

(THIS IS A DRAFT.
PLEASE DON'T QUOTE WITHOUT PERMISSION)

Nietzsche on the Superficiality of Consciousness

Mattia Riccardi

Introduction¹

Nietzsche famously states that “consciousness *is* a surface” (EH *Why I am so clever* 9). This is not only a highly provocative claim, but also a very puzzling one. How are we to make sense of such a striking contention? In this chapter I tackle the challenge posed by this question by showing that the view on consciousness underlying the perplexing claim expressed in *Ecce Homo*—Superficiality, for short—is philosophically well motivated, though unintuitive and probably less palatable to most of us. In particular, I will focus on two more specific characterizations of consciousness—both to be found in aphorism 354 of *Gay Science*—as it seems to me that they provide the key to Nietzsche’s endorsement of Superficiality. First, Nietzsche maintains that consciousness is “basically superfluous” (GS 354): the fact that we can explain one’s behaviour without appealing to one’s consciousness indicates that “consciousness is not causally efficacious in its own right”, as Leiter (2002: 92) puts it. Second, Nietzsche argues that consciousness involves “a vast and thorough corruption, falsification, superficialization, and generalization” (GS 354), since, far from revealing the motives of our own actions, it rather tends to distort them in a way which—he suggests—we have good reason to consider confabulatory. I will refer to these two main features of Nietzsche’s position as to the “superfluousness claim” (SC) and to the “falsification claim” (FC).

In a seminal paper on this topic, Paul Katsafanas has recently offered a reading of Nietzsche’s view on consciousness which perceptively

¹ I would like to thank João Constâncio, Manuel Dries and Paul Katsafanas for their comments on previous drafts of this paper. Shortly before I prepared the final version of this paper, André Itaparica (in conversation) made me aware of some ambiguous formulations, which I tried to expunge. I presented (parts of) the paper in Porto, Lisbon, Belo Horizonte and São Paulo, where I could benefit from the stimulating discussion.

addresses both (SC) and (FC). On the one hand, he provides an interpretation of Nietzsche's endorsement of (FC) by arguing that, by turning conscious, the content of a given mental state gets articulated conceptually and that such conceptualization is the source of the falsification² Nietzsche ascribes to consciousness in GS 354. On the other hand, Katsafanas denies that Nietzsche holds (SC) altogether, since—he argues—this second claim is at odds with several descriptions and explanations of psychological phenomena he offers elsewhere in his works.

The reading developed in what follows challenges the treatment Katsafanas offers of both claims. The deepest dissent will be about superfluosity, for I will argue that (SC) plays a crucial role in Nietzsche's case for Superficiality. In particular, I will maintain that Nietzsche endorses a weak, but still substantive version of *epiphenomenalism* with regard to *conscious causation*. As regards (FC), I will agree with Katsafanas' main thesis that the kind of falsification Nietzsche is concerned with in GS 354 is ultimately due to the way in which conscious content is conceptualized. However, I will dissent from Katsafanas' further claim according to which—for Nietzsche—"a mental state is conscious if its content is conceptually articulated, whereas a state is unconscious if its content is nonconceptually³ articulated" (Katsafanas 2005: 2). I will instead show that Nietzsche allows for *unconscious conceptual content*. Therefore, if he takes consciousness to involve some kind of falsification, this has to depend on some proprietary form of conceptualisation. My proposal will be that *socially mediated propositional articulation* is here the relevant, intrinsically conscious, form of conceptualisation.

Nietzsche's Leibnizian Story

Aphorism 354 of *Gay Science* starts by laying out the puzzle of superfluosity:

The problem of consciousness (or rather, of one's becoming conscious [*des*

² The general claim that for Nietzsche conceptualisation is a source of falsification is also defended by Hussain 2004. In Riccardi (forthcoming) I, too, argue in favour of this claim. In both these pieces, however, the stress is quite different from that of the present paper.

³ More precisely, unconscious content is "phenomenally articulated" (Katsafanas 2005: 4).

Sich-Bewusst-Werdens] of something) first confronts us when we begin to realize how much we can do without it; and now we are brought to this initial realization by physiology and natural history (which have thus required two hundred years to catch up with *Leibniz's* precocious suspicion). For we could think, feel, will, remember and also 'act' in every sense of the term, and yet none of all this would have to 'enter our consciousness' (as one says figuratively). All of life would be possible without, as it were, seeing itself in the mirror; and still today, the predominant part of our lives actually unfolds without this mirroring—of course also our thinking, feeling, and willing lives, insulting as it may sound to an older philosopher. *To what end* does consciousness exist at all when it is basically superfluous? (GS 354, translation changed)

Nietzsche argues that contemporary developments in natural science have provided support to Leibniz's insight according to which a great part of our mental life is not conscious. A first problem is how to make sense of Nietzsche's reference to Leibniz. Here, Leibniz's famous talk of "*petites perceptions*" might be the most likely association to come to mind. On closer scrutiny, though, this option does not seem to harmonise well with the context of Nietzsche's aphorism.

Consider, as Lanier Anderson invites us to do,⁴ Leibniz's example of our hearing of the ocean's roar. The idea is that we do not perceptually experience all the tiny sounds each wave produces, but rather a 'unified' acoustic property we typically describe as the ocean's roar. However, Leibniz claims that we do have perceptions corresponding to each one of the stimuli, which conjointly generate our acoustic experience of the ocean's roar. Only, these perceptions are too infinitesimal to directly become the object of our awareness, and must thus be unconscious. Despite being such a straightforward option, it is not clear how we could sensibly extend the model of "*petites perceptions*" so as to cover also mental attitudes like beliefs, desires or emotions, which appear in fact to be the main concern of GS 354. For what is here supposed to play the very role that, in the case of sensory experience, Leibniz ascribes to the tiny perceptions?

Fortunately, help comes from aphorism 357 of *Gay Science*, where Nietzsche refers to "*Leibniz's* incomparable insight" according to which "consciousness (*Bewusstheit*) is merely an *accidens* of representation"⁵

⁴ See Anderson 2002.

