The Affective Mind:

Essays on Nietzsche's Philosophical Psychology



Christopher Charles Fowles
The Queen's College, University of Oxford

A thesis submitted for the degree of D.Phil. in Philosophy
Trinity Term, 2019

Contents

Contents	i
Abstract	iii
Acknowledgements	v
Abbreviations	vii
Introduction	1
1. The Heart of Flesh	9
2. Conscious and Unconscious Thought	47
3. The Social Structure of Drives	83
Conclusion	127
Bibliography	129

The Affective Mind:
Essays on Nietzsche's Philosophical Psychology

Christopher Charles Fowles
The Queen's College, University of Oxford

A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy
Trinity Term, 2019

Abstract:

This thesis comprises three stand-alone papers, each concerning the philosophical psychology that can be drawn from the writings of Friedrich Nietzsche. The thread running throughout the thesis is a focus on *affective states*.

The first paper ("The Heart of Flesh") concerns the question of what exactly Nietzsche thinks emotions and moods *are*. I argue, *contra* received opinion, that Nietzsche had some rather sophisticated and substantive views about affects, comprising a *somatic* account of the role of the body and an *interpretive* account of the process through which we identify our own states. I also demonstrate some of the ways in which these views surface in works such as *On the Genealogy of Morality* and *Twilight of the Idols*.

The second paper ("Conscious and Unconscious Thought") provides the first reconstruction of Nietzsche's enigmatic claims about thinking. Starting from §354 of *The Gay Science*, I draw on the view of affects laid out in the preceding paper to explicate Nietzsche's (broad-brushstroke, but philosophically interesting) account of non-linguistic, unconscious cognition. I then turn to his remarks on conscious thinking, addressing two influential readings of the claim that conscious thought alone "*takes place in words*" (GS §354). Neither, I argue, is entirely successful. I propose instead an alternative that draws on work on inner speech.

The third paper ("The Social Structure of Drives") concerns Nietzsche's suggestion in *Beyond Good and Evil* that the soul should be understood as a "social structure of drives and affects" (BGE §12). I first examine the only extant treatment of this metaphor in the literature, arguing that it is philosophically and textually problematic. In its place, I suggest a new reading, which draws on Nietzsche's doctoral notes and his subsequent engagement with physiology. By the time of his remarks on the formation of the state in the *Genealogy*, I contend, Nietzsche's view has developed considerably into a precise and interesting view.

Word Count: 50,326 words

Supervisors: Peter Kail (St. Peter's College); Stephen Mulhall (New College)

Why I write such good papers

This thesis concerns my two greatest philosophical interests. The first of these is the work of 19th century German philosopher, Friedrich Nietzsche. The second is contemporary philosophical and empirical work on emotion. My introduction to both of these areas came during my time as a B.Phil. student, and the topics seemed to dovetail rather nicely. Although I have been fortunate in the six years since I moved to Oxford to study a broad range of philosophical topics with many fascinating people, nothing prompted in me the same interest as Nietzsche's work. I wrote my first paper (on self-creation, which I was recommended as a "good topic to start with") in Trinity term 2014, and started my B.Phil. thesis on his philosophy of mind in Hilary 2015. Since then, Nietzsche's writings on philosophical psychology, and on emotion in particular, have occupied my working hours—and a fair few of my non-working hours—near continuously. This thesis is the end result of five years of trying to better understand Nietzsche's often-brilliant, sometimes-frustrating, yet always engaging reflections on human psychology. I sincerely hope that it makes a contribution to debates on these topics.

I would like to acknowledge the Arts and Humanities Research Council for providing a doctoral studentship, as well as the Master and Fellows of The Queen's College for election to the Neda-Agha Soltan Graduate Scholarship in Philosophy during my time as a B.Phil. student. I also received generous support for workshop and conference attendance from both my college and the philosophy faculty, for which I remain most appreciative. The administrative staff at the philosophy faculty have been superb throughout my time at Oxford, and it would be remiss of me not to convey my appreciation to Iris Geens, Sarah Panko, and Sally Baume for knowing better than I do where I am supposed to be, and what I am supposed to be doing.

All of the material in this thesis has benefitted enormously from the feedback I received at a number of workshops over the course of my time as a doctoral student. I have presented versions of this material at the Nietzsche seminar at Birkbeck, University of London; the 'Nietzsche in the North East' workshop at John Cabot University, Rome; the International Society for Nietzsche Studies Annual Workshop in London; and at the Post-Kantian Research Seminar here in Oxford. I would like to thank Ken Gemes and Andrew Huddleston, Manuel Dries, Tom Bailey, and the ISNS organizing committee for the invitations to present my work. I would also like to express my gratitude to the audiences at each of these events for their perceptive and insightful comments that spared me many an otherwise unnoticed error. Thanks are also owed to the *Journal of the History of Philosophy* and *Inquiry* for permission to present papers one and two in this thesis, which were published in those journals respectively. The Nietzsche community more broadly has provided a remarkably welcoming and supportive environment for a young scholar, and I would particularly like to thank Ken, once more, in addition to Brian Leiter, and Mattia Riccardi for their continued encouragement and support.

