

Virtuous Homunculi: Nietzsche on the Order of Drives

MATTIA RICCARDI

Universität Bonn

This is the penultimate draft of a paper forthcoming in *Inquiry*. Please quote the printed version.

1) Introduction

It is now widely held that drives are the primary explanatory items of Nietzsche's philosophical psychology.¹ Though great progress has been made in recent years towards a proper understanding of this notion, Nietzsche's descriptions of how the drives interact with each other have proved particularly impervious. The interpretive problem is made vivid by Nietzsche's famous characterization of the 'soul' as a 'society constructed out of drives and affects' (BGE 12).² This suggests that the drives are ordered in a way resembling that in which societies are ordered. Indeed, Nietzsche often talks of the drives' order as of some kind of hierarchical arrangement. Unfortunately, it is far from clear in which sense the relations among the basic constituents of human psychology are supposed to be similar to the hierarchical relations on which societies are based. Moreover, this interpretive puzzle immediately raises a deep philosophical issue. Whereas social relations occur between persons, relations among whatever constituents one thinks the mind is made up of occur at what philosophers usually call the subpersonal level. But postulating relations among subpersonal items that mimic those among persons seem to lead one directly to the most specious form of homuncularism. For doesn't this mean that the mind ends up just being

¹ See, among others, Richardson, *Nietzsche's New Darwinism*; Anderson, 'Nietzschean Self'; Janaway, 'Morality, Drives, and Human Greatness'; Clark and Dudrick, *Soul*; Leiter, 'Sign-Language of the Affects'; Welshon, *Nietzsche's Dynamic Metapsychology*; Katsafanas, *Nietzschean Self*. See, however, Stern, 'Against Nietzsche's "Theory" of the Drives', for a more skeptical take on about the import of this notion in Nietzsche's writings.

² This characterization of the soul does not only mention drives, but also affects. More on affects in section 6.

populated by mysterious tiny agents? So it seems that no model derived from the sphere of social transactions between persons can be meaningfully applied to the mind's subpersonal constituents. This means that no satisfactory solution of the interpretive puzzle can avoid engagement with the philosophical issue concerning nature and scope of subpersonal explanations.

The aim of my paper is to offer a construal of Nietzsche's conception of the drives' order that avoids committing him to vicious forms of homuncularism. I shall proceed as follows. Section 2 sets the stage by briefly introducing Nietzsche's notion of the drives, whereas section 3 clarifies what makes a certain view of the mind a case of vicious homuncularism. Sections 4 and 5 critically survey two recent interpretations of the drives' order, which I call the vitalistic reading and the normative reading. Section 6 develops my own Hume-inspired dispositional reading.

2) Drives: A Sketch

Drives are the primary explanatory *posita* of Nietzsche's psychology. But what are such drives? Most scholars agree that at bottom they are behavioral dispositions toward specific patterns of goal-directed behavior.³ For instance, the hunger drive disposes the agent toward searching for food. The cruelty drive disposes the agent toward inflicting pain on sentient creatures. Moreover, as Richardson (*Nietzsche's New Darwinism*) argues, drives are dispositions rooted in one's biological constitution. As an unpublished note puts it, a drive is constituted by '[s]eparate parts of the body telegraphically connected' (KSA 10, 308). Of course, there are many behavioral dispositions matching this description we would not count as drives, such as contracting and dilating the pupils according to light conditions. So what makes a certain behavioral disposition a drive? First, drives have a characteristic urging character: when hungry I don't merely wait for food to appear in front of me, but I feel a craving that makes me actively look for it. Second, drives produce 'affective orientations' that shape one's experience of the world (Katsafanas, *Nietzschean Self*, 94; see also Clark and Dudrick, *Soul*, 124 and 128). For instance, the sex drive causes a characteristic arousal that makes me alert to potential mating partners. Third, and precisely in virtue of so being affectively loaded, drives embody an evaluative perspective: 'each 'drive' is the drive to 'something good', as seen from a certain standpoint; there's valuation (*Werthschätzung*) in it' (KSA 11, 167). Hunger, for instance, makes the ham sandwich on the kitchen table appear as

³See Richardson, *Nietzsche's New Darwinism*; Janaway, 'Morality, Drives, and Human Greatness'; Welshon, *Nietzsche's Dynamic Metapsychology*; Katsafanas, *Nietzschean Self*.

‘something good’. Fourth, drives are best understood as having both an aim and an object (see, again, Katsafanas, *Nietzschean Self*, 101). Taking again the hunger drive, its aim is nourishment, whereas the ham sandwich is its current object. This means that whereas the drive’s aim identifies the pattern of behavior typical of a certain drive, its object only plays a contingent and merely punctual role: it’s just the item that satisfies the drive on a certain occasion. Moreover, it is in virtue of having an aim distinct from its objects that drives act as a constant source of motivation. They are structuring determinants of human agency.

Let me conclude by stressing two points that will prove relevant to my further discussion. First, drives have a physical as well as a mental side. On the one hand, they are dispositions biologically realized in one’s organism. On the other hand, the fact that they embody a specific evaluative perspective is sufficient to make them mental *qua* representational. Second, drives can be functionally identified in terms of the typical pattern of behavior they incline us to and of the typical affects they induce. As Nietzsche somewhat obscurely puts it in an unpublished note, they are ‘actions, sensations (*Empfindungen*) and feelings grown together, self-organizing and self-nourishing’ (KSA 10, 304).

This brief sketch should suffice to adequately address the interpretive puzzle concerning the nature of the drives’ mutual arrangement. However, before I go on to this task it will be helpful to clarify how homuncularism threatens Nietzsche’s view of the drives’ order.