⁵ Nauckhoff translates *Vorstellung* as "power of representation", which seems wrong to me. Nietzsche is saying that a representation, taken as a mental state token, can be either conscious or unconscious. Thus, with *Vorstellung* he does not mean some faculty

(*Vorstellung*) and *not* its necessary and essential attribute; so that what we call consciousness (*Bewusstsein*) constitutes only one state of our mental and psychic world [...] and *by no means the whole of it*" (GS 357, translation altered). Arguably, this is the same view he has in mind when writing the opening lines of GS 354. Let us then take a closer look.

To start with, note that this characterization of Leibniz's position is almost literally borrowed from Otto Liebmann's *Analysis der Wirklichkeit*—a book Nietzsche studied avidly.⁶ In particular, Liebmann praises Leibniz's "psychological discovery" according to which "'to have representations' and 'to be oneself conscious of them' is by no means the same", for "there are in us many latent and unconscious representations" (Liebmann 1880: 212). To make his point clearer, Liebmann explicitly refers to the Leibnizian notion of "*connaissance virtuelle*", rather than to that of "*petites perceptions*".⁷ Importantly, the former notion is immune to the problem raised by the latter, for it non-controversially applies to mental attitudes like beliefs, desires and emotions. Hence, it is this notion that turns out to be the Leibnizian idea pertinent to the context of GS 354.

Moreover, a later chapter from Liebmann's book entitled "Human and animal cognition (*Menschen- und Tierverstand*)" provides significant clues to what Nietzsche might have in mind when he refers to discoveries in the fields of physiology and natural history that have carried Leibniz's intuition further. In a passage underlined in Nietzsche's own copy of Liebmann's work we read that "the non-linguistic animal as well as the cognitively still incapable child, too, judge *in concreto* and draw wordless inferences. As many researchers (Schopenhauer, Helmholtz, Wundt, Sigwart) affirm or recognize, the activity of sensory intuition already involves a hidden logical activity of the intellect, a tacit but very fast occurring judgment and inference" (Liebmann 1880: 498). The main reference here is to the theory of "unconscious inferences" which was first formulated by Schopenhauer⁸ and later fully developed by Helmholtz—a

or power—*Vermögen*, in German—of representation. This is confirmed by the passage from Otto Liebmann from which Nietzsche borrows his description of Leibniz's thesis, as we will see in a minute.

⁶ This is convincingly demonstrated by Loukidelis 2006.

⁷ One might argue that the two different points made by Leibniz are not actually that different. This, however, would not be correct. Instructively, William James considers both Leibnizian ideas, i.e. tiny perceptions and latent representations, as each being the starting point for two different—and in his eyes equally flawed—arguments in favour of unconscious mental states. See James 1902: 164–68, *First Proof* and *Seventh Proof* respectively.

⁸ See in this context GS 99, where Nietzsche refers to Schopenhauer's 'doctrine of the intellectuality of intuition' as to 'his immortal doctrine'.

view the young Nietzsche was already well acquainted with.⁹ Thus, the way in which physiology and natural history¹⁰ substantiate the Leibnizian view that most of our mental attitudes would occur and sustain our agency even if they were unconscious is by showing how the sophisticated cognitive capacities exhibited by animals do not require them to be conscious.

So far, however, it is still not clear in which sense the term “conscious” is here to be understood. Distinguishing between different ways in which we can talk about consciousness should help us to illuminate this point. Firstly, we may use the term to pick out the qualitative, first-person character of experience, i.e. the what-it-is-like to be in a certain mental state. The standard qualification in this case is “phenomenal”. Secondly, we may talk about consciousness in terms of awareness, as when one says that one is perceptually conscious of something in one’s visual field. Thirdly, we may also take the term to refer to the more specific and complex kind of consciousness we normally ascribe only to human beings, namely self-consciousness. Which of these different notions of consciousness is at stake in GS 354?¹¹

It seems plain that neither phenomenal consciousness, nor awareness would work here. To appreciate this, recall that Nietzsche’s argument implies that the kind of consciousness he is concerned with is such that we cannot ascribe it to animals. This immediately rules out our first candidate, for phenomenal consciousness is usually understood as given together with sentience. Awareness, however, fares no better, for we intuitively allow for animals to be conscious in this sense—at least in the perceptual case. Thus, if we are looking for a notion of consciousness that is suitable for drawing a divide between animals and human beings, we have to exclude both options. Self-consciousness, however, seems to be a much more promising candidate. Firstly, consider that Nietzsche describes the kind of consciousness he is dealing with as “one’s becoming conscious [*Sich-Bewusst-Werden*] of something”, a formulation which implies the kind of

⁹ On this, see now Reuter 2009.

¹⁰ Here, “natural history” translates Nietzsche’s more straightforward “*Tiergeschichte*”. This roughly corresponds to what we now call evolutionary biology.

¹¹ Of course, the list of putative candidates could go on, including for instance access consciousness and, in particular, monitoring consciousness. In my opinion, access consciousness should be ruled out because of the reflexivity implicit in Nietzsche’s description (in general, see Block 1997 for this notion). Monitoring consciousness, on the contrary, is a *prima facie* viable candidate in favour of which Welshon 2002 makes a strong case. See note 17 below for the reasons why we should prefer self-consciousness over monitoring consciousness in this context.

reflexivity¹² we normally capture with the notion of self-consciousness. Support for this reading is also lent by Nietzsche's "mirror" metaphor, which, too, suggests some sort of reflexivity. Self-consciousness, moreover, is such a high cognitive ability that we can easily expect it to serve as differentiating us from other animals. Indeed, there is the strong intuition that animals lack the kind of reflexivity implied by a genuine notion of self.

