, it should be agreed upon by the proponents of both the homuncularist approach as well as the dispositional one. On the one hand—to take the latter view first—in saying that drives are dispositions typically issuing in actions of a certain kind one is offering a functional characterization of them. On the other hand, Dennett's brand of homuncularism is explicitly committed to functionalism about the relevant cognitive systems.

Second, I assume that drives are mental in that they are—perhaps in some minimal sense—intentional. In other words, they are directed towards certain aspects of reality and can consequently impact our experience of it. For instance, as Katsafanas (forthcoming) suggests, drives can make a determinate feature of reality become perceptually salient by directing my attention to it.[21] Again, it seems to me that this is a claim we can uncontroversially ascribe to Nietzsche.

Given this sketch of Nietzsche's conception of the drives, let us now turn to a point which is more germane to our present concern. Once again, it nicely emerges from what Nietzsche writes in aphorism 115 of *Daybreak*. As you might recall, there he writes that "[l]anguage and the prejudices upon which language is based are a manifold hindrance to us when we want to *explain* inner processes and drives" (D 115, my italics). This time, I would like to focus more narrowly on Nietzsche's claim that our use of folk-psychological terms prevents us from formulating suitable *explanations* of our "inner processes" and "drives"—that is why the word "explain" appeared italicized in this last rehearsal of the passage. In my view, Nietzsche is thereby pointing out a kind of substantial *discrepancy* between the mental states picked out by our current mental

---

[20] I say something more—still quite sketchy, though—on Nietzsche's notion of drive in my (forthcoming-b).

[21] A similar view is put forward also by Clark, Dudrick (2012: 145-46). See, however, Richardson (2004: 36), who denies that drives are mental.

vocabulary and those states and processes that should on the contrary appear in genuine psychological explanations. This needs some unpacking.

Nietzsche argues that the members of a given community tend to conceptualize their conscious states in conformity to the folk-psychological vocabulary they acquire through social interaction. Such vocabulary is typically quite impoverished, as it allows us to conceptualize, and consequently become introspectively aware of, only "*extreme*" and "*superlative*" inner states. However, folk-psychology's constraints on self-knowledge are not confined to its providing us with mental categories which are meager and phenomenologically inadequate. For we also acquire a *set of beliefs* about the *rol*e each of the relevant state-type typically plays in the overall economy of one's mind, in particular with regard to the causation of one's actions. At a general level, we learn, for instance, how beliefs, desires and emotions are supposed to interact to form intentions and how these usually issue in actions. At a more fine-grained level, we learn how to see a certain type of behavior as caused by a certain type of mental states—one's tapping on the desk as caused by one's being nervous, or one's refraining from eating a certain sort of food from one's holding a certain religious belief. Note that for Nietzsche this is reflected in the way in which we are conscious of ourselves. On the one hand, the mental states we are conscious of are the kind of linguistically (propositionally) articulated states postulated by folk-psychology. More specifically, according to him the Jamesian "stream of consciousness" we are introspectively aware of is a succession of *discrete* and linguistically (propositionally) articulated states. On the other hand, we naively conceive of the relations obtaining between the states we introspectively access in accordance to what our socially-acquired theory of the mind tells us about their specific role.

Contrary to the picture suggested by self-experience, Nietzsche maintains that what actually goes on is the *continuous* interplay between our drives occurring at the unconscious level. His position is exemplified by aphorism 333 of *Gay Science*.