3) The Spectre of the Homunculus

That the drives are conceived as subpersonal items immediately raises the problem of homuncularism. In philosophical jargon, homunculi are human-like psychological entities introduced in order to explain a certain cognitive function. Explanations that appeal to such entities are usually looked at with suspicion, as they easily commit the so-called homunculus fallacy. This can happen in two different ways.⁴

The first version of the fallacy occurs when one ascribes to a subpersonal constituent properties or states that are necessarily personal-level. The problem results in this case from a category mistake: properties or states belonging to the whole are ascribed to one of its part. For this reason, this fallacy is sometimes called the *mereological* fallacy. Of course, its possible range will depend on which properties one takes to be necessarily personal. Uncontroversial candidates seem to be phenomenal states like pains. As Dennett stresses in

⁴ The following discussion is greatly indebted to Drayson, ‘Uses and Abuses’ and ‘The Personal/Subpersonal Distinction’.

his original discussion of the personal-subpersonal distinction, ‘when we abandon the personal level in a very real sense we abandon the subject matter of pains as well’ (Dennett, *Content and Consciousness*, 93-4). The reason why talk of pains and sensations proves inadequate when one moves to the subpersonal level is simple: states of that kind can be meaningfully ascribed only to persons (or other sentient creatures), and not to subsystems of their cognitive architecture, nor to parts of their brain. Things are arguably more controversial when it comes to different kinds of mental properties and states, like intentional ones.

The second version of the homunculus fallacy occurs when one explains a certain personal-level capacity by positing a subpersonal entity that already possesses the capacity to be explained. The problem here is that any account of this form is simply explanatorily idle, for it presupposes the very capacity it is supposed to elucidate. For this reason I shall call this the *explanatory* fallacy.

In Dennett’s own words, providing subpersonal explanations consists in ‘analyzing a person into an organization of subsystems ... and attempting to explain the behavior of the whole person as the outcome of the interaction of these subsystems’ (Dennett, *Brainstorms*, 152). Importantly, this strategy is *not* meant to rule out that mentalistic vocabulary be meaningfully used in subpersonal explanations. Quite to the contrary, it has been introduced precisely to allow for such a use (see Drayson, ‘Uses and Abuses’, on this point). Therefore, as long as one avoids either version of the fallacy, one is allowed to posit homuncular constituents of the mind and describe their behavior with the same vocabulary we usually employ to describe that of whole persons. This approach can thus be called *virtuous homuncularism*.

Let me now turn to Nietzsche. To start with, note that he seems to be aware that theories of the human mind may commit either version of the homunculus fallacy. On the one hand, the critique Nietzsche levels against Kant’s transcendental ‘faculties’ is clearly motivated by the idea that they are explanatorily idle *posita*: ‘How are synthetic judgments *a priori* possible? Kant asked himself, – and what really was his answer? *By virtue of a faculty*, which is to say: *enabled by an ability*’ (BGE 11). On the other hand, the fact that he warns his readers against different forms of anthropomorphism—most notably, in GS 109—suggests that he rejects views that consider the mind as actually composed by little human-like agents.⁵

Nonetheless, some scholars have argued that his treatment of the drives suggests that he explicitly endorses such a homuncular picture of human psychology. Peter Poellner, in particular, has been very clear on this point. He claims that Nietzsche ‘ultimately treats drives

⁵ Thanks to Paul Loeb, who made me appreciate this point.

not as attributes of agents ..., but as agents themselves' (*Nietzsche's Metaphysics*, 174). This is shown by the fact, he argues, that 'when it comes to specifying the actual mode of operation or agency of these drives, ... Nietzsche invariably uses intentional-mentalistic terms' (215). From this, Poellner—and, more recently, Gardner ('Nietzsche and Freud', 375)—concludes that Nietzsche's view of the drives amounts to a rather confused form of fallacious homuncularism.

Though Poellner's portrayal of how Nietzsche typically talks of the drives is correct, the conclusion he draws seems overhasty. As we saw, use of personal-level vocabulary in subpersonal explanations doesn't entail that one commits the homunculus fallacy (in any of its two versions). Nonetheless, as we shall see in the next section, whenever Nietzsche tries to describe how the drives interact his propensity to mentalistic talk becomes indeed alarming. Importantly, this does not only mean that, at least at face value, the homunculus fallacy constitutes a serious threat to Nietzsche's conception of the drives' order. It also poses a crucial challenge to interpreters, for any reading that construes this conception as perpetrating that fallacy would arguably condemn to failure his core view of the human mind.

One could try to block the homuncular worries by arguing that Nietzsche's mentalistic talk shouldn't be taken at face value at all. This strategy is pursued by Richardson (*Nietzsche's New Darwinism*), who maintains that Nietzschean drives are not mental items, but rather selected biological dispositions. The main motivation behind Richardson's purely biological reading is that it allows to make sense of the drives' apparent intentionality merely in terms of their adaptive goal-directedness and hence to dispense with any kind of problematic mentalistic talk.

Richardson is right in rejecting those readings that construe the drives not only as 'conscious', but even as "“previewing” or “preconceiving” their outcomes unconsciously' (*Nietzsche's New Darwinism*, 36). Nonetheless, this hardly suffice to show that they are strictly non-mental *posita*. In fact, several considerations strongly support that drives are mental. First, it is hard to see how they could otherwise be the basic *posita* of *psychological* explanations. Second, Nietzsche's mentalistic talk seems too pervasive to be interpreted away so thoroughly. Third, as we saw, that a drive embodies a certain evaluative attitude seems to indicate that it is in some—perhaps minimal—sense representational and, consequently, mental. Fourth, that drives are biologically realized doesn't rule out that they also possess distinctively mental properties. Freud, for instance, stresses precisely their hybrid organic-cum-psychological nature: as the 'psychic representative of the stimuli coming from the inner parts of the body and reaching into the soul', the drive is a 'border concept between the psychological (*Seelischem*) and the somatic' (Freud, 'Triebe und Triebchicksale', 85).

Another suggestion to sidestep homuncular worries is due to Katsafanas (*Nietzschean Self*, 97), who argues that instead of thinking that ‘drives, *considered in isolation*, can reason, evaluate and interpret’, we should rather hold that they are able of so doing only as ‘*embodied drives*’, ‘as part of a whole organism’. Though Katsafanas is right that we should conceive of the drives not as free-floating loci of mentality and agency, but as constitutively integrated within the organism, this doesn’t rule out fallacious forms of homuncularism. As Keil, (‘Homunkulus-Fehlschluß’) clearly demonstrates, views of the human mind so different like those of Plato, Aristotle, Kant and Freud are all plagued by the problem of homuncularism. The underlying reason is that as soon as one partitions the human mind into different parts—that may be called faculties, systems, capacities, departments or, indeed, drives—, one is irremediably confronted with it. From then on there’s no easy way out.