If this is correct, we can formulate the main thesis conveyed by Nietzsche's Leibnizian story as follows:

(LT): a mental state can be either self-conscious or non-self-conscious.¹³

At this point, a question arises: how can an unconscious mental state become conscious in the specific sense Nietzsche has in mind? In current philosophy of mind, variations of (LT) are often associated with what has become known as higher-order approaches to consciousness. The basic idea is that a mental state *M* turns conscious when it is "indexed" by a higher-order representation (HOR) of some kind, which signals to one that one is in *M*—a view which Nietzsche clearly entertains in the following *Nachlass* passage:

"Consciousness"—to what extent the represented representation, the represented will, the represented feeling (*which alone is known to us*) is completely superficial! "Appearance" (*Erscheinung*) also in our *inner* world! (NL 1884, KSA 11, 26 [49])

It is not easy to tell which kind of accompanying HOR Nietzsche has in mind here. In contemporary debates, several alternatives are being explored: the HOR accompanying a mental state is taken to be either an (inner) experience, or an (inner) perception, or a thought. Even if it is not entirely clear which one among these candidates is most appropriate to grasp Nietzsche's view, he seems to endorse something like a higher-order thought (HOT) version of the more general HOR approach. This is

¹² Luca Lupo calls this kind of reflexivity-involving consciousness "secondary consciousness" and distinguishes it from the more basic "primary consciousness" we share with animals and which roughly corresponds to mere environmental awareness. See Lupo 2006: 192–3.

¹³ From this point on, I will simply drop the prefix "self" for short and thus revert to the terms "conscious" and "consciousness". However, unless specified otherwise, the terms should always be read as "self-conscious" and "self-consciousness" respectively. Also, whenever emphasis is needed, I will still resort to the prefix "self-".

suggested by his description of one's being conscious in terms of the ability "to 'know' what distressed him, to 'know' how he felt, to 'know' what he thought" (GS 354). Further support for this suggestion comes from the striking similarities between Nietzsche's position and the account of consciousness defended by David Rosenthal, the most famous proponent of HOT theory in contemporary philosophy of mind.¹⁴ According to Rosenthal's version of the HOR approach, in order to become conscious, a mental state M has to be accompanied by a HOT to the effect that one is in M.¹⁵ In particular, it will be useful here to underscore three key points made by Rosenthal, for they smoothly correlate with important aspects of Nietzsche's account.

First, Rosenthal stresses that not all mental states are conscious. Indeed, the main motivation of his theory is to articulate a view which does not conflate 'mental' and 'conscious'. According to his theory, an unconscious mental state M can become conscious *only* by being accompanied by the relevant HOT. This thesis nicely expresses the Leibnizian point made by Nietzsche according to which we have latent mental representations, which can at some moment "enter our consciousness" (LT).

Second, from Rosenthal's definition of conscious mental state follows that it involves reference to the subject who is in it. Thus, reflexivity turns out to be an essential characteristic of conscious mental states. As Rosenthal puts it, the "content [of HOTs] must be that one is, oneself, in that very mental state" (Rosenthal 1997: 714).

Third, and crucially, Rosenthal stresses that the fact that mental states can occur both in conscious and unconscious form raises the problem of superfluousness already exposed by Nietzsche: for "[w]hat, if any, function do conscious versions of these states have that nonconscious versions lack?" (Rosenthal 2007: 829). It is about time to start tackling this problem.

Consciousness and Language

According to Nietzsche, once we grasp that all our mental attitudes can—and most of the time in fact do—occur and determine our agency without becoming conscious, i.e. without being accompanied by any suitable HOT,

¹⁴ This point has been stressed by Brian Leiter, in particular in his incarnation as a blogger. See also Abel 2001: 10.

¹⁵ The classic paper is Rosenthal 1997.

the obvious question arises: why has consciousness evolved, given that it is superfluous in this precise sense? His answer is that consciousness emerged as a result of the fact that human beings had to join and live in society in order to survive:

That our actions, thoughts, feelings and movements—at least some of them—even enter into consciousness is the result of a terrible “must” which has ruled over man for a long time: as the most endangered animal, he *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)

Nietzsche’s main thesis, thus, is that consciousness is closely related to the “*ability to communicate*” (GS 354). It is not completely clear, however, how we are to make sense of this claim. Central to his view seems to be the quite plausible idea that among the many things the members of a given community would need to communicate about, the mental states they are in would in any case be fundamental. This, however, requires that one “knows” the mental state one is currently in. Moreover, the communication-driven character of conscious states is clearly indicated for Nietzsche by their linguistic nature: “conscious thinking *takes place in words, that is, in communication signs [Mittheilungszeichen]*” (GS 354). Now, provided that consciousness is born out of our need to communicate, it seems natural to assume that it bears some relevant connection to language. And yet, again, we will need to qualify this view further in order to get a clearer picture.¹⁶ Let us start by highlighting two main aspects.

The first point concerns the relation between language and consciousness. In an aphorism from *Beyond Good and Evil* that is tightly linked to GS 354, Nietzsche notes that to share a language is not sufficient to guarantee flawless communication. For “[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

¹⁶ The fact that Nietzsche directly links consciousness and language confirms that the kind of consciousness at issue in GS 354 is self-consciousness. On the one hand, it seems plausible to argue that self-consciousness depends on language as long as the capacity to self-refer implicit in it requires one’s mastery of pronouns like “I” or “mine”. On the other hand, it confirms that Nietzsche is concerned with a kind of consciousness that—being essentially linguistic—cannot be ascribed to animals. As argued by Rosenthal 2005b, our ability to verbally *report* the mental states we are in offers an important clue of the close link between consciousness and language.

have to have the same experience *base*” (BGE 268). The issue addressed here is the problem of a private language: if mental terms refer to first-person properties of our inner states, how can we be sure that they convey genuine, i.e. inter-subjectively understandable meaning? How can mental talk not be private? Remarkably, Nietzsche’s answer is that language-mediated social intercourse has *de facto* made uniform the inner life of people belonging to the same community, thereby enabling mental terms to actually denote states of the same type. The “genius of the species” Nietzsche refers to in both aphorisms¹⁷ consists therefore in the mental vocabulary shared by the members of a certain linguistic community.¹⁸

In Nietzsche’s eyes, however—and we are now on the second point—this has fatal consequences. For “each of us, even with the best will in the world to *understand* ourselves as individually as possible, ‘to know ourselves’, will always bring to consciousness precisely that in ourselves which is ‘non individual’, that which is average” (GS 354). In other words, Nietzsche seems to hold that we interpret our own mental states in light of a socially developed “theory of mind”: we attribute to ourselves the same types of mental states we also attribute to others. Crucially, this is the main reason which substantiates (FC), since it is the fact that “our thoughts themselves are continually as it were *outvoted* and translated back into the herd perspective” which causes “all becoming conscious” to bring about “a vast and thorough corruption, falsification, superficialization, and generalization” (GS 354). Since one ends up ascribing to oneself attitudes identical to those one currently ascribes to others, the uniqueness of what we think, desire and feel is, if not completely blanked out, at least significantly blurred. Hence, to understand the dynamics of this re-interpretation will be crucial to make sense of (FC). Before we turn to this task, it will be helpful to briefly pause and see where we have gone so far.