I come, now, to the title of this acknowledgements section, which I hope the reader appreciates was typed with my tongue lodged *very* firmly in my cheek. The answer, in short, is that whatever value or interest might be found in the following pages—and I hope that there is some—it is as much a consequence of the brilliant, kind, and supportive people with whom I have been so fortunate to be surrounded as anything I have done. I have been very lucky at every stage to have had wonderful supervisors. I would like to thank Clayton Littlejohn at King's College, London for his guidance during my 'conversion' to philosophy, and Guy Kahane, for welcoming me to Oxford as a B.Phil. student with such patience and kindness. I am particularly thankful to my D.Phil. supervisors, Peter Kail and Stephen Mulhall. I have learned a great deal over the last four years working with both of them, and hope that this thesis is an adequate return for the time they have so generously spent reading and my discussing work. I wish to thank Peter in particular, for it was his series of lectures in 2014 that sparked my interest in Nietzsche. He has been unable to rid himself of me for over five years now, and yet still remains as generous, encouraging, and committed a supervisor as one could ask for.

I would like to conclude this acknowledgements section by thanking the people to whom I am most indebted: my friends and family. Oxford has been a wonderful place to live and to do philosophy, and I have benefitted enormously both in and outside of work by being able to spend my days surrounded by wonderful people. I would like to offer particular thanks to Milena Bartholain, Laurenz Casser, Sam Clarke, Luke Davies, Alive Evatt, AJ Gilbert, James Kirkpatrick, Annina Loets, and Tom Walton for their friendship, humour, and warmth. Most of all, I would like to express my sincerest thanks to David and Hilary Fowles, my brother Richard, John and Jeanne Griffiths, and Ruth O'Loughlin, both for their love and support in my years as a graduate student, and their relentless, unyielding lack of interest in 19th century German philosophy, which has kept me sane.

Christopher Charles Fowles

Iffley,

October 2019

Abbreviations

I use the following standard abbreviations for Nietzsche's writings (in translation and in German) throughout the thesis:

A:	*The Antichrist*
BGE:	*Beyond Good and Evil*
CW:	*The Case of Wagner*
D:	*Daybreak*
EH:	*Ecce Homo*
eKGWB:	*Digitale Kritische Gesamtausgabe Werke und Briefe*
eKGWB/BVN:	*Nachgelassene Fragmente*, *Briefe von Nietzsche*
GM:	*On the Genealogy of Morality*
GS:	*The Gay Science*
HH:	*Human*, *All Too Human*
KGW:	*Kritische Gesamtausgabe Werke*
TI:	*Twilight of the Idols*
TL:	*On Truth and Lies in the Extra-Moral Sense*
WTP:	*The Will to Power*
Z:	*Thus Spake Zarathustra*

Introduction

> The whole opinion of the rank of passions: as if the right and normal thing were to be led by *reason*—while the passions are abnormal, dangerous, half-animal and, moreover, according to their aims, nothing more than *desires for pleasure*… […] The misjudgement of passion and *reason*, as if the latter were an entity in its own right, as opposed to a relation-state of various passions and appetites; and as if it were not so that every passion had its quantum of reason…
>
> (eKGWB 1887: 11 [310]).

1.

This thesis concerns the philosophical psychology found in the writings of Friedrich Nietzsche. Nietzsche's reflections on the mind have attracted considerable attention in the last decade or so, as commentators have argued convincingly that his thoughts here evince a sophistication and richness previously overlooked. Despite this increase in attention, and some very fine work on the topic having been produced, there remain important issues that have been unduly neglected in the literature to date. The aim of this thesis is to reconstruct and examine Nietzsche's views on several topics that I believe to be important to his broader psychology, and of relevance to ongoing debates elsewhere in Nietzsche scholarship. The thread connecting the three independent papers in this thesis is a focus on Nietzsche's views on *affective states*, and their role in various psychological processes.

Few issues have occupied philosophers for longer than that of the relationship between our passions and the rest of our mental lives. Yet, despite this enduring interest and the familiarity of moods and emotions, questions regarding what emotions *are*, and what *role* they do or should play in human psychology, remain stubbornly controversial. Are passions akin to judgements or feelings, attitudes or ways of seeing? Do they aid or hinder us? Should emotions and moods be brought under the command of reason, or ought reason only to be the "slave of the passions",

as Hume so famously put it (Hume, THN II.3.3: 415)? Historically, most philosophers have approached these questions by making a firm distinction between the rational and emotional parts of the mind. Emotions are commonly presented as cruder, often disruptive, psychological episodes; vestiges of an animal past, at odds with the distinctly human faculty of reason. This view, however, has come under *considerable* recent pressure, especially in the sciences (see, e.g., Damasio, 1994; Asma & Gabriel, 2019), necessitating a reassessment of the standard picture.