To conclude, there seems to be no straightforward way to argue that Nietzsche’s drive psychology is immune to the homunculus fallacy. The challenge this poses to any view that breaks the mind into several components is serious and requires a careful treatment. In the remainder of the paper I shall explore whether an understanding of the drives’ order is available that succeeds in eschewing both its versions.

4) The Vitalistic Temptation

As we saw, according to Nietzsche our ‘soul’ is to be understood as a ‘society constructed (*Gesellschaftsbau*) out of drives and affects’ (BGE 12). Nonetheless, concrete illustrations of how such a ‘society’ is supposed to look like are frustratingly rare in his published work. BGE 19, where Nietzsche sketches a view of human volition as resulting from the combination of different psychological ingredients, is a notable exception. However, this passage does not directly illuminate the characterization of the soul offered in BGE 12. Rather, it is the source of further puzzles. First, though in BGE 19 Nietzsche explains how volition works in term of command-obedience relations among different psychological ingredients, no mention of the drives is made in this aphorism. Second, Nietzsche does not identify the resulting hierarchical order with the soul, as in BGE 12. Rather, and quite strikingly, he claims now that the ‘body’ is a ‘society constructed (*Gesellschaftsbau*) out of many souls’ (BGE 19).

It is not at all clear how these strangely intertwined characterizations of soul and body are supposed to relate. The soul, which is the *definiendum* of BGE 12, is now mentioned in the definition of what a body is: a collection of souls. Moreover, Nietzsche uses the notion of social structure or order (*Gesellschaftsbau*) in both cases, but it is not obvious it picks out the same phenomenon. Even more puzzling is perhaps the fact that he *explicitly* chose *not* to

reemploy in BGE 19 the original characterization of the soul as consisting in the social structure of our drives and affects.⁶ Why provoke such a confusion in the first place?

I shall come back later to this puzzle. For now, let me start by focusing on a certain way to make sense of the specific idea that the body is a society constituted by many souls, as stated in BGE 19. As the published work doesn't offer any clear guidance here, it is tempting to look at the *Nachlass* from 1884-1886 in the hope to gather some helpful insight. A double-edged reward awaits those pursuing this strategy. On the one hand, one will indeed find many unpublished notes describing the body as a multiplicity of souls. On the other hand, such notes display the kind of marked propensity to an alarming mentalistic talk stressed by Poellner. What follows is a representative sample of such passages.

Nietzsche writes that the human body is 'a prodigious alliance of living beings, each dependent and subservient and yet in a certain sense also commanding and acting out of its own will' (LN, 29-31). Later in the same note, he claims that such a 'prodigious synthesis of living beings and intellects' as we are can work only 'once that subtle system of connections and mediations, and thus lightning-fast communication between all these higher and lower beings, has been created'. Even more strikingly, he characterizes such 'beings' allegedly constituting ourselves as both 'intelligences' and 'consciousnesses'. In another note, he exhorts to 'understand the ruler and his subjects as being *of the same kind*, all feeling, thinking, willing—and that wherever we see or sense movement in the body, we learn to infer a kind of corresponding, subjective, invisible life' (LN:44). 'Fellow-feeling—we read in another note—exists only in social formations [*socialen Bildungen*]', 'one of which is the human body, whose individual living beings "feel with" one another' (LN:50).

Of course, recruiting these unpublished notes to shed light on the published aphorisms leads one to ascribe to Nietzsche something like an incredible form of vitalism that inevitably commits him to the homunculus fallacy. The one put forward by Wotling ('What Language Do Drives Speak?') is a recent example of this kind of *vitalistic* reading of the drives' order, as I shall call it. Wotling maintains that Nietzsche's notion of drive involves 'the idea of a perceptive apparatus, but an elementary, infra-conscious apparatus' (74). Accordingly, he ascribes to the drives the 'capacity to perceive or to be affected by other drives' (75). Such 'perceptions' underlie the kind of 'communication' occurring among the drives, which in turn enables them to enter command-obedience and agreement relations. Though he recognizes that 'one cannot avoid reading these lines [those of the aforementioned unpublished passages]

⁶ In BGE's print-manuscript, the characterization of the soul originally offered in BGE 12 appears also in BGE 19 (see KSA 14: 350). It was only at the very last stage of book production that Nietzsche decided to alter the text and expunge this second occurrence.

with astonishment', Wotling takes them as expressing Nietzsche's considered view. Consequently, he does not refrain from attributing to the drives mental capacities like consciousness and perception.

Some historical consideration may be adduced in support of this kind of vitalistic reading. The picture suggested by the *Nachlass* materials resembles Leibniz' view about the composition of the human body: '[e]ach living body has a dominant entelechy, which in the animal is the soul; but the limb of this living body are full of other living beings, plants, animals, each of which also has its entelechy or dominant soul' (Leibniz, *Essays*, 222; see Look, 'Monadic Domination'). What we have here is an open-ended hierarchy of animated beings—from the macroscopic level of the whole organism down to its most microscopic parts. Perhaps surprisingly, this Leibnizian view of the (organic) world were still alive in the second half of the 19th century. Nietzsche, who read and was at least partially influenced by the work of some neo-Leibnizian philosophers, was surely familiar with these ideas. In order to give a sense of this historical context, I shall briefly consider one such work that proves particularly pertinent.

The author I have in mind is the French philosopher and sociologist Alfred Fouillée. In his 1880 book *Science sociale contemporaine*, which Nietzsche knew, Fouillée puts forward a sociological theory based on the traditional analogy between organisms and states. He thus dwells on the functioning of organisms, including the human animal. By explicitly taking up ideas of Leibniz and Diderot, he claims that 'in every organized individual contemporary science [...] shows us a world of other organized beings' (Fouillée, *Science sociale*, 84). Accordingly, organisms should be understood as 'societies of tiny animals (*sociétés d'animalcules*)' (91). Moreover, Fouillée's book is pervaded by the tendency to ascribe personal-level features to such *animalcules*. For instance, he boldly asserts that 'the possibility to feel and somehow to feel oneself feeling' as well as 'some more or less latent consciousness' are also to be found 'in all the elements of a living body' (221). Of course, that we find the same kind of bizarre claims Nietzsche makes in the *Nachlass* in a book he read doesn't prove much. However, this kind of vitalistically flavored talk—together with the panpsychism it seems to entail—was far from uncommon among Nietzsche's contemporaries. For instance, in a lecture he gave in 1878, Ernst Haeckel maintained that 'we can ascribe an autonomous soul even to every cell, a cell soul (*Zellseele*)' (Haeckel, *Zellseelen*, 15).