In the light of the analyses developed in the last two sections, we can start by giving a more precise formulation of the two main claims Nietzsche endorses with regard to consciousness:

(SC): mental attitudes could—and most of the time actually do—occur and sustain our agency without becoming conscious.

¹⁷ “On ‘genius of the species’” is the very title of GS 354. Here is how the expression occurs in BGE 268: “Fear of the ‘eternal misunderstanding’: this is the benevolent genius that so often keeps people of the opposite sex from rushing into relationships at the insistence of their hearts and senses—and *not* some Schopenhauerian ‘genius of the species’—!”

¹⁸ Lupu has a nice discussion of this point (2006: 189-190).

(FC): in becoming conscious, the content of mental attitudes is re-translated in light of a socially developed and acquired “theory of mind”.¹⁹

More importantly, we are now in a position to address the two difficulties raised by Katsafanas’ rendering of Nietzsche’s account of consciousness. The first problem concerns (FC). According to Katsafanas’s reading, Nietzsche holds that, by turning conscious, the content of a mental state becomes conceptualised and that such conceptualization is responsible for falsification. However, as suggested by Nietzsche’s commitment to the idea of Helmholtzian—and Schopenhauerian—“unconscious inferences”, Katsafanas’ further claim to the effect that conceptualisation *always* goes together with consciousness seems to misconstrue his view.²⁰ My own suggestion will be to argue that the kind of falsification involved in conscious mental attitudes is not due to conceptualisation *in general*, but rather results from the both *socially mediated* and *propositionally articulated* form in which mental attitudes are typically re-translated.

The second difficulty regards (SC). For does not Nietzsche’s answer to the problem of superfluosity indicate that consciousness has indeed a profound and manifest influence on the course of our life? Despite the *prima facie* conclusiveness of this observation, I will defend that there is a strong sense in which consciousness is for Nietzsche superfluous, for he holds that a mental state has the causal powers it happens to have quite independently of it being conscious. As Welshon correctly points out,²¹ this gives us sufficient room for a weak, but still substantive version of epiphenomenalism about conscious causation.²²

¹⁹ The motivation for (FC) worked out above can be summarized as follows: (a) consciousness serves the goal to communicate the mental states we are in; (b) the vehicle of communication are words (or, more generally, signs); (c) mental terms, in order to convey meaning, have to pick out the same type of mental state. From (a), (b) and (c) follows that we express and report the mental attitudes we are in by adopting the same kind of mental talk we use to make sense of others’ inner states.

²⁰ See section (1) above.

²¹ See Welshon 2002. For Welshon’s proposal see note 31 below.

²² As we have already seen at the beginning of this paper, this is also the main thesis held by Leiter 2002, who writes that “conscious states are only causally effective in virtue of type facts about that person” (91), where “type-facts” are facts regarding one’s psycho-physical constitution.

Conceptualisation and Falsification, Unconscious and Conscious

According to Katsafanas' account, Nietzsche holds that conceptualisation is indissolubly associated with consciousness. Given that he takes conceptualisation to be responsible for falsification, the two theses follow that (a) unconscious mental states are not-yet-falsified *qua* still non-conceptualised, and that (b) conscious mental states are falsified *qua* conceptualised. In this section I will argue against (a) and show that (b) requires some substantive qualification if it is to accurately capture Nietzsche's view. Since Katsafanas sees his own reading confirmed by BGE 192, I will focus on this aphorism.

Nietzsche stresses here the inaccuracy of our perception, for instance, of a tree, how little we see it "precisely and completely, with respect to leaves, branches, colors, and shape" (BGE 192). Quite on the contrary, he notes, "[w]e find it so much easier to imagine an approximate tree instead" (BGE 192). Katsafanas' treatment of this example is as follows:

Nietzsche's idea is that our perceptions sometimes represent objects in a way that is not sensitive to all of the detail of the object, but is instead sensitive only to the general type to which the object belongs. This type of perception represents the tree as an instance of the concept TREE, rather than representing it in its individual detail; it does so by emphasizing certain general features of trees at the expense of the individual details of this particular tree. (Katsafanas 2005: 7)

Surely, Katsafanas' rendering of Nietzsche's point in BGE 192 is mostly correct. To put it more succinctly, the kind of conceptualisation responsible for falsification at the perceptual level is *generalisation*. What seems problematic, however, is the further assumption according to which such a generalisation requires a mental state to be conscious and consequently—given Nietzsche's view—language-dependent.²³ Is this correct?

I think the answer is No, for the kind of generalisation illustrated by the tree example seems to be the result of unconscious processes, which transform what Nietzsche refers to in the *Nachlass* as the "chaos of sensation" into a full-fledged perception.²⁴ This is confirmed by Nietzsche's general characterisation of the way in which our perception

²³ Recall that for Nietzsche "conscious thinking *takes place in words*" (GS 354).

²⁴ See NL 1887, KSA 12, 9[106]. On Nietzsche's "chaos of sensation", see Riccardi forthcoming.

falsifies: “[g]iven some stimulus, our eyes find it more convenient to reproduce an image that they have often produced before than to register what is different and new about an impression” (BGE 192).²⁵ Here, the fact that Nietzsche uses the term “image” suggests that no linguistic concept needs to be involved in the relevant cases. Indeed, examples like the tree perception seem to apply also to non-linguistic animals.