It is in light of this that Nietzsche's writings on emotion and the mind acquire new importance. As the quotation at the beginning of this introduction makes rather clear, Nietzsche challenges what I called the 'standard picture'—the venerable opposition of passion and reason—at every turn. I hope to demonstrate in the coming pages that Nietzsche's scepticism about the traditional view of the mind is in fact well-thought out and philosophically well-motivated. Nietzsche's preference for aphoristic writing and short, dense paragraphs is apt to suggest to readers that his positions are sometimes wholly conjectural—asserted without argument or evidence. While there are a multitude of reasons for the relative neglect of Nietzsche's theoretical writings on emotion and psychology, some of which are rather understandable, it is plausible to think that *part* of this of this neglect can be traced back to an enduring belief that Nietzsche simply doesn't do theory, doesn't attempt to explicate the positions he adopts, and thus offers us something the value of which is hard to gauge.

Now, there is quite clearly some truth to the claim that Nietzsche doesn't often present his theoretical views before making use of them, and that he doesn't make his reasons for adopting them in the first place terrible clear either. Nevertheless, this is quite different from *not having* substantial theoretical positions, or *lacking* reasons in support of them. What surfaces as a paragraph, a line, sometimes even a single word in Nietzsche's published works, often has many years of careful, rigorous reflection behind it. I intend, in this thesis, to make clear what Nietzsche's views were on three key topics, and, in doing so, to demonstrate that the claims made in the passage quoted at the beginning of this introduction are philosophically well-grounded, and worth taking seriously.

2.

It is important, here, to make some methodological clarifications. My task in these papers is primarily *exegetical*. I ascribe to Nietzsche a range of views I believe him to have held, and for which I take myself have sufficient textual evidence to make a strong case. While he might have had theoretical views, Nietzsche was not a systematic philosopher, and was prone to producing particularistic analyses of given phenomena from which one has to derive his position, as opposed to general theoretical statements that are then applied to particular cases. A consequence of this is that there are occasions where Nietzsche simply does not address a question that we might like him to. There are some issues that arise in relation to views Nietzsche advances that are of considerable importance to contemporary debates on the topics, but which are not discussed in any of Nietzsche's writings. I have endeavoured, throughout these papers, to be clear whenever such a situation is encountered, sketching instead the options that would have been open to Nietzsche had he considered the lacuna in question, and highlighting the ways in which contemporary philosophers of a similar bent have broached the problem.

In each paper, the task of reconstructing Nietzsche's views has often required me to draw upon the *Nachlaß*. To be candid about my approach to using notebook materials, I have no objection to drawing on them in cases where I think the passages illuminate published claims, or clarify remarks Nietzsche makes that are somewhat ambiguous (of which there are, of course, a few). I will, on occasion, make quite considerable use of Nietzsche's unpublished writings. When doing so, however, I always attempt to connect what is being written in the notebooks to Nietzsche's published works from the period. There are a range of views on topics addressed in this thesis that arise in the *Nachlaß*, some of which can be found in multiple notes over long periods of time. In cases where I have been unable to find corresponding published material utilizing the ideas directly, or making claims that could only be explained by his holding the view in question, however, I have omitted the views from discussion.

3.

(i)

The opening paper ("The Heart of Flesh") presents a reconstruction of Nietzsche's understanding of the emotions, and of the mechanisms of self-knowledge through which we come to identify such states. I argue that close attention to Nietzsche's reflections on the psycho-physical states and sub-personal processes putatively involved in affective episodes in fact reveals the account of 'affect' with which he was working. His position, I contend, is more substantial (and, indeed, worked out) than previously assumed. Two distinct elements emerge from his remarks on affects. The first concerns the role of the body in such episodes. Nietzsche's view, I suggest, represents a form of *somatic* account, informed by the physiology of his day. On such views, one's bodily states play a causal role in the generation of emotions and moods. On Nietzsche's rather idiosyncratic version, occurrent affects are complex psycho-physiological episodes, comprising one's bodily states and the subsequent modulation of specific psychological mechanisms and processes. Emotions and moods start as conditions of the body, the sensations of which modulate the ensuing psychology, influencing (*inter alia*) the features of the environment that appear as salient, the behavioural opportunities that are highlighted, and the evaluative attitudes we take towards the things with which we are confronted.

This view of the body's role in emotion is then paired with the second strand of Nietzsche's overall picture, which concerns the process of identifying one's condition as *this-or-that* affect. The subsumption of given states under folk-psychological concepts and categories is, for Nietzsche, an *interpretive* process. One's bodily changes are connected to this interpretive process by a psychological mechanism, viz. the *Ursachentrieb*, which extracts a putative 'cause' for the state from one's surroundings. We can interpret our state as "anger", "resentment", "jealously", and so on, he insists, *only* once a cause has been posited. I demonstrate how Nietzsche's reflections on language and consciousness explain this connection, and this somewhat confounding claim. I also demonstrate that some of Nietzsche's most prominent

claims in *On the Genealogy of Morality* and *Twilight of the Idols* grow from his reflections on affective states.