Though considerations of this kind gives the vitalistic reading some historical footing, I believe we should strenuously resist it. The obvious reason is that the position one thereby ascribes to Nietzsche is utterly unattractive. Moreover, further considerations speaks against it. First, by comparing unpublished with published materials it becomes clear that Nietzsche

shifts from an arguably confused biological view of the composite nature of the human self to a more plausible psychological one.⁷ *Nachlass* notes like those quoted above typically describe the human body as a hierarchically organized collection of some sort of tiny ‘living beings’ or even ‘cells’. Remarkably, no mention of the drives is made. In his published works, on the contrary, the drives figure as the basic constituents of the self. Second, Nietzsche’s clear rejection of teleological explanations and, more generally, his project of a ‘dehumanization of nature’ is at odds with the vitalism of Leibniz’s natural philosophy and of its 19th century descendants (KSA 9:525; see also GS 109). That’s the reason why in another *Nachlass* note he points out ‘a disguised polytheism in the monads’ (KSA 9:522). This leads to the conclusion that Nietzsche entertained a somewhat vitalistic picture of the human organism only to soon realize its dim prospects.

Of course, the problem remains of how to make sense of the puzzling phrasing we found in BGE 19. Why does Nietzsche opt for writing there that the ‘body’ is a ‘society constructed out of many souls’ instead of repeating the more kosher formulation already provided in BGE 12? Doesn’t this indicate that he is still sympathetic with the kind of vitalistic picture I am trying to quarantine him off? Not necessarily. As we saw, drives are best understood as having both a physical and a mental side. In virtue of their being biologically realized, they can be seen as subsystems of the human body. In virtue of their being psychological dispositions, they can be seen as subsystems of the human soul. My suggestion is thus that the phrasing adopted in BGE 19 aims at stressing that the *body* can be divided into subsystems the identification of which is only possible at the *psychological* level. Conversely, the phrasing adopted in BGE 12 can be read as stressing that the human *soul* can be divided into subsystems that are essentially embodied, i.e. biologically realized. To put it differently, both formulations are intended to convey the idea that drives are both physical and mental. Far from proving a form of vitalism, the resulting picture of the human mind is best understood as a version of non-reductive physicalism according to which the basic psychological *posita*, though biologically realized, can only be properly described in mental terms. This means that, though I am body ‘through and through’ and ‘soul is merely a word for something about the body’ (Za, ‘On the Despisers of the Body’, translation changed), as Zarathustra has it, psychological explanations are autonomous and cannot be reduced to merely biological and physical ones.

⁷ Sommer (*Kommentar*, 152), who carefully considers the textual genesis of BGE 12, comes to a similar conclusion.

To conclude, though BGE 19's puzzling phrasing may tempt one to take the vitalistic-sounding *Nachlass* from 1885 to express Nietzsche's considered position, there are good reasons—both philosophical and textual—to resist this move.

5) The Normative Reading

Clark and Dudrick—C&D, from now on—have offered the most sustained exploration of Nietzsche's conception of the drives' order. The proposal they put forward on this specific matter is part of the broader project of offering what they call a *normative* reading of Nietzsche's philosophy as presented in BGE. A crucial point in their argument is the claim that Nietzsche's appeal to the notion of soul indicates that his philosophical psychology cannot be understood in purely naturalistic terms. Given that Nietzsche identifies the soul with the social arrangement obtaining among one's drives, C&D, argue that such arrangement cannot be a merely *causal* one. Rather, it should be understood as a genuinely normative one.

It is important to stress the scope of C&D's normative reading. Given that for Nietzsche a 'person's valuations reveal something about the *structure* of his soul and what the soul sees as its condition of life' (BGE 268), there surely is a normative dimension to his view of the soul. Moreover, as we saw, Nietzsche claims of any drive as such that it is 'the drive toward 'something good'' and thus involves some kind of 'valuation' (N 1884 26[72] KSA 11:167). However, what the normative reading aims at establishing is the stronger claim that the normative dimension which is part Nietzsche's philosophical psychology is *irreducibly* so. The claim is thus that whatever normative properties are ascribed to the soul conceived as the social structure of our drives cannot be traced back to non-normative—in particular, causal—properties of that structure and of its constituents. More specifically, the irreducibility of this normative dimension is framed by appeal to Sellars's famous distinction between the 'space of reasons' and the 'space of causes'.⁸ As C&D put it, the 'object of psychological investigation, the human soul, is not a naturalistic entity', but rather 'a normative entity, which exists only in and through the space of reasons' (C&D, *Soul*, 139). Importantly, this normative claim is not confined to the personal level of the soul, but extends to the subpersonal level of the drives and of their arrangement: 'the drives are arranged not merely in a causal order but in a political one' (175). So it is not just the soul that inhabits the 'space of reasons'. Its subpersonal constituents, too, possess irreducibly normative properties.

How does the Sellarsian framework apply to the specific case of Nietzsche's conception of the drives' order? Or, more precisely, what does it mean to say that that order is

⁸ For explicit reference to Sellars's distinction, see for instance C&D, *Soul*, 131.

normative in the substantive, irreducible sense advocated by C&D? A notion they appeal to in order to convey their basic idea is that of ‘recognition’. They write, for instance, that ‘one drive has a higher rank than another ... not in virtue of its causal efficaciousness, its ability to win in case of conflicts, but in virtue of being recognized as having a *right* to win such cases’ (150). Similarly, they claim that any episode of willing presupposes ‘that the commanded drives recognize the authority of the commanding drives’ (193). Recognition, in turn, should be understood in terms of rational appreciation: lower-ranked drives are not simply forced to subordination, but accept it as legitimate. This presupposes that the transaction between the drives are, in some sense, mediated by reasons.