Even if this was accepted, one could still accommodate Katsafanas’ original proposal. One strategy would be to accept that there is low-level falsification due to the unconscious mechanisms that govern our perceptions, yet at the same time refuse to qualify such processes as genuine conceptualisation. One could therefore salvage the main claim according to which only (language-dependent) conscious content is conceptualized.²⁶

How might one respond to this move? The only strategy we can pursue is to show that perceptual generalisation counts as a genuine kind of conceptualisation. To start with, note that Nietzsche’s characterisation of concepts as “pictorial signs [*Bildzeichen*] for sensations that occur together and recur frequently” (BGE 268) clearly indicates that mental images could do the required conceptual work. Indeed, according to this description, it seems that perceptual concepts are something like “sensory templates” that we form when we first come across some object O and then reactivate on successive encounters with objects of the same kind.²⁷ A sensory template, thus, works as a recognitional concept. Can we say that such a recognitional concept genuinely represents a given O as being a particular of a certain type, even if it operates under the threshold of consciousness? For the answer to this question, it will be helpful to briefly go back to the historical context of Nietzsche’s theory.

Recall that Nietzsche endorsed the by his time mainstream Helmholtzian account according to which our perceptions are the result of

²⁵ Note that the kind of inaccuracy Nietzsche is willing to point out is a quite general one. Indeed, as suggested by the subsequent example according to which, while reading a text, we do not actually read “all the individual words (or especially syllables) on a page” (BGE 192), to perceive something as instantiating some general type is only one way in which our senses can be inaccurate.

²⁶ I am grateful to João Constâncio for this point. See also Constâncio 2011, section 2. Katsafanas, too, notes: “conscious perception involves a classifying awareness, whereas unconscious perceptions involve only a discriminatory ability, only a perceptual sensitivity to features of the environment” (Katsafanas 2005: 9). It is worth remembering, however, that according to Katsafanas the tree example is indeed an example of *conscious* perception.

²⁷ See Riccardi forthcoming for more details on this point. The notion of “sensory template” is borrowed from Papineau 2007.

an unconscious processing, which is best described as an instance of thought. Thus, given that sensory templates are recognitional concepts and that they plausibly play some central role in such unconscious mechanisms, we obtain a consistent notion of ‘unconscious concept’. Interestingly, we find a similar view in Liebmann’s *Analysis der Wirklichkeit*—the work to which Nietzsche owes essential insights into the Leibnizian story that proves crucial for his understanding of consciousness in GS 354. As we have already seen, Liebmann maintains that the behaviour exhibited by several animal species presupposes a quite sophisticated mind. In particular—and decisively for our present concern—he holds that the “simple recognition of the objects of sensible intuition is but the primitive type of affirmative judgment” (Liebmann 1880: 498). This shows that the recognitional ability provided by one’s possession of a given sensory template suffices for one to perceptually represent an O as instantiating the corresponding type. Crucially, such an ability qualifies as conceptual, although one’s exercise thereof requires neither mastery of a language nor self-consciousness.

We can therefore conclude that generalisation is a kind of falsification-involving conceptual capacity which falls on the wrong side of the divide Nietzsche draws at the beginning of GS 354, namely on the side also populated by animals. It follows that whatever type of conceptualisation might be relevant in our context needs to satisfy the quite general constraint that it must not already occur at the level of unconscious conceptualisation which is typical for perceptual experience.²⁸

Unfortunately, this is still much too vague. To start working towards a viable solution, recall that GS 354 is concerned, in particular, with conscious mental attitudes like beliefs, desires, and emotions, the content of which is thus *propositional*. Suitable examples are states like: “I think that p”, “I want that p” and “I feel that p”. Furthermore, Nietzsche tells us that one typically acquires the ability to conceptually articulate such attitudes through linguistic intercourse with other members of one’s society. It seems to me that these two characteristics give us important clues as to how to work out the *peculiar* kind of falsification which is peculiar for consciousness.²⁹ Let us take a closer look.

A first feature of the relevant kind of falsification derives from the

²⁸ Consider also that Nietzsche holds consciousness to be language-dependent, whereas he does not take concepts to (necessarily) be language-dependent, since, as we saw, he defines them as “pictorial signs” (BGE 268).

²⁹ Of course, the most general kind of falsification, i.e. generalisation, might be at work also in such cases. The point is that we need to understand in which proprietary way self-conscious and propositionally articulated states falsify.

mental vocabulary we use to consciously articulate our mental life. Recall Nietzsche's position: through social, i.e. basically linguistic interaction with the other members of our community we learn how to read one's mind—an ability we apply to ourselves as well. Thus, we tend to self-ascribe the same types of mental states we attribute to others. In Nietzsche's eyes, however, the way in which our own inner states are re-interpreted in terms of this mental vocabulary obscures their nature.

A second feature is due to the propositional structure of conscious content itself. In particular, all mental states we attribute to ourselves involve reference to the "I"—something that should not surprise us, for we are dealing with states that are self-conscious. This gives us a powerful hint as to how to make sense of the kind of falsification which, according to Nietzsche, is typical for propositionally-articulated mental attitudes: we are led to believe that there is an "I" which acts as the bearer of the relevant mental attitudes. In Nietzsche's eyes, this is due to the syntactical structure of our conscious thought. However, he argues, on this point language simply misleads us.

There are two main senses in which we are thus misled: the first is that the propositional structure of conscious mental attitudes like "*I think that p*", "*I desire that p*", and "*I feel that p*" inoculates the belief that there is actually something to which the indexical "I" refers: a soul, or a subject. As Nietzsche puts it, "people used to believe 'in the soul' as they believe in grammar and the grammatical subject: people said that 'I' was a condition and 'think' was a predicate and conditioned—thinking is an activity, and a subject *must* be thought of as its cause" (BGE 54). As this passage already suggests, the second relevant aspect is that the soul, or subject, so posited is conceived of as being causally efficacious. Crucially, both aspects are for Nietzsche immediately related to the fact that the mental attitudes we are concerned with are conscious. In other words, the soul, or subject, that we take to be substantial and efficacious is also believed to be intrinsically conscious. At face value, it is hard to see how we are to make sense of the relations between the quite different properties here ascribed to the soul. Since this will give us the crucial clue as to how to interpret (SC), I will leave this problem for the next section. For now, let me briefly recapitulate the main points elaborated so far.