(ii)

My second paper ("Conscious and Unconscious Thought") examines Nietzsche's enigmatic remarks on *thinking*. This topic is of considerable importance to several further debates, including those on Nietzsche's view of agency, responsibility, and the causal efficacy of mental states. This paper presents the first reconstruction of Nietzsche's views on thought. In §354 of *The Gay Science*, Nietzsche writes that we "think constantly, but [do] not know it", and that conscious thinking alone "*takes place in words*". These are suggestive, albeit rather opaque, remarks. In the opening half of the paper, I unpack the first of these claims, namely that there is near-constant unconscious thought. This assertion, I contend, is grounded in three further views that can be drawn out from Nietzsche's writings: (i) that the representations through which unconscious thinking operates are image-like as opposed to language-like; (ii) that affects have an integral role in thinking; and (iii) that affects are pervasive features of mental life. These combined, I argue, offer a remarkably plausible outline of what non-linguistic, unconscious thinking might be.

In the second half of the paper, I address Nietzsche's views on the difference between conscious thought and its unconscious counterpart. I focus specifically upon two influential interpretations of Nietzsche's views on *consciousness*. The first (owing to Paul Katsafanas) argues that conscious states alone have conceptually articulated content. The second (defended by Mattia Riccardi) claims that Nietzsche's remarks evince a form of higher-order thought (HOT) theory, on which consciousness of a state is a matter of having an unconscious thought about it. Neither of these, I contend, succeeds as a reading of Nietzsche. Instead, I propose an alternative interpretation that draws on contemporary work on *inner speech*. Nietzsche's remarks about conscious thinking taking place "*in words*", I suggest, refer to the articulation of thoughts in *public language*. This view, I claim, is textually well-supported. It also comfortably

accommodates the good features of the views of Katsafanas and Riccardi, without their accompanying problems. Nietzsche's picture, in turns out, bears resemblance to positions advanced by several leading contemporary philosophers and psychologists engaged in debates on the nature of non-linguistic thought and the role of affective states in broader cognition.

(iii)

The third paper ("The Social Structure of Drives") addresses Nietzsche's suggestion in *Beyond Good and Evil* that the belief in the "soul" as something "indestructible, eternal, indivisible" is an idea of which science must rid itself (§12). He offers instead the idea of the soul "as a social structure [*Gesellschaftsbau*] of drives and affects". This paper begins by addressing the sole examination of this metaphor in the literature (owing to Maudemarie Clark and David Dudrick). Although they are right to stress the importance of this choice of imagery, Clark and Dudrick's account, I contend, runs into some serious problems. In the second half of the paper, I move to argue that the social metaphor is a rich one that draws upon different sources, and develops over the course of Nietzsche's working life. What we end up with is a complex account of organic unity, and the unity of given person's psychology. The reconstruction of this picture begins with Nietzsche's preparatory notes, composed in 1868 for a doctoral dissertation on Kant and teleology (that he in fact never commenced writing). These notes reveal Nietzsche's interest in the idea that robust notions of teleology could be replaced with a naturalized account, on which 'organized beings' comprise interactions between 'living' parts.

Having established Nietzsche's interest in the issue of *Zweckmäßigkeit*—purposiveness or appropriateness—in nature, and the general framework he had developed, I move to compare this picture to the discussion of the formation of a social structure (viz. the first states) in *On the Genealogy of Morality*. Nietzsche develops his early idea considerably in light of the material he consumed in the interim. In the *Genealogy*, the structure is no longer conceived of as simply containing living parts, but as a *Rangordnung* necessarily requiring domination and subjugation. The relations described in the *Genealogy* provide a basis for us to flesh out the idea of

drives/affects being 'structured'. Nietzsche's eventual picture distinguishes between what we might refer to as the mind—the fluctuating relations between occurrent states and processes *at a given time*—and the self or soul, understood as a complex causal-dispositional structure comprising dispositions, desires, impulses, etc.

1

The Heart of Flesh†

Abstract: Nietzsche's philosophical psychology draws upon a rich and extensive terminology. Integral to this terminology is the concept of an 'affect'. Although less widely discussed than the concept of a 'drive' or 'instinct', affects are arguably every bit as important to Nietzsche's mature writings. Little attention, however, has been paid to the question of what Nietzsche thought affects were, such that they could do the considerable explanatory work he required of them. In this paper, I argue that by focusing on his reflections on both the psycho-physiological states to which we apply affect-words, and the sub-personal processes putatively involved in such episodes, we can draw out Nietzsche's understanding of 'affect'. The picture that emerges comprises a form of 'somatic' account, paired with an 'interpretive' account of the mechanisms that subsume affective states under folk-psychological concepts and categories. Such a view, I hope to demonstrate, underpins some of Nietzsche's most prominent psychological claims in works such as *Daybreak*, *On the Genealogy of Morality*, and *Twilight of the Idols*.

Keywords: Nietzsche, affect, somatic theory of emotion, interoception, interpretive accounts of self-knowledge

INTRODUCTION

In a *Nachlaß* fragment of 1888, Nietzsche refers to psychology as "Affektenlehre"—the doctrine or theory of the affects (eKGWB 1888: 13 [2]).[1] Given his contention elsewhere that psychology represents the "path to the fundamental problems" (BGE §23), it should come as no surprise

† Copyright © 2020 Journal of the History of Philosophy. This article first appeared in *Journal of the History of Philosophy*, Volume 58, Issue 1, January 2020, pp. 113–139. Published by Johns Hopkins University Press.