"*What knowing means.—Non ridere, non lugere, neque detestari, sed intelligere!* says Spinoza as simply and sublimely as in his wont. Yet in the final analysis what is this *intelligere* other than the way we become sensible (*fühlbar*) of the other three? A result of the different and conflicting drives to laugh, lament, and curse? Before knowledge is possible, each of these impulses must first have

16

presented its one-sided view of the thing or event; then comes the fight between these one-sided views, and occasionally out of it a mean, an appeasement, a concession to all three sides, a kind of justice and contract […]. Since only the ultimate reconciliation scenes and final accounts of this long process rise to consciousness, we suppose that *intelligere* must be something conciliatory, just, and good, something essentially opposed to the drives, when in fact *it is only a certain behavior of the drives towards one another*." (GS 333)

In this passage, Nietzsche focuses on a single state of mind, namely knowledge.[22] He starts by considering Spinoza's view according to which one's knowing something requires one to suspend certain affective attitudes towards that thing and then argues that this view is wrong. First, he maintains—indeed, quite to the contrary—that rather than being independent from the drives indicated by Spinoza, the psychological state we refer to as "knowledge" is in fact *constituted* by a certain relation among them. Second, he argues that we take that state to be "opposed to the drives" as a consequence of our becoming aware "only [of] the ultimate reconciliation scenes and final accounts of this long process". Therefore, we falsely conceive of knowledge as a purely cognitive and discrete mental state, whereas it results instead from the continuous and affective-loaded interplay between certain drives.

The picture Nietzsche offers of the specific case of knowledge generalizes, I submit, if not to the entirety of our mental states, at least to the linguistically (propositionally) articulated attitudes, as beliefs, desires, emotions—for such are the states which are (self-)conscious in the Nietzschean sense of the term. Therefore, when he writes elsewhere that "thinking (*Denken*) is only a relation between [the] drives" (BGE 36), we should understand "thinking" as a general term which is intended to cover, at least, all attitudes of this kind.

As our main concern is how drives contribute to Nietzsche's account of action, this is the issue to which I shall turn next. An important point will be the discrepancy between the causal role the drives display in determining our actions and the kind of experience we have of ourselves as agents. In particular, to appreciate how (IO) applies to agency, it will be helpful to focus on his view of volitional states.

---

[22] Here, Nietzsche does not seem to be interested in knowledge as an epistemic state, but rather as a merely psychological one. This seems fair to Spinoza's own definition.

6) The Introspecting Agent

Nietzsche thinks that the drives are the primary explanatory tokens we should appeal to in accounting for our actions. More specifically, he believes that the volitional states which usually issue in actions—willings as I shall call them—are constituted by a certain interplay between one's drives. The kind of relation he takes to hold between the drives is one of command and obedience, as suggested by the following note:

> "Each thought, each feeling, each willing (*Wille*) is not born out of one determinate drive, but is rather a total state (*Gesamtzustand*), the overall surface of the whole consciousness, and results from the momentary power-arrangement (*Macht-Feststellung*) of all drives which constitute ourselves—just of the dominant drive and of those obeying and disputing it alike." (KSA 12: 26)

Again, note that the model on offer applies to psychological states of quite a different sort. However, here Nietzsche also explicitly names volitional states. Two features of his characterization of such states are most relevant to our discussion. We are told that a willing is to be conceived of, first, as a "total state" resulting from the specific "power arrangement of all drives" and, second, as occupying the "overall surface of the whole consciousness". **W**hat Nietzsche means by this is far from clear. My proposal is that he is offering a model according to which a conscious willing, though it actually results from the interplay between one's unconscious drives, is experienced by the subject as if it were a simple—in the sense of uncomposed—state. Moreover, his saying that such a willing occupies the "overall surface of the whole consciousness" conveys the idea that it is, at that moment, the only element of our inner life we are consciously aware of. Thus, what becomes conscious are not the drives as such, but rather the "total" state constituted by their current arrangement. This picture, though crucial to the issue of agentive introspection, is in need of further elucidation.[23]

In the previous section, I argued that for Nietzsche the folk-psychological framework we acquire through social interaction impacts in two fundamental ways on self-knowledge. First, the access we have to our own mind is constrained by the reservoir of mental terms provided by the framework. Second, one assumes that one's

---

[23] A third important feature of Nietzsche's characterization is the claim that the arrangement between one's drives which constitutes a certain willing has some drive dominating the other ones. As this point has no bearing on the problem being discussed in this paper, I shall simply ignore it. This aspect of Nietzsche's view has recently received a thorough treatment by Clark, Dudrick (2012).