Nietzsche does not bind consciousness and conceptualisation together as tightly as argued by Katsafanas. The perception of a tree, for instance, typically involves generalisation, an operation he takes to be genuinely conceptual although it does not require one to be self-conscious of the perception one is having. Hence, in order to make sense of (FC) it won't help to appeal to conceptualisation indiscriminately. Rather, we need to individuate the *specific* form of conceptualisation responsible for the

content of our mental attitudes being falsified as soon as they turn conscious. According to the proposal put forward here, the kind of conceptualisation we are looking for consists in our mental attitudes (a) being re-interpreted according to a socially acquired “theory of mind” (b) and being conceptualised in such a way that the “I” figures not only as the bearer, but also as the ostensibly authentic originator of our beliefs, intentions and volitions. In short, the relevant kind of conceptualisation is *socially mediated propositional articulation*.

The Illusion of Conscious Causation: Superfluousness Vindicated

As we have seen above, consciousness encompasses a proprietary form of falsification. In order to work out (SC), the falsifying element we need to focus on is the idea that conscious states are causally efficacious *qua* conscious, a view Nietzsche clearly rejects in *Twilight of the Idols*: “the conception of a consciousness (‘mind’) as cause, and then that of the I (the ‘subject’) as cause are just latecomers that appeared once causality of the will was established as given, as *empirical*... Meanwhile, we have thought better of all this” (TI *The four great errors* 3). It will be impossible to deliver a detailed account of Nietzsche’s position here. Rather, I will concentrate on the aspects most relevant for the problem of superfluousness. In particular, two theses need to be discussed. The first is the view according to which real psychological causality is at the level of unconscious dispositions Nietzsche conceives in terms of drives. The second is that we usually confabulate about our own mental life, thus construing false explanations of our being in a certain (conscious) mental state. Let us start with the first claim.

Nietzsche thinks that we become introspectively aware of just a few of the inner states we are in. Thus, only a very small part of our mental attitudes become conscious. Such conscious states, however, are causally produced by psychological processes which do not themselves “enter our consciousness”. The consequence that follows from this is twofold:

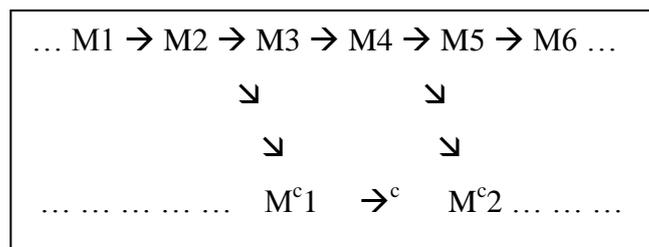
[W]hat becomes conscious is subject to causal relations which are completely withholden from us,—the succession of thoughts, feelings, ideas in consciousness does not mean (*ausdrücken*) that this sequence is causal: it is *apparently* so, though, and at the utmost level. (NL 1887, KSA 12, 11[145])

On the one hand, there are causal connections holding between our mental states of which we are not aware. On the other hand, given that we have conscious access only to a very small number of our inner states, we

come to feel that there is some different, distinctive causal link obtaining *only* between those conscious states. Here, in Nietzsche's eyes, is where the (wrong) picture we have of our own agency originates. Interestingly, Rosenthal offers a very similar account:

Because our mental states are not all conscious, we are seldom if ever conscious of the mental antecedents of our conscious states. And conscious desires and intentions whose mental antecedents we are not conscious of seem to us to be spontaneous and uncaused. The sense we have of free agency results from our failure to be conscious of all our mental states. (Rosenthal 2005c: 361)³⁰

We can summarize this idea in the following model:



There is a series M1-M6 of causally connected unconscious mental states. Only M3 and M5 become conscious, whereby their contents become propositionally articulated. M^c1 and M^c2 are the conscious counterparts of M3 and M5. (The different numeration is meant to highlight that M^c1 and M^c2 are the only two introspectively accessible states, for they are the only two states to become conscious.) Once we introspect and find that we are in M^c2, "we want there to be a *reason why* we are in the particular state we are in" (TI *The Four Great Errors* 4), as Nietzsche puts it. Therefore, since all that we find by introspection is that our actual (conscious) state M^c2 was preceded by the (conscious) state M^c1, we take that there is some

³⁰ See also Dretske: "If what makes an experience or a thought conscious is the fact that S (the person to whom it occurs) is, somehow, aware of it, then it is clear that the causal powers of the thought or experience [...] are unaffected by its being conscious. Mental states and processes would be no less effective in doing their job [...] if they were all unconscious" (2000: 186). Dretske, however, takes this to be an unacceptable consequence of HOR theories of consciousness.

conscious causal connection (\rightarrow^c) between them.³¹ Here is how we construe this connection according to Nietzsche:

The memory that unconsciously becomes activated in such cases is what leads back to earlier states of the same type and the associated causal interpretation, — *not* their causality. Of course, memory also interjects the belief that representations (*Vorstellungen*), the accompanying processes of consciousness (*Bewusstseins-Vorgänge*), had been the cause. This is how a particular causal interpretation comes to be *habituated*; this interpretation in fact inhibits an *investigation* into the cause and even precludes it. (TI *The Four Great Errors*, 4, translation changed)

The scenario Nietzsche sketchily presented in this passage is complicated. For the purpose of this paper, it will suffice to focus on the two main elements of falsification implicit in it. The first one is the conscious causation we posit as linking together only the mental states we are conscious of, and which works, as it were, as a general schema. The second element of falsification has to do with the peculiar conceptualisation our inner states undergo by becoming conscious. As we have seen above, this process results in propositionally articulated mental attitudes. Now, such propositional attitudes are the kind of states we recur to in order to fill in the general causal schema we use to make sense of our mental life, for they typically figure in our explanations as the *motives* which bring about the actual state we are in. According to the view recently defended by Peter Carruthers—another HOT theorist—“all active intentions and acts of intention-formation are self-attributed via a process