This line of reading obviously raises serious homuncular worries. In saying that the drives behave in a way that can only be captured in terms of rational justification, C&D seem to construe them as tiny little agents already equipped with the kind of personal-level capacities Nietzsche’s drive psychology is supposed to explain. To block objections of this sort, C&D attempt to qualify their view. First, by appealing to Dennett’s defense of the homuncular strategy they contend that ‘the activities in which the drives engage are *simpler* than the activities of the person that they are ‘rung to explain’’ (198). Second, to further substantiate this point they propose to take the kind of hierarchical organization characteristic of primate societies as a suitable analogous to the drives’ order. The idea is that the behavior of both drives and primates involves irreducibly normative phenomena, for it depends on rank relations. At the same time, however, as to make sense of primates’ behavior we don’t need to ascribe to them conscious beliefs about their social standing, to make sense of the drives’ behavior we don’t need to think they are ‘*conscious* of their political situation’ either (199). In fact, we can safely assume that drives are ‘not conscious *at all*’ (199). To recap, C&D aim to show that although the drives are simpler than the persons they constitute, their behavior cannot be fully captured in merely causal terms. Like that of primates, it is constitutively normative.

The appeal to primate societies strikes me as fascinating. It’s also possible that reflection on such cases has actually contributed in some way to shape Nietzsche’s conception of the drives’ order. Consider, in particular, the descriptions he found in the book *Des sociétés animales* by French psychologist Alfred Espinas, which Nietzsche read in German translation. In discussing primate societies, Espinas writes for instance that they are based on a form of ‘subordination’ that ‘requires a leader (*Führer*) that both commands (*befiehlt*) and guides’ (Espinas, *Sociétés animales*, 485-486). So Nietzsche might have recognized certain analogies between the hierarchical arrangement of the drives and that obtaining between the leader and the subordinated members in a group of apes. Nonetheless, it doesn’t seem to me that pointing

out analogies of this sort actually helps the kind of reading C&D are proposing. Let me explain why.

First, though it is probably true that primates cannot entertain full-fledged reflexive beliefs about their own social standing, there is a wide range of cognitive capacities they surely possess. Espinas—Nietzsche’s own source on such matters—argues for instance that the emergence of rank relations among apes presupposes ‘transmission of thoughts by means of signs’ (Espinas, *Sociétés animales*, 486). More trivially, we can safely assume that primates undergo conscious perceptions and affects. Of course, no hierarchical structure at all could evolve in primate societies *independently* of mental capacities of this kind. But (some of) these are precisely the kind of necessary personal-level capacities we should *not* ascribe to the drives if we want to avoid the mereological fallacy.⁹ Second, C&D’s appeal to primate societies seems in tension with the distinction they embrace between the ‘space of reasons’ and the ‘space of causes’. Those endorsing this distinction typically assume that inhabitants of the ‘space of reasons’ are endowed with a robust form of rationality—roughly, with what McDowell calls ‘responsiveness to reasons *as such*’ (McDowell, ‘Conceptual Capacities’, 128). This kind of rationality is thought to constitutively depend on human linguistic capacities and thus to distinguish us from the rest of the animal realm. Apes, however clever, seem thus unsuitable candidates for normative transactions supposed to take place in the ‘space of reasons’. Consequently, if we assume that hierarchical relations among the drives are mediated by this full-blown form of rationality, they cannot possibly be modeled on those obtaining among primates.

In my view, these difficulties stem from a deeper problem at the heart of C&D’s normative reading and ultimately deriving from their oscillation between the naturalistic model of the mind defended by Dennett and the anti-naturalistic one underlying the Sellarsian framework. On the one hand, the followers of Sellars who argue that psychological descriptions cannot be further analyzed into causal ones typically *deny* that they can be applied at the subpersonal level. In other words, they propose that the normative-causal distinction be straightforwardly aligned with the personal-subpersonal distinction.¹⁰ This crucial feature of the standard Sellarsian story, however, is clearly rejected by C&D, for they want to ascribe normative features to the subpersonal drives. On the other hand—and precisely to find support for the idea that agent-like features can also be found at the subpersonal level—, they fall back on Dennett’s understanding of the personal-subpersonal distinction. However, though Dennett does argue that personal-level predicates can also be

⁹ See also Katsafanas, *Nietzschean Self*, 82-3, for a similar point.

¹⁰ See, for instance, Hornsby, ‘Personal and sub-personal’. On Hornsby’s position, see Drayson, ‘The Personal/Subpersonal Distinction’, sec. 3.

meaningfully applied to a person's parts, he *allows* descriptions so couched to be further analyzed into causal ones. As he writes:

Each homunculus in turn is analyzed into *smaller* homunculi, but, more important, into *less clever* homunculi. When the level is reached where the homunculi are no more than adders and subtractors ... they have been reduced to functionaries 'who can be replaced by a machine.' The aid to comprehension of anthropomorphizing the elements just about lapses at this point, and a mechanistic view of the proceedings becomes workable and comprehensible. (Dennett, *Brainstorms*, 80-1)

Thus, at list in principle, there is for Dennett no limit to the further decomposition of the mind's functions all the way down to elementary processes manageable of causal description. There is no place for irreducible normativity in this picture, nor for any strict compartmentalization between 'space of reasons' and 'space of causes'.

To sum up, there seems to be a clear conceptual tension at the heart of the very notion of psychological normativity C&D adopt. Commitment to Sellars' distinction makes the idea that irreducibly normative (*qua* reason-giving) psychological descriptions can also be supplied of subpersonal states and processes completely inviable. Acceptance of Dennett's more liberal understanding of the personal-subpersonal distinction, on the contrary, does leave no room for the idea that certain states and processes are irreducibly normative. To my eyes, C&D's attempt to combine these two diverging conceptions of human psychology is what leads them to ascribing to Nietzsche a fallacious form of homuncularism. This result alone, however, suffices to make their interpretation of the drives' order unconvincing.

6) A Humean Account

My aim in this section is to produce an alternative reading of the drives' order inspired by Hume's deflationary account of relationships of authority. In a passage from the *Treatise* he writes:

We may ... remark, not only that two objects are connected by the relation of cause and effect, when the one produces a motion or any action in the other, but also when it has a power of producing it. And this we may observe to be the source of all the relations of interest and duty, by which men influence each other in society, and are plac'd in the ties of government and subordination. A master is such-a-one as by his situation, arising either from force or agreement, has a power of directing in certain particulars the actions of another, whom he calls servant. (Hume, *Treatise*, I.1.V.12)

What Hume suggests is that relations of authority among human beings should be understood in dispositional terms. That someone has authority over someone else means that the former individual has a power that she can put to use over the latter one in certain circumstances. The way this power is displayed can be made sense of in standard causal terms, though the account will arguably be much more complex than in textbook cases such as the collision of two billiard balls. Hume's point seems thus to be that we should not underestimate the explanatory power of causal accounts even when applied to the extremely intricate cases of social interaction.