[1] Nietzsche's published works are cited using the abbreviations listed in the bibliography, followed by Arabic numerals. Where works comprise several volumes, books, or essays (such that sections are not numbered consecutively), either Roman numerals or abbreviated section titles precede Arabic numerals. For example, GM I, §1 is *On the Genealogy of Morality*, Essay I, Section 1; and TI "Errors", §4 is *Twilight of the Idols*, "The Four Great Errors", Section 4. Citations to the *Nachlaß* refer to year, group, and section/fragment number (eKGWB year: group [note]). Quotations from the published works are from the translations listed in the bibliography (with occasional amendments). Translations from the *Nachlaß* are my own.

that Nietzsche makes reference to affects in numerous prominent passages, and throughout some of his most important works.[2] Yet, as Peter Poellner has claimed, one might "feel that not much is gained by [Nietzsche's] assertions in the absence of a detailed account of what 'affects' are supposed to be" (Poellner, 2007: 229). Literature on the topic has focussed predominantly on the roles ascribed to affects in his moral psychology and broader philosophy, granting less attention to what Nietzsche thought affective states were, such that they could play these roles.[3]

Indeed, the situation might be even more unhappy than Poellner's remark suggests. The central explanatory concept employed by Nietzsche is that of the 'drive' or 'instinct'. In a recent, influential interpretation, Paul Katsafanas has argued that drives are, *inter alia*, "dispositions that induce affective orientations" (Katsafanas, 2013: 740). This suggestion has much to recommend it, but it does make Poellner's concern all the more pressing. If drives are dispositions that generate affective orientations, the explanatory efficacy of the concept of a drive is, in no small part, dependent upon the understanding of 'affect' we take Nietzsche to be employing. This places affects at the heart of Nietzsche's often-celebrated moral psychology, but leaves us needing to say more about what they are "supposed to be" (Poellner, 2007: 229).

My principal aim in this paper is to do just that. More precisely, I intend to focus predominantly upon some of the psychological claims that Nietzsche makes regarding affects. Such reflections are, I submit, more substantial (and, indeed, thought-out) than previously believed. Nietzsche pays considerable attention to the nature of the psycho-physiological states to which affect-

---

[2] There are over 500 references to *Affekt* or cognate words/variations thereof in the published works and notebooks, as well as around ninety instances of *Passion* etc. and over 900 uses of *Leidenschaft* etc. Of the eighty-five passages in the published works that refer explicitly to 'affect' (and so on), fifty-four appear in just four books: *Human, All Too Human*, *Daybreak*, *Beyond Good and Evil*, and *On the Genealogy of Morality*. There are places (predominantly in the *Nachlaß*) where Nietzsche appears to treat 'affect' and 'passion' as subtly distinct, with the latter sometimes seeming to refer to enduring (affective) orientations of love/hate for particular things (e.g. for justice; see eKGWB 1882: 3 [1], aphorism 40). His general usage, however, suggests the terms are broadly interchangeable.

[3] Much of the work on affects has focussed predominantly on their relation to values (see Peter Poellner, 2007; Brian Leiter, 2013; and Paul Katsafanas, 2013 and 2015). Christopher Janaway (in his 2007) emphasizes the importance of affects in relation to both the interpretation of the *Genealogy* and to Nietzsche's broader philosophy. For work that touches on the question of what affects are for Nietzsche, see Leiter, 2013; Poellner, 2007; and Rex Welshon, 2014: 126–29.

words are applied, and, I shall claim, to the sub-personal processes putatively involved in such episodes. It is his remarks on these topics that will command our attention for the majority of this paper,[4] for they reveal much about the understanding of 'affect' with which he was working. Two distinct positions emerge. The first concerns Nietzsche's account of affects themselves, which I shall argue is 'somatic' in character. The second concerns self-ascription, and the mechanisms that subsume affective states under folk-psychological descriptions. Nietzsche's view here, we shall see, is an 'interpretive' one.

Nietzsche often discusses these positions together as they present his picture of the psychological processes that take us from exposure to a stimulus through to reflective awareness of our being in a particular state. Reconstructing this account provides us with something to say in response to Poellner's worry, of relevance to both the exegesis and evaluation of Nietzsche's philosophical psychology. But this is not all. Nietzsche's remarks present us with a reason to reassess some of the features commonly attributed to his view in the literature. His considered position, I shall argue, is somewhat less orthodox (and more emendatory) than has been appreciated. Nietzsche draws heavily on the physiology of his day, affording a prominent role to the body in affective states. He also develops a picture on which affects pervade mental life, with the majority of such episodes being non-conscious. In addition, the interpretive account of self-ascription that Nietzsche favours raises the possibility of being radically mistaken about the affects one undergoes. As I hope to bring out in the course of unpacking his somatic and interpretive views, these positions are related in important ways to claims found in works such as *Daybreak*, *Twilight of the Idols*, and *On the Genealogy of Morality*.