³¹ The model proposed here is similar to the one put forward by Welshon 2002: 123–24. As I see it, the main differences between my own account and Welshon’s version of epiphenomenalism result from his reading consciousness as *monitoring* consciousness. As he puts it, “psychological events cause other psychological and non-psychological events because of their non-monitoring conscious properties” (2002: 123). I agree with this general way of construing Nietzsche’s epiphenomenalism. Only, I think we should drop the qualification “monitoring”, and rather understand consciousness as self-consciousness. The main gain in so doing is that we can make sense of (FC) far more naturally. On the contrary, if we construe consciousness as some kind of monitoring of our inner states, it is not easy to see why such monitoring should involve falsification. For on the most natural reading, monitoring is a “neutral” operation, which does not affect what is being monitored in the way consciousness is supposed to do according to Nietzsche. Moreover, the relevant kind of falsification includes the way in which, by articulating our mental attitudes propositionally, we factor in the “I” as the bearer they depend on. Again, this indicates that the kind of consciousness Nietzsche is dealing with entails a constitutive reference to the self.

of self-interpretation” (Carruthers 2007: 205).³²

We are now in a position, I think, to better assess (SC). According to Nietzsche, there is some kind of psycho-physiological causation, which determines the inner states we are in. The conscious causation we ascribe to the propositionally articulated states we are introspectively aware of is, on the contrary, fictitious. This seems to suggest that a mental state M’s power to cause some other mental state M* is insensitive to M’s being conscious or not. From this follows a robust sense in which consciousness is superfluous: the fact that a mental state turns conscious does not lend any additional causal efficacy to it.

At this point, however, there is a strong difficulty we have to face. Katsafanas forcefully points it out by underscoring that the “way in which a state becomes conscious has the most diverse and far-reaching range of consequences” (Katsafanas 2005: 23). In particular, he considers some pertinent examples taken from Nietzsche’s own work. Take, for instance, the thought of eternal recurrence. It seems quite obvious that the way one would normally get acquainted with this thought is by consciously forming and entertaining it, as most philosophy students in fact do after reading Nietzsche or after being told about ancient Pythagoreanism. Another example on which Katsafanas pauses at length is that of *bad conscience*. According to his rendering of Nietzsche’s position, bad conscience “names an unconscious state of profound suffering” which “is conceptualized as guilt: that is, the unconscious bad conscience gives rise to the conscious emotion of guilt” (Katsafanas 2005: 21). Crucially, this process of conscious conceptualisation has tremendous impact on the entire mental life of the subject. If this is correct, what about (SC)? Should we say that consciousness is not superfluous, after all?

My view is that the proper answer to this last question is Yes and No. The reason for the ambiguity is due to the fact that considerations such as those put forward by Katsafanas tend to conflate two different ways in which we may understand superfluosity. Since, according to Nietzsche, linguistic communication requires a subject to be conscious, it follows that consciousness plays a fundamental role in our *acquisition* of public or

³² There is an interesting connection between this and the belief in a substantial subject I mentioned briefly above. Two things strike me as important: (i) it seems that we are “immediately aware” of conscious thoughts, and (ii) in such thoughts always figures a reference to the “I”. See Rosenthal: “And, by seeming subjectively to be independent of any conscious inference, HOTs make it seem that we are conscious of our conscious states in a direct, unmediated way. But that very independence HOTs have from conscious inference also makes it seem that we are directly conscious of the self to which each HOT assign its target” (333–4).

cultural representations in general. Moreover, it is undisputable that such representations have an enormous impact on what we think and do. Nonetheless, I cannot see how this point should rule out the relevant kind of superfluousness we have been concerned with so far. For the fact that consciousness plays a crucial role in our acquisition of a wide range of representations is compatible with the physio-psychologically causal role of those representations being independent from consciousness.³³ Let me explain this point in more detail.

Consider again the example of bad conscience. Someone in the relevant “unconscious state of profound suffering” is being told “things” which induce conceptualising her actual state as guilt. How does this happen? A quite general answer to this problem is that the “things” one is told are public representations which need to be *internalized* if they are supposed to have some bearing on what one thinks and does. In our example, thus, one would need to internalize the belief that the distressing state she is in is the consequence of some misdoing she has committed earlier. Only once this belief has been internalized would one’s state be effectively conceptualised as “guilt” and become behaviourally relevant as such. The crucial factor, thus, is the kind of psychological mechanism responsible for the internalization. Again, Rosenthal makes the relevant point:

The role that thoughts and desires can play in our lives is a function of their causal relations to one another and to behavior. And presumably those causal relations are due solely, or at least in great measure, to the intentional contents and mental attitudes that characterize the states. (Rosenthal 2005c: 362)

Likewise, the picture sketched by Nietzsche does not indicate that the process through which some acquired representation becomes psychologically efficacious involves consciousness.³⁴ Rather, he often talks of a mechanism of cognitive “assimilation” which resembles physiological processes such as digestion. In the *Genealogy*, for instance, Nietzsche tentatively suggests for such a mode of internalization the term

³³ As Katsafanas (private correspondence) correctly points out, he takes that “conscious state” is for Nietzsche tantamount to “conceptual state”. Given this, the claim that no state is causally efficacious *qua* conscious would be tantamount to the claim that no state is causally efficacious *qua* conceptualized, which is a very strong and unpalatable claim. However, I am not myself committed to this claim, since I do not share the premise according to which conscious is tantamount to conceptual.

³⁴ For a distinction similar to the one I am here advocating between the *acquisition* and the *internalization* of a public representations, see Constância (2011).

“inanimation (*Einverseelung*)”, which explicitly parallels physiological “incorporation (*Einverleibung*)” (GM II 1: 35).³⁵ Relevantly, this view is in tune with the claim to be found in a note from 1882 where “morality” is defined as the “quintessence (*Inbegriff*) of all our incorporated (*einverleibten*) valuations” (4[151]: KSA 11). Accordingly, the moral—and, more generally, cultural—representations and beliefs we acquire socially can work as causally efficacious mental states only once they have been internalized and thus integrated into the relevant psycho-physiological mechanisms. More specifically, Nietzsche holds that the content of conscious mental attitudes—like my conscious desire to quit smoking—needs to be reshaped so as to figure as the content of the unconscious and intentionally structured drives which actually determine my agency. This process realizes in his view the kind of internalization he often refers to with metaphorical variations on the theme of *Einverleibung*.³⁶

Conclusion

By arguing for Superficiality, Nietzsche maintains both that consciousness is superfluous and that it involves falsification. In this paper, I have tried to make sense of these two claims (SC and FC respectively). In order to better frame the problem, I started by considering Nietzsche’s notion of consciousness and argued that it should be read as corresponding to self-consciousness. Given this—and also considered the Leibnizian story he tells us at the beginning of GS 354—the most accurate rendering of Nietzsche’s position brings him in the vicinity of contemporary HOT theories of consciousness. What, then, about superfluosity and falsification?