(conscious) mental states play the role which is typically ascribed to them by the relevant framework—both at a general and a more fine-grained level. This, of course, applies to willings as well.

The folk-psychological role we conceive willings as having is obviously tightly related to action. Suppose I get up, go to the fridge and take a beer out of it. Were I to explain my action, I would say that I acted the way I did because I wanted to drink a beer—whereby my wanting to do such-and-such constitutes the relevant state of willing. This is to say that willings are the folk-psychological explanatory tokens we usually appeal to when asked to account for our own actions. In short, we naïvely conceive of willings as the proximal *causes* of our own actions.

First, Nietzsche argues that the will, or, more precisely, a case of willing, is a "multifarious thing that people have only one word for" (BGE 19). Contrary to what is suggested by our undifferentiated way of talking, he maintains that actual willing is "something *complicated*" as it both involves mental elements of a different sort and results from the interplay between one's drives—as Nietzsche writes in the *Nachlass* passage quoted above, it is a "total state".[24] However, given that the conscious experience of our own willings is shaped by our linguistic practice, their real nature remains introspectively inaccessible. In other words, we simply fail to see that our willings are constituted by the complex and continuous interplay between our (unconscious) drives, and take them to be instead simple and discrete (conscious) states. In this sense, the fact that we become aware of our volitional states only in a linguistically-mediated way is but an instance of the general fact that the reach of consciousness goes hand in hand with that of language, as Nietzsche argues in D 115 and GS 354.

Second, and more relevantly in the way in which we are conscious of our volitional states is embedded a wrong picture of how our actions are caused. Here we finally come to the Lichtenbergian point I started with. As Lichtenberg puts it, the cause of a certain action is already "too big" when we become aware of it. What appears to one as the cause of one's action is a conscious willing, i.e., according to Nietzsche, a (linguistically) propositionally articulated volitional state. His way of putting

---

[24] In BGE 19, Nietzsche offers a view of the will which aims at capturing this complexity. For two alternative readings see Leiter (2007) and Clark, Dudrick (2012).

Lichtenberg's point, thus, is that self-experience, by presenting states of this sort as the *causes* of our actions, does not reveal those elements which are in fact causally efficacious. Such elements, as it should now be clear, are the drives operating at the unconscious level.

One could argue that this view does not take into account the fact that at least some of our willings *appear* to us to be in fact "something *complicated*", as they are arrived at only through a laborious process of (conscious) deliberation. In those cases, we try to figure out the possible outcomes of the alternative actions we could perform, to weigh the situation we are in by taking different points of view on it, etc. Thus, we may have a picture of our coming to want to do such-and-such as a process which involves conflicting *conscious motives*. Furthermore, such motives may be seen as the relevant explanatory tokens to which one may appeal to in accounting for one's actions. On this basis, one could claim that the conscious access we have to the processes constituting an instance of deliberation provides us with a reliable clue as to the nature of both the involved explanatory tokens—conscious motives—and of the transactions between them. Were this true, we would be left with no real motivation for any substantive (IO) claim, as the way things appear to us in introspection would indeed offer an accurate enough picture of what a volitional state is and how it gets formed.

Nietzsche, however, thinks that this is *not* true. He holds that the transactions between our drives are quite different from those we take to obtain between conscious motives. In an aphorism from *Daybreak*, he clearly separates the "*consequences* of various acts" which "step into our reflexive consciousness" as we deliberate and which thus constitute the *conscious* motives of our actions from the "different species of motives" by which our actions are "often enough determined" (D 129). This second class of *causally efficacious* motives includes, *inter alia*, "the way we habitually expend our energy", "some slight instigation from a person whom we fear of honor or love", "our indolence", "an excitation of our imagination brought about by some immediate, very trivial event", or just "incalculable physical influences" as well as "caprice and waywardness" (*ibid.*). Crucially, as "*we are accustomed* to *exclude* all these unconscious processes from the accounting [of our actions] and to reflect on the preparation for an act only to the extent that it is conscious", we also fail "to learn which motive has therewith actually proved victorious" (*ibid.*). This suggests that those conscious states we usually take to be the motives of our actions in fact lack the relevant

causal power. On the contrary, the causally efficacious tokens that should figure in a genuine explanation of our actions are the processes situated at the unconscious levels. Crucially, the "struggle itself" which takes place between such goings-on remains "hidden from me" (*ibid.*)—which is just the kind of picture Nietzsche suggests when he argues that willings result from the unconscious interplay between one's drives.