The paper comprises three stages. In the first of these (§1), I lay out some of the claims made about Nietzsche's account of affects in the contemporary literature. I move on in the second

[4] This paper exhausts neither Nietzsche's interesting thoughts on affects, nor his psychological reflections on such states. He offers insightful remarks on the relationship(s) between affects and pleasure/displeasure, valuations, and judgements, as well as on the phenomenology of affective experience—all of which I shall studiously avoid for the most part here. (For some brief remarks, see n. 31.)

stage to tease out his somatic and interpretive views, and to consider them in detail (§2). The third stage is an examination of the process of affective states becoming conscious (§3). In the concluding section, I present the fleshed-out picture of Nietzsche's account.

1.

Our first task is to highlight some features commonly attributed to Nietzsche's understanding of 'affect', and to sketch out roughly how the account I am proposing diverges from such views. If we look to contemporary work on affects (both philosophical and psychological), we encounter a wide range: emotions, moods, various pleasures and displeasures, as well as states like hunger and more abstract desires, are all described as 'affective'. As I suggested earlier, attention to what affects were for Nietzsche has been reasonably limited. One possible explanation for this relative neglect is a scepticism about whether he had any substantial thoughts on the topic. Nietzsche's references to affective states are generally to straightforward moods and emotions, which seems congenial to the suggestion that he was working with a reasonably common-sense conception of 'affect'.

Since one could hardly accuse Nietzsche of peppering his readers with necessary and sufficient conditions for the terminology he employs, commentators have looked to draw out a position from his usage of 'affect'. Peter Poellner's own response to the concern quoted at the beginning of the paper is that affects, for Nietzsche, should be understood as "any mental episode which constitutively involves a pro- or con- attitude", with "a distinctive phenomenology—some experienced attraction or repulsion". This would include "a feeling of shame, an occurrent *desire* for something absent, as well as a *bodily sensation* experienced unqualifiedly as painful or pleasant" (Poellner, 2007: 229). If we look elsewhere, Rex Welshon claims that "affective character is, typically at least, *consciously experienced* as a particular feeling with a distinctive phenomenology" (Welshon, 2014: 127; cf. 128). Indeed, "all affects", he suggests, "have distinctive default phenomenologies" (Welshon, 2014: 129). It is this conscious qualitative character that putatively provides the "incentivizing and appetitive aspect" of the drives, acting

as "a stimulus, an impetus, a spur" (Welshon, 2014: 120). Not dissimilarly, Brian Leiter, in his treatment of the topic, suggests that affects and feelings refer "usually to the same kind of mental state for Nietzsche" (Leiter, 2013: 239).[5] Nietzsche's position, Leiter argues, is "basically *noncognitivist*: he thinks our basic affects of inclination and aversion are marked by a distinctive conscious, qualitative feel", (Leiter, 2013: 242–43) and involve a conative element, or "*motivational oomph*" (Leiter, 2013: 240).

Despite differing from one another in certain (occasionally substantial) respects, these interpretations share some ground. Poellner, Leiter, and Welshon each argue that occurrent affects are a broad class of phenomenally rich, (at least typically) conscious mental states, intimately related to feelings, inclination and aversion, and perhaps conative 'oomph'. This picture (hereafter, the 'extant view') certainly qualifies as reasonably common-sense, and surely gets much right. There is ample evidence that Nietzsche sees *some* connection between affects and states of feeling, and I have no objection to claims about inclination and aversion, or conative 'oomph'. Nevertheless, I have reservations about whether even this minimal position gets things completely right. Attention to Nietzsche's affect-psychology, I shall claim, suggests a more revisionary picture than the extant view.

Nietzsche in fact contends that we could well misunderstand affects, as he thinks we do other mental phenomena. He writes in the *Nachlaß*, for instance, that we "take ourselves to *know*, in the sensation of envy, hatred etc., *what* envy, hate etc., *is*—an error!" This is followed by a comparison: "Just as in thinking: we believe we know what thinking is". In reality, however, "we experience some *symptoms* of an illness significantly unknown to us, and opine that the illness simply *consists in this*" (eKGWB 1880: 6 [444]; cf. D §119).[6] This is a suggestive remark

---

[5] Cf. Janaway, 2007: 214; John Richardson, 1996: 37; and Lanier Anderson, 2012: 221.

[6] Nietzsche's complaints about how we conceive of thinking are broadly two-fold: first, that we present it as the activity of a substantive subject (see BGE §17; eKGWB 1887: 11 [113]; 1888: 14 [122]); and second, that we take thinking to consist in conscious thoughts standing in causal relations to one another (see eKGWB 1885: 1 [61]; 1887: 11 [113]; 1888: 14 [152]). (The latter concern is plausibly the basis of the above comparison.)

upon which Nietzsche does not really elaborate. The note itself is brisk and greatly condensed, with Nietzsche gesturing towards a number of distinct concerns. The comparison should, however, be taken seriously. Nietzsche's view, I submit, is that thought and affect are alike not simply in being misunderstood, but are misunderstood in many of the same ways.