With regard to (FC), I argued that the relevant kind of falsification is due to the mental vocabulary as well as to the propositional form which govern the way in which our mental life is consciously articulated—something Nietzsche tracks back to our linguistically mediated acquisition of a “theory of mind” we apply not only in order to make sense of others’ behaviour, but also use to ascribe mental attitudes to ourselves. It is therefore true that conceptualisation is the ultimate source of the

³⁵ Also BGE 230 speaks in favour of the reading proposed here.

³⁶ Nietzsche’s position differs from Rosenthal in one relevant respect, for he seems to deny that there must be any strong continuity between the acquired content which characterises our conscious attitudes and the internalized content as it figures in the psychological mechanisms which causally determine our agency.

falsification Nietzsche takes to be implied by consciousness, as claimed by Katsafanas. However, the kind of conceptualisation which is pertinent to consciousness is *not* mere generalisation, which we can find at work also in un(self)conscious perceptions, but rather *socially mediated propositional articulation*—hence, a quite peculiar and complex kind of conceptualisation.

Concerning (SC), I maintained that Nietzsche endorses a weak, but still substantive version of epiphenomenalism about consciousness, for he claims that the causal powers of a given mental state *M* do not depend on *M*'s being or not conscious. I also defended this reading against the arguments put forward by Katsafanas arguing that all that his considerations prove is that, for Nietzsche, consciousness plays an important role in our linguistically mediated *acquisition* of beliefs and, in general, public representations which may become behaviourally efficacious. However, and crucially, this is *not* incompatible with the epiphenomenalist reading proposed here: indeed, Nietzsche's account of cognitive *internalization* in terms of incorporation (*Einverleibung*) suggests that the mechanisms through which representations are interiorised and thus acquire psychologically relevant causal powers work at the unconscious level. Therefore, there is no reason to deny Nietzsche's endorsement of weak epiphenomenalism as construed here.

References

- Abel, Günter (2001). *Bewusstsein–Sprache–Natur*. Nietzsche's Philosophie des Geistes. *Nietzsche-Studien* 30: 1–43.
- Anderson, R. Lanier (2002). Sensualism and Inconscious Representations in Nietzsche's Account of Knowledge. *International Studies in Philosophy* 34: 95–117.
- Block, Ned (1997). On a Confusion about a Function of Consciousness. In: N. Block, N. Flanagan and G. Güzeldere (eds): *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Carruthers, Peter (2007). The Illusion of Conscious Will. *Synthese* 96:197–213.
- Constâncio, João (2011). On Consciousness: Nietzsche's Departure from Schopenhauer. *Nietzsche-Studien* 40: 1–42.
- Dretske, Fred (2000). What Good is Consciousness. In: F. Dretske, *Perception, Knowledge and Belief: Selected Essays*, Cambridge: Cambridge University Press.
- James, William (1902). *The Principles of Psychology*. New York: Henry Holt and Company.
- Hussain, Nadeem (2004). Nietzsche's Positivism. *European Journal of Philosophy* 12: 326–368.
- Katsafanas, Paul (2005). Nietzsche's Theory of Mind. Consciousness and Conceptualization. *European Journal of Philosophy* 13: 1–31.
- Leiter, B. (2002). *Nietzsche on Morality*. London: Routledge.

- Liebmann, Otto (1880). *Zur Analysis der Wirklichkeit*, 2nd edition. Strasbourg: Karl J. Trübner.
- Loukidelis, Nikolaos (2006). Nachweis aus Otto Liebmann ‚Zur Analysis der Wirklichkeit‘. *Nietzsche-Studien* 35: 302–3.
- Lupo, Luca (2006). *Le colombe dello scettico. Riflessioni di Nietzsche sulla coscienza negli anni 1880–1888*. Pisa: ETS.
- Nietzsche, Friedrich (1882/1887). *The Gay Science*, ed. B. Williams, trans. 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. R.-P. Horstmann and J. Norman, trans. J. Norman. Cambridge: Cambridge University Press, 2002.
- Nietzsche, Friedrich (1887). *On the Genealogy of Morality. A Polemic*, trans. M. Clark and A. Swensen. Indianapolis: Hackett, 1998.
- Nietzsche, Friedrich (1888). *Twilight of the Idols*, in F. Nietzsche, *The Anti-Christ, Ecce Homo, Twilight of the Idols and Other Writings*, ed. A. Ridley and J. Norman, trans. J. Norman. Cambridge: Cambridge University Press, 2005.
- Nietzsche, Friedrich (1980). *Kritische Studienausgabe in 15 Bänden*, ed. G. Colli and M. Montinari. Berlin, New York: Walter de Gruyter.
- Papineau, David (2007). Phenomenal and Perceptual Concepts. In: T. Alter and S. Walter (eds). *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.
- Reuter, Sören (2009). *An der „Begräbnisstätte der Anschauung“*. *Nietzsches Bild- und Wahrnehmungstheorie in ‚Ueber Wahrheit und Lüge im aussermoralischen Sinne‘*. Basel: Schwabe.
- Riccardi, Mattia (forthcoming). Nietzsche’s Sensualism. *European Journal of Philosophy*.
- Rosenthal, David (1997). A Theory of Consciousness. In: N. Block, N., Flanagan, and G. Güzelde (eds). *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Rosenthal, David (2005a). *Consciousness and Mind*. Oxford: Clarendon Press.
- Rosenthal, David (2005b). Why Are Verbally Expressed Thoughts Conscious. In: Rosenthal 2005a: 282–305.
- Rosenthal, David (2005c). Unity of Consciousness and Self. In: Rosenthal 2005a: 339–364.
- Rosenthal, David (2008). Consciousness and Its Function. *Neuropsychologia* 46: 829–40.
- Welshon, Rex (2002). Conscious and Unconscious Perspectives. *International Studies in Philosophy* 34: 119–33.