Let us take stock. In section 2, I spelled out first two points in favor of (IO), both regarding how the psychological vocabulary we acquire through socialization underlies the way we conceptualize our mental states. In particular, Nietzsche argues that our use of such vocabulary prevents phenomenological accuracy and something like genuine individuality. These two points, I argued, hardly suffice to substantiate (IO). However, as soon as we factor in Nietzsche's view about the drives, a wider range of reasons emerge in support of his phenomenalist take on introspection. First, we appreciate that—in addition to the previous two points—the way in which we folk-psychologically conceptualize the states we are in is such that it completely obscures their *nature*. As Nietzsche's paradigmatic treatment of states of knowledge illustrates, we take such states to be possible only once detachment from one's drives and affects obtains. Quite to the contrary, however, one's state of knowing something is constituted by a certain overall orientation of (some of) one's drives and affects towards that thing. Second, the folk-psychological framework we defer to also provides us with (typically) wrong beliefs about the *role* played by each type of mental state. For instance, we take that willings are caused by conscious motives which sometimes—i.e., in presence of conflicting motives—are weighed within a process of conscious and rational deliberation. Instead, Nietzsche argues that willings are constituted by a certain arrangement of our drives and that the transactions between members of this "different species of motives" do not mirror the kind of rational patterns of practical inference we take to obtain between conscious motives. What goes on is, rather, a "struggle" which remains hidden from us.

If it is true that introspection affords us awareness neither of the real *nature*, nor of the real *role* of our psychological states—and, in particular, of those states which are involved in the production of our own actions—it seems fair to say that (IO) is vindicated. In point of fact, as Lichtenberg has it, our willings—those volitional states which usually issue in action—are already "too big" when they become "noticeable to us".

7) The Fate of Conscious Attitudes

Nietzsche argues that there is a substantial discrepancy between conscious psychological states and what goes on at the unconscious level of the drives. At the conscious level we find states which are conceptualized according to the folk-psychological framework we learn through intercourse with the other members of our community. This includes both a certain psychological vocabulary as well as a set of beliefs about the role each type of state is supposed to play. At the unconscious level, we find states and processes which result from the interplay between the drives. At this point, we should therefore ask what, according to Nietzsche, we are to make of conscious attitudes.

To address this question, it is helpful to start by focusing on the status of folk-psychology. According to Nietzsche, the kind of personal-level mental states postulated by folk-psychology are *not* the explanatory tokens that should figure in a genuine account of how our actions are produced. The relevant tokens are, on the contrary, the processes and states occurring at the sub-personal level of the drives. Nor—due to the discrepancy between these two levels—are there any prospects of providing a successful mapping between conscious attitudes and unconscious goings-on. This means that accounts couched in terms of folk-psychological conscious attitudes do not offer any genuine explanation of one's actions, as states of this kind simply lack the causal efficacy we naively take them to possess.

The first point which seems to flow from this concerns the *function* of folk-psychology. As Nietzsche insists, phylogenetic evolution and ontogenetic acquisition of a certain folk-psychological framework are intimately related to one's communication-mediated intercourse with other individuals. Thus, folk-psychology has emerged as a tool which facilitates the mutual understanding among the members of a given society. Its success, however, does not depend on its explanatory and predictive power. Rather, Nietzsche takes the adoption of a shared folk-psychological framework to determine that "[i]n all souls an equal number of frequently recurring experiences have gained an upper hand over ones that occur less frequently" (BGE 268). This, in turn, considerably bears on the kind of *normative* attitudes one is likely to adopt, for "[w]hat group of sensations in a soul will be the first to wake up, start speaking, and making demands is

decisive for the whole rank of its values, and will ultimately determine its table of goods" (*ibid.*). Thus, rather than aiming at offering accurate explanations and predictions of actions, folk-psychology works as a mechanism by which the members of a certain society come to *internalize* shared normative attitudes.[25]