The claim I want to defend is that Hume's strategy is particularly apt for making sense of the interaction among the drives, for it provides a deflationary strategy for dealing with the vocabulary of authority and hierarchy often employed by Nietzsche. This yields a clear advantage *vis-à-vis* the alternative vitalist and normative readings, which—as I have argued—cannot avoid committing the homunculus fallacy. For a dispositionalist construal of the drives' order does not need to ascribe to them *necessarily* person-level properties in order to make sense of their mutual interaction. Moreover, this kind of dispositionalist construal is in tune with the widespread dispositional and/or functional account of the drives. If this account is true of drives taken in isolation, it seems plausible to think it is true also of the way in which they interact and are hierarchically arranged.

To flesh out the deflationary strategy, I shall pick up an important suggestion made by Ken Gemes, who argues that rather than supposing 'recognition capacities on the part of the drives' one should think of their hierarchical order as resulting from their success in 'commanding resources' (Gemes, 'Freud and Nietzsche', 50). To illustrate this idea, Gemes provides the following example:

Thus consider an individual who sees a moderately large animal in his or her proximity. The drive to nourishment might prompt the individual to interact with that animal as if it were a potential source of food; so, for instance, under the influence of that drive perceptual capacities would be trained to help figure the quickest route toward that animal. On the other hand, if the drive for survival is stronger perhaps it will orient the perceptual capacities to help figure the quickest escape route from that animal's vicinity. Here it is not a case of a drive recognizing another drive and deciding to not let it be expressed but, rather, simply a matter of the stronger drive grabbing resources and thus preventing a competing drive from grabbing those same resources and using them to achieve its aim. (50)

Gemes' suggestion not only has the clear advantage of construing the drives' interaction in a way that avoids committing any version of the homunculus fallacy. The claim that cognitive resources can be recruited by the drives is also well supported by Nietzsche's texts. For instance, after listing five different methods to obtain mastery over a drive, D 109 concludes by stressing that the exercise of such methods is not something one is actually free to choose: 'that one desires to combat the vehemence of a drive at all, however, does not stand within our power; nor does the choice of any particular method; nor does the success or failure of this method. What is clearly the case is that in this entire procedure our intellect is only the blind instrument of another drive which is a rival of the drive whose vehemence is tormenting us' (D 109). Accordingly, the interaction among the drives is (at least in part) mediated by the capacity each drive has to activate other cognitive functions and processes. The same point is implicit in Nietzsche's psychological characterization of fanaticism as 'a type of hypnosis of the entire sensual-intellectual system to the benefit of the excessive nourishment (hypertrophy) of a single point of view and feeling which is now dominant' (GS 347; see also GM II 3 for an analogous description of ascetism). Here, the dominating 'point of view' and 'feeling' are arguably the evaluative perspective typical of and the affective orientation brought about by a uniquely ruling drive. This rule is secured by the constant and absolute control over the whole cognitive system. Therefore, Nietzsche explains fanaticism as a case of extremely imbalanced and thus unfortunate allocation of cognitive resources, which are all engaged by one single dominant drive.

So far I have been following Gemes' suggestion and argued that the hierarchical arrangement among the drives will depend (at least in part) on each drive's ability to compete with the other drives in mobilizing one's cognitive resources. A perhaps even more crucial point concerns the relation between drives and affects. As we saw, the characterization of the soul as a 'social structure' offered in BGE 12 does not only mention the drives, but also the affects. So what is the role affects play in shaping the soul's order?

As we saw, Nietzsche contends that the evaluative perspective embodied by each drive is affectively charged. More precisely, it is part of what a drive is to dispose one to undergo affective experiences of a certain kind any time one encounters a certain kind of item. For instance, it is part of what the sexual drive is to dispose one to experience a distinctive arousal any time one encounters a mating partner. Thus, at least basic drives seem to be coupled by default with corresponding affective experiences. However, drives and affects arguably interact also in more complex ways. The specific claim I want to attribute to Nietzsche is that affects may either facilitate or inhibit the discharge of a certain drive. Consequently, causing (and maybe recruiting) a suitable affect is a way for a drive to either hasten its own discharge

or to block that of a competing drive.¹¹ Nietzsche's description of how criminals' aggressive drives are inhibited by being stably associated with aversive affects offers a clear illustration of this point:

The criminal type, this is ... a strong person made ill. He needs a wilderness, a nature and form of existence that is somehow freer and more dangerous; this is where all the arms and armour of a strong person's instincts *rightfully belong*. His *virtues* are ostracized by society; *his liveliest drives quickly fuse with depressive affects, with suspicion, fear, dishonour*. [my italics] But this is almost the *recipe* for physiological degeneration. When somebody is ... forced to be cautious and sly for a long time just to do what he does best and likes do most, he will become anemic; and because he only ever experiences danger, persecution, and disaster from his instincts, even his feelings turns against these instincts—he feels them fatalistically. (TI, Skirmishes, 45)¹²

According to this picture, criminals are typically human beings equipped with a set of aggressive instincts systematically curbed by society. At the psychological level, this practice of social repression is effected by associating the criminal's aggressive drives with aversive affects. Any time one of the aggressive drives urges the individual towards a socially sanctioned action, one of such degrading affects will be triggered. As they become a constant source of negatively-valenced feelings, the result in the long run is complete alienation from this set of drives: they will be experienced as something extraneous to the self.¹³

We can now appreciate how the soul's order emerges from this interaction among drives and affects. In BGE 268, Nietzsche writes that 'which group of sentiments (*Empfindungen*) in a soul will be first to wake up, start speaking, and give command (*Befehl geben*) is decisive for the whole rank order of its values' (translation changed). Let me spell out this in some detail. Nietzsche believes that the drives' hierarchical order is structured through command-obedience relations among them. This passage tells us something about the way in which those relations are brought about: the relevant commands are issued by certain affective states. The emerging picture, thus, is as follows. Conflicting drives dispose one towards different patterns of behavior. Which drive will eventually succeed in orienting the whole agent towards its own aim is (at least in part) a matter of each drive's impact on one's

¹¹ That drives cause affects is implicit in Katsafanas' claim that drives 'induce' what he calls 'affective orientations'. Clark and Dudrick also refers to the 'production of affects ... that encourage or inhibit the behavior originally prompted by the drives' (Clark and Dudrick, *Soul*, 168). That drives can also 'recruit' affects is suggested by Anderson, 'Nietzschean Self'.