Nietzsche's main contention is that (in both cases) we are led astray by considering only those things of which we become conscious. He follows the remark about experiencing the symptoms of an unknown illness with the claim that we "gauge and name all moral states according to *what* we consciously sense *thereof* [w a s *wir* d a b e i *bewußt empfinden*]" (eKGWB 1880: 6 [444]). Nietzsche's writings on thought suggest a number of errors purportedly consequent upon this shortcoming, two of which I shall mention here. The first is the error of failing to appreciate the *non-conscious* processes involved in a phenomenon such as thinking. Nietzsche insists that 'thinking' refers to a process more extensive than just the concatenation of thoughts experienced consciously. The "events which are actually linked", he writes, "play out beneath our consciousness", with conscious thoughts and feelings merely "symptoms of the actual event" (eKGWB 1885: 1 [61]; cf. 1885: 1 [28]).[7] The second error I wish to highlight concerns the *pervasiveness* of thinking. Nietzsche writes that "man, like every living creature, is constantly thinking but does not know it; the thinking which becomes *conscious* is only the smallest part" (GS §354). In considering just the symptoms (e.g. conscious thoughts, and so on), we fail not only to note the non-conscious components of the psychological phenomenon in question, but also its ubiquity.

Nietzsche, I believe, takes us to be guilty of the same errors in the case of affects. We misunderstand such states, at least in part because we conclude that they consist simply in their

[7] A similar picture to this emerges from Nietzsche's remarks on willing. See, for instance, D §129: "*we are accustomed not* to touch upon all [the] unconscious processes, and to think of the preparation for an act only to the extent that it is conscious" (cf. eKGWB 1886: 7 [1]; GS §360). In the note where he claims that we believe we "know what thinking is", Nietzsche compares our folk-psychological understanding of affects to our understanding of willing, claiming that we "misunderstand" willing in terms of "purposes" (eKGWB 1880: 6 [444]) (cf. n. 29).

'symptoms'. In doing so, we arrive at a conception of affects that overlooks the non-conscious elements of such episodes, and which is based upon broadly unrepresentative instances (i.e. those of which we become conscious). If this is correct, one might worry that even if the extant view succeeds in characterizing conscious affective experience, it would be hasty to conclude that it provides a full answer to the question of what affects, for Nietzsche, are "supposed to be" (Poellner, 2007: 229). The features ascribed to conscious affects might not hold for all components of the broader psychological process to which 'affect' refers, or generalize to all (or even most) affective episodes. Understanding Nietzsche's view requires us to consider the 'illness' in addition to the 'symptoms'. The latter will plausibly include conscious feelings and thoughts. The former, I propose, refers to the broader psycho-physiological process *in toto*. It is in relation to this that Nietzsche's reflections on affect-psychology, sub-personal processes, and physiology come into play—to which we shall now turn.

2.

Much philosophical work on emotion in the century or so after the end of Nietzsche's creative life relied predominantly upon introspection and conceptual analysis, with great weight placed upon the intuitions of competent speakers as to the meanings and referents of emotion words.[8] By contrast, Nietzsche's approach is far more naturalistic in character, both in regard to affects and to psychological concepts more generally. The label 'naturalistic' (in the context of discussions of emotion) simply conveys that the approach engages with the relevant contemporary sciences, perhaps constructing a theory from empirical evidence or integrating scientific results into philosophical theorizing. Such approaches might well offer correctives to folk-psychological conceptions of given states, or to the usage of particular psychological

[8] See, for example, Anthony Kenny, 1963; and Robert Solomon, 1984.

terminology.[9] For modern naturalistic theorists, the relevant sciences are generally psychology and neuroscience. For Nietzsche, and his like-minded contemporaries, it was physiology.

If we look to the *Nachlaß*, we find a wealth of fascinating remarks on affects that reflect Nietzsche's interest in examining the psychological terms he employed in light of an understanding of the body gleaned from physiology. He writes in 1880, for instance, of the "heart of flesh", insisting that "the viscera are active in the affects", and that this "roughly corresponds to Schopenhauer's 'Will'" (eKGWB 1880: 4 [218]). Nietzsche expands upon this line of thought somewhat in a passage from the following year, which clearly betrays the influence of the material he was reading. In this case, the work in question is Wilhelm Roux's *Der Kampf der Theile im Organismus* (published that year). As part of an attempted explanation of the appearance of purposiveness in nature, Roux posited a struggle between the parts of the body, with complex organic structures emerging through competition between the cells, tissues, and organs for nutriment and space in cycles of excitation (at a stimulus) and (over-compensatory) response. Nietzsche notes that now "one has rediscovered the *struggle* everywhere, and speaks of the struggles of the cells, tissues, organs, organisms". He goes on to write that "one *can* find all of our conscious affects" in these struggles, and that on such a picture, "what's really going on during the activity of our human affects are those physiological movements, and the affects (struggles, etc.) are only intellectual interpretations". Towards the end of the passage, Nietzsche states that science is "now on the path to making clear the smallest processes", and providing a "*kind of language* for those processes" (eKGWB 1881: 11 [128]).