A similar line of reasoning has been recently put forward by Tadeusz Zawidski:

"Human beings are distinguished from other mammals by their extreme sociality. Because of this, solving problems of coordination with our fellows is our most pressing ecological task. Many lines of evidence strongly suggest that folk psychological practices like mental state ascription play an important role in this." (Zawidski 2008: 198)

However, as Zawidski notes, the way in which such practices satisfy this role is primarily by "molding human behavior" and not by "accurately describing its causes so as to correctly predict it" (199). In short, he suggests that the function of folk-psychology is not (merely) "mind-reading", but rather "*mind-shaping*", i.e. the setting up of "regulative ideals" (194) bound to socially approved forms of conduct. If we assume this to be correct, it should come as no surprise that the attitudes typically posited by folk-psychology should not figure as the explanatory tokens in a genuine account of our actions.

A second point concerns the *status* of the folk-psychological posits. As we have just seen, Nietzsche maintains that they lack the power to causally determine one's behavior. Furthermore, recall that in his view conscious states are conceptualized in conformity with the psychological framework we have learnt to apply in order to make sense of others' behavior. Ultimately, therefore, the only form of self-knowledge available to us is of *interpretive* nature. This is, indeed, the core view underlying the kind of phenomenalism about self-experience which is captured by (IO). Now, should we then conclude that Nietzsche defends some version of epiphenomenalism or even eliminativism about conscious attitudes? Eliminativism strikes me as too strong a thesis. After all, Nietzsche still thinks that, however misleading the picture they give us of our

---

[25] The view that folk-psychology's primary function is not explanatory has also been defended by some contemporary philosophers and psychologists. See, for instance, Hutto (2004). The fact that the way in which folk-psychological concepts are used is sensible to the moral feature of the context seems to support the view that one's adoption of a certain folk-psychological framework bears on one's normative attitudes. See Knobe (2007) on the morally-loaded aspects of folk-psychology.

mind's working and however confused we might be about their nature and role, conscious states nonetheless qualify as actually occurring mental events.[26] Rather, his main concern is with their lack of explanatory power. Consequently, some version of epiphenomenalism seems a more appropriate candidate. This, however, is material for another paper.[27]

Literature

Carruthers, Peter. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Campioni, Giuliano *et alia* (eds.). 2003. *Nietzsches persönliche Bibliothek*. Berlin, New York: De Gruyter.

Clark, Maudemarie and Dudrick, David. 2012. *The Soul of Nietzsche's Beyond Good and Evil*. Cambridge: Cambridge University Press.

Danto, Arthur. 2005. *Nietzsche as Philosopher. Expanded Edition*. New York: Columbia University Press.

Hurley, Susan and Chater, Nick. 2005. 'Introduction: The Importance of Imitation', in *Perspectives on Imitation. Volume 1. Mechanisms of Imitation and Imitation on Animals*. Cambridge (MA): MIT Press: 1–52.

Hutto, Daniel. 2004. 'The Limits of Spectatorial Folk Psychology'. *Mind and Language* 19/5: 548-573.

Katsafanas, Paul. 2005. 'Nietzsche's Theory of Mind. Consciousness and Conceptualization'. *European Journal of Philosophy* 13: 1–31.

Katsafanas, Paul. 2014. 'Nietzsche on the Nature of the Unconscious'. *Inquiry. Special Issue on Niezsche and Moral Psychology*, ed. by. B. Leiter.

---

[26] For a recent eliminativist view of conscious attitudes, see Carruthers (2011).
[27] I argue that Nietzsche defends a weak version of epiphenomenalism about conscious causation in my (forthcoming-a). For a similar epiphenomenalist reading, see also Welshon (2002). For disagreement, see Katsafanas (2005).