¹² See also GM II 24, where Nietzsche argues that a major effect of socialization and acculturation consists in the tight psychological association of our natural inclinations with bad conscience.

¹³ If negatively-valenced affects can effectively block drive discharge, it seems plausible to assume that positively-valenced affects can have the opposite effect of facilitating it.

affective life. More precisely, for a drive to become and remain dominant it needs to succeed in consistently causing (or recruiting) affects through which they can issue commands to the other drives.

As we saw, that drives form a ‘social structure’ based on command-obedience relations underlies the picture of human volition sketched in BGE 19. Here I cannot provide any remotely accurate reading of this quite dense aphorism. I simply want to register that among the many mental ingredients Nietzsche argues are involved in episodes of willing he mentions a ‘commandeering thought’ and, in particular, a distinctive ‘affect of command’ (BGE 19).¹⁴ Though the nature and role of such states remains unclear, what Nietzsche says here seems to confirm that the command-obedience relation among one’s drives is mediated by a specific kind of cognition and, in particular, affect. To put it differently, that a higher-ranked drive controls a lower-ranked drive requires that the former be able to cause a ‘commandeering thought’ and, in particular, an ‘affect of command’ targeted at the latter.

Let us now suppose that such a deflationary account is convincingly supported by the textual evidence and arguments provided so far. However, one may doubt it helps dispelling the homuncular worries raised against Nietzsche’s very notion of the drives’ order. After all, you surely noticed that sometimes I have too reverted to personal-level vocabulary to describe the drives’ interaction. So how can my dispositional reading avoid the homunculus fallacy?

Recall that Dennett’s understanding of the personal-subpersonal distinction *allows* for the use of personal vocabulary to describe subpersonal states and processes. Problems only arise if either one ascribes to the subpersonal processes and states properties that are necessarily personal-level (mereological fallacy) or if the proposed explanation proves to be circular (explanatory fallacy). As I have argued, both the vitalistic reading and the normative one fail by committing the first kind of fallacy: the former one straightforwardly so, as it ascribes to the drives properties like sensation and consciousness; the latter one due to the irreducible normative status it takes agential vocabulary to possess. None of this, however, applies to the dispositional construal on offer here. On the one hand, on my reading the drives do not need to be illegitimately equipped with personal-level capacities like sensation and consciousness. On the other hand, I do not assume that Nietzsche’s vocabulary—his talk of drives ‘dominating’, ‘commanding’, ‘obeying’, etc.—is irreducibly normative. For my strategy consists precisely in showing that, for all Nietzsche says, the states and processes he describes by appeal to such normative terms can be further analyzed into simpler states and processes describable in non-normative, dispositional terms. For instance, I suggested that a

¹⁴ The prominent role of this ‘affect of command’ over the other mental ingredients involved in human volition is made clear by Nietzsche’s claim that willing is ‘fundamentally’ that affect.

drive A ‘dominating’ another drive B is (in part) a matter of A’s being able to cause (or recruit) affects that inhibit the discharge of B. On the contrary, on the normative reading defended by C&D the relations among the drives Nietzsche describes in agential terms cannot be analyzed into merely causal interactions.

At this point, those sympathetic to the normative reading may ask why Nietzsche would use normative terms at all if they were simply dummies that lump together many simpler causal processes. My answer to this question appeals to the fact that this practice was, as it still is, extremely common not only among philosophers, but also among psychologists avowedly in the business of providing causal accounts. As an illustration, consider how Georg Schneider—whose work directly influenced Nietzsche’s conception of the drives¹⁵—describes how volition works:

in a case of choice (*Wahl*) the representation (*Vorstellung*) that presently appears as the relatively most purposive always leads to the decision (*Entschluss*), precisely because the representation of what is relatively most purposive also causes the strongest feeling and the strongest drive, and the strongest drive also gains dominance (*Herrschaft*) over the others and necessarily leads, and must lead, to action. (Schneider, *Der menschliche Wille*, 329)

There are two reasons this passage proves particularly relevant to my historically-minded argument against a literal reading of Nietzsche’s normative talk. First, the same kind of vocabulary—domination—is applied to precisely the same kind of psychological phenomenon—the interaction among the drives.¹⁶ Second, Schneider is clearly using this vocabulary to describe a merely causal interaction: the drive that dominates is just the one that happens to be stronger. Any literal reading of Schneider’s apparently normative terms is thus ruled out.

This example shows how late 19th century psychologists employed the normative vocabulary of obedience-dominance to describe psychological processes of causal nature. My suggestion is that Nietzsche conforms to this pervasive practice. Thus, his use of that vocabulary provides no evidence against the dispositional understanding of the drives’ order I am proposing.

¹⁵ The notebook MIII4b from 1883 (see KSA 10:235-324) documents Nietzsche’s extensive engagement with Schneider’s *Der thierische Wille*. Nietzsche also possessed Schneider’s *Der menschliche Wille*.

¹⁶ Interestingly, Hume already employs a similar vocabulary: ‘I cannot compare the soul more properly to any thing than to a republic or commonwealth, in which the several members are united by the reciprocal ties of government and subordination’ (Hume, *Treatise*, I.4.VI.261). Thanks to Peter Kail for this.

7) Conclusion

Not only is Nietzsche's characterization of the drives' order in terms of dominance-obedience relations obscure. According to the Hume-inspired reading put forward in this paper, the drives' order is best understood in dispositional terms. This proposal has two merits. First, it is well supported by Nietzsche's occasional descriptions of how drives interact with other mental states and cognitive capacities. Second,—and contrary to the alternative vitalistic and normative interpretations—, it avoids committing Nietzsche to any form of vicious homuncularism. Drives turn out, indeed, to be homunculi, but ones of the virtuous sort.