Such passages evince Nietzsche's conviction as to the relevance of physiology to our thinking about affects. Thankfully, his discussions of the role of the body go beyond noting that physiological arousal is "what's really going on" (whatever we take that to mean). The remarks we are going to examine in the coming sections present to us an idiosyncratic form of 'somatic'

[9] The term 'naturalistic' is borrowed from Paul Griffiths' 1997.

account. As the name suggests, somatic accounts afford a prominent role to the body in affective states. More precisely, advocates of somatic theories are interested in the role of the body beyond the brain (to which 'body', etc., will refer hereafter). Of course, countenancing the involvement of bodily states in emotional episodes is hardly the preserve of somatic theorists. A common way of incorporating the presence of physiological perturbation into accounts of the emotions has been to claim that emotions *cause* bodily states. On such a view, some stimulus *S* is encountered by the individual in question and causes an affective state *A*, which in turn causes a physiological state *P*. My perception of a snake, *S*, causes the emotion of fear, *A*, which causes a series of bodily changes, *P*, such as increased heart rate or shaking. People often claim that they are shaking *because* they are scared, or that their heart is beating forcefully *because* they are excited.

Somatic accounts contest this order. The claim characteristic of the somatic approach is that bodily perturbation makes "a causal contribution" to the generation of affective mental states (Barlassina and Newen, 2014: 640). Instead of *S*, *A*, *P*, somatic theories claim the correct order to be *S*, *P*, *A*: my perception of the snake, *S*, causes a series of bodily changes, *P*, which are then causally related to the affective mental state, *A*. This is what we find in Nietzsche's writings. As above, one's encountering a stimulus, *S*, triggers a range of physiological changes, *P*. So far, so good. The idiosyncratic element of Nietzsche's picture appertains to *A*. The somatic causal claim leaves room for disagreement among theorists on a number of fronts, including which part of the process should be properly identified as the emotion (or 'affect').[10] As we saw briefly

[10] Somatic accounts can be 'pure' or 'impure', a distinction that hinges on the acceptance (pure) or rejection (impure) of what we could call the 'somatic restriction claim'—the claim that emotions are exhausted by the bodily element or its perception, broadly construed (cf. Luca Barlassina and Albert Newen, 2014: 640). Examples of pure theories are William James' 'feeling' theory (1884; 1890; 1894), and Jesse Prinz's view, on which the emotion is the perception of bodily perturbation (2004). On the above construal of the restriction claim, Antonio Damasio's account would also be a 'pure' theory. Damasio identifies the "*essence*" of emotion as the "collection of changes in body state", which is separate from the perception of "all changes that constitute the emotional response" (1994: 139). (Hence my use of the phrase "affective *mental state*" [emphasis added] to characterize *A*.) Barlassina and Newen argue for an impure account on which emotions are multi-modal states (Barlassina and Newen, 2014). Nietzsche's view, on my reading, would be a form of impure somatic theory. One distinctive feature of Nietzsche's position is that he resists the restriction claim both in relation to the multiple elements synchronically occurrent during emotional episodes (*à la* Barlassina and

in §1 (and will see again in §3), Nietzsche is deeply sceptical about the adequacy of our conception of 'affect', and appears reluctant to identify a single point in the process as the 'affect' proper, to the exclusion of all else. Instead, his preference is to amend and revise our conception of 'affect' to cover a broader psycho-physiological process. This reflects not only his naturalistic approach to the topic, but also his doubts about whether language and our intuitions about the extension and application of particular concepts offer much insight into the phenomena themselves.

On the view I ascribe to Nietzsche, an occurrent 'affect', *A*, is a complex psycho-physiological process, comprising one's bodily state (standardly one of perturbation in response to a stimulus or stimuli)[11] and the subsequent modulation of specific psychological processes resultant upon the sensation of the bodily state.[12] This is what I take him to have in mind when he writes, for instance, of "*Passions*" as "states of our organs and their retro-active effect on the brain—with a *search* for *release* [A u s l ö s u n g]" (eKGWB 1882: 4 [219]).[13] Nietzsche's writings reveal a number of ways in which he thinks such sensations modify or stimulate mental function (which we will come to in §2.2). One mechanism in particular that he claims is triggered by the sensations of bodily arousal, and to which he returns consistently, is the mechanism that seeks a 'cause' for one's condition. As we shall see, this plays an important part in Nietzsche's affect-psychology, and it is upon this feature of his account that much of our attention shall rest.

Newen), and in relation to the components of the temporally extended psycho-physiological process to which 'affect' refers.

[11] I say "standardly" since Nietzsche is sensitive to the possibility of affective states (most probably mood-like) that are *not* responses to proximal stimuli, but are instead the result of an altered physiological state (such as tiredness or hunger) modulating one's psychology in the appropriate fashion.

[12] It should be stressed that, in articulating the view this way, I do not mean to suggest an overly hard distinction between the psychological and the physical for Nietzsche. Indeed, I take it that the complex interaction of psychological processes and the body proper in affective episodes partly underpins Nietzsche's ready application of physiological language to psychological topics. (Thanks to Paul Katsafanas for raising this issue.)

[13] I take the idea of a 'search' to be figurative here, conveying that the affective disposition is sensitive to the constant stream of sensory information, and responsive to stimuli. This will become clearer in §2.1, where we will also discuss *Auslösung*.

Nietzsche's view is in fact more nuanced still, as in addition to *S*, *P*, *A*, he also addresses the points at which one becomes aware of