Katsafanas, Paul. Forthcoming. 'Nietzsche's Philosophical Psychology', in J. Richardson and K. Gemes (eds.), *The Oxford Handbook of Nietzsche*. Oxford: Oxford University Press.

Knobe, Joshua. 2007. 'Folk Psychology: Science and Morals', in D. Hutto and M. Ratcliffe (eds.), *Folk Psychology Reassessed*. Dordrecht: Springer.

Leiter, Brian. 2002. *Nietzsche on Morality*. London: Routledge.

Leiter, Brian. 2007. 'Nietzsche's Theory of the Will'. *Philosophers' Imprint* 7: 1–15.

Leiter, Brian and Knobe, Joshua. 2008. 'The Case for Nietzschean Moral Psychology', in B. Leiter and N. Sinhababu (eds.), *Nietzsche and Morality*. Oxford: Oxford University Press: 83–109.

Lichtenberg, Georg Christoph. 1867. *Vermischte Schriften. Erster Band*. Göttingen: Dieterich.

Lichtenberg, Georg Christoph. 1973. *Schriften und Briefe. Erster Band. Sudelbücher I*, ed. by W. Promies, 2nd ed. Munich: Carl Hanser Verlag.

Nietzsche, Friedrich. 1878/1880. *Human, All Too Human. A Book for Free Spirits*, trans. by R. J. Hollingdale, intr. by R. Schacht. Cambridge: Cambridge University Press, 1996.

Nietzsche, Friedrich. 1881. *Daybreak. Thoughts on the Prejudices of Morality*, ed. by M. Clark and B. Leiter, trans. by R. J. Hollingdale. Cambridge: Cambridge University Press, 1997.

Nietzsche, Friedrich. 1882/1887. *The Gay Science*, ed. by B. Williams, trans. by J. Nauckhoff and A. Del Caro. Cambridge: Cambridge University Press, 2001.

Nietzsche, Friedrich. 1886. *Beyond Good and Evil: Prelude to a Philosophy of the Future*, ed. by R.-P. Horstmann and J. Norman, trans. by J. Norman. Cambridge: Cambridge University Press, 2002.

Nietzsche, Friedrich. 1980. *Kritische Studienausgabe in 15 Bände*, ed. by G. Colli and M. Montinari. Berlin, New York: De Gruyter.

Riccardi, Mattia. Forthcoming-a. 'Nietzsche on the Superficiality of Consciousness', in M. Dries (ed.), *Nietzsche on Consciousness and the Embodied Mind*. Berlin, New York: De Gruyter.

Riccardi, Mattia. Forthcoming-b. 'Nietzsche on the Embodiment of Mind and Self', in J. Constâncio (ed.), *Handbook on Nietzsche on the Self*. Berlin, New York: De Gruyter.

Richardson, John. 2004. *Nietzsche's New Darwinism*. Oxford. Oxford University Press.

Risse, Mathias. 2008. 'Nietzschean "Animal Psychology" versus Kantian Ethics', in B. Leiter and N. Sinhababu (eds.), *Nietzsche and Morality*. Oxford: Oxford University Press: 57–82.

Schwitzgebel, Eric. 2008. 'The Unreliability of Naive Introspection'. *Philosophical Review* 117/2: 245–273.

Sully, James. 1882. *Illusions. A Psychological Study*, 2nd ed. London: Kegan Paul, Trench & Co.

Sully, James. 1883. *Les illusions des sens et de l'esprit*. Paris: Germer Baillière.

Tomasello, Michael. 2008. *Origins of Human Communication*. Cambridge (MA): MIT Press.

Welshon, Rex. 2002. 'Conscious and Unconscious Perspectives', *International Studies in Philosophy* 34: 119–33.

Zawidski, Tadeusz (2008): 'The Function of Folk Psychology: Mind Reading or Mind Shaping?'. *Philosophical Explorations* 11/3: 193–210.