References

- Anderson, Lanier R. 'What is a Nietzschean Self?'. In *Nietzsche, Naturalism and Normativity*, eds. C. Janaway and S. Robertson. Oxford, New York: Oxford University Press, 2012.
- Clark, Maudemarie and Dudrick, David: *The Soul of Nietzsche's Beyond Good and Evil*. Cambridge: Cambridge University Press, 2012.
- Dennett, Daniel. *Content and Consciousness*. 2nd ed. London: Routledge, 1986.
- Dennett, Daniel. *Brainstorms. Philosophical Essays on Mind and Psychology*. Cambridge MA: Bradford Books, 1978.
- Drayson, Zoe. 'The Uses and Abuses of the Personal/Subpersonal Distinction'. *Philosophical Perspectives*, no. 26 (2012): 1-18.
- Drayson Zoe. 'The Personal/Subpersonal Distinction'. *Philosophy Compass*, no. 9.5 (2014): : 338-46.
- Espinas, Alfred. *Die thierischen Gesellschaften. Eine vergleichendpsychologische Untersuchung*. Vieweg: Braunschweig, 1879.
- Fouillée, Alfred. *La science sociale contemporaine*. Paris, Hachette: Fouillée, 1880.
- Freud, Sigmund. 'Triebe und Tribschicksale'. In *Studienausgabe. Band III. Psychologie des Unbewußten*. Frankfurt: Fischer, 2000.
- Hornsby, Jennifer. 'Personal and Sub-Personal: A Defence of Dennett's Early Distinction'. *Philosophical Explorations*, no. 3.1 (2000): 6-24.

- Hume, David. *A Treatise of Human Nature*, ed. L.A. Selby-Bigge. 2nd ed. Oxford: Clarendon Press, 1978.
- Gardner, Sebastian. ‘Nietzsche and Freud: the ‘I’ and Its Drives’. In *Nietzsche and the Problem of Subjectivity*, ed. J. Constâncio, M.J. Mayer Branco and B. Ryan. Berlin and Boston: De Gruyter, 2015.
- Gemes, Ken. ‘Freud and Nietzsche on Sublimation’. *Journal of Nietzsche Studies*, no. 38.1 (2009): 38-59.
- Haeckel, Ernst. *Zellseelen und Seelenzellen*. Leipzig: Kröner, 1909.
- Hornsby, Jennifer. ‘Personal and sub-personal: A defence of Dennett’s early distinction’. *Philosophical Explorations*, no. 3.1 (2000): 6-24.
- Janaway, Christopher. ‘Nietzsche on Morality, Drives, and Human Greatness’. In *Nietzsche, Naturalism and Normativity*, eds. C. Janaway and S. Robertson. Oxford, New York: Oxford University Press, 2012.
- Katsafanas, Paul. *The Nietzschean Self. Moral Psychology, Agency, and the Unconscious*. Oxford, New York: Oxford University Press, 2016.
- Keil, Geert: ‘Über den Homunkulus-Fehlschluß’. *Zeitschrift für philosophische Forschung*, no. 57 (2003): 1-26.
- Lampl, Hans-Erich. ‘Flair du livre: Friedrich Nietzsche und Théodule Ribot’, *Nietzsche-Studien*, no. 18 (1989): 573-86.
- Leibniz, Gottfried Wilhelm. *Philosophical Essays*, ed. R. Ariew and D. Garber, Indianapolis: Hackett, 1989.
- Leiter, Brian. ‘Moralities Are a Sign-Language of the Affects’. *Social Philosophy and Policy*, no. 30.1-2 (2013): 237-58
- Look, Brandon. ‘On Monadic Domination in Leibniz’s Metaphysics’. *British Journal for the History of Philosophy*, no. 10.3 (2002): 379 – 99.
- McDowell, John. ‘Conceptual Capacities in Perception’. In *Having the World in View: Essays on Kant, Hegel, and Sellars*. Cambridge (MA): Harvard University Press, 2009.
- Nietzsche, Friedrich. *Daybreak. Thoughts on the Prejudices of Morality*, ed. M. Clark and B. Leiter. Cambridge: Cambridge University Press, 1997.

- Nietzsche, Friedrich. *On the Genealogy of Morality*, ed. M. Clark, A.J. Swensen. Indianapolis: Hackett 1999.
- Nietzsche, Friedrich. *The Gay Science*, ed. B. Williams. Cambridge, Cambridge University Press 2001.
- Nietzsche Friedrich. *Beyond Good and Evil. Prelude to a Philosophy of the Future*, ed. R.-P. Horstmann and J. Norman. Cambridge: Cambridge University Press 2002.
- Nietzsche, Friedrich. *Thus Spoke Zarathustra. A Book for Everyone and Nobody*, ed. G. Parkes. Oxford, New York: Oxford University Press, 2005.
- Nietzsche, Friedrich. *Twilight of the Idols. In The Anti-Christ, Ecce Homo, Twilight of the Idols and Other Writings*, ed. by A. Ridley and J. Norman. Cambridge: Cambridge University Press, 2005.
- Nietzsche, Friedrich. *Kritische Studienausgabe in 15 Bände*, ed. G. Colli and M. Montinari. Berlin, New York: De Gruyter, 1980.
- Nietzsche, Friedrich. *Writings from the Late Notebooks*. Ed. R. Bittner. Cambridge: Cambridge University Press, 2003.
- Poellner, Peter. *Nietzsche and Metaphysics*. Oxford: Clarendon Press, 1995.
- Richardson, John. *Nietzsche's New Darwinism*. Oxford, New York: Oxford University Press, 2004.
- Schneider, Georg Heinrich. *Der thierische Wille*. Leipzig: Abel, 1880.
- Schneider, Georg Heinrich. *Der menschliche Wille vom Standpunkte der neueren Entwicklungstheorien (des Darwinismus)*. Berlin: Dümmlers, 1882.
- Sommer, Andreas Urs. *Kommentar zu Nietzsches Jenseits von Gut und Böse*. Berlin, Boston: De Gruyter, 2016.
- Stern, Tom. 'Against Nietzsche's "Theory" of the Drives'. *Journal of the American Philosophical Association*, no. 1.1 (2015): 121-40.
- Welshon, Rex. *Nietzsche's Dynamic Metapsychology. This Uncanny Animal*. Basingstoke: Palgrave MacMillan, 2014.

- Wotling Patrick. ‘What Language Do Drives Speak?’. In *Nietzsche on Instinct and Language*, eds. J. Constâncio and M. J. Mayer Branco. Berlin, Boston: De Gruyter, 2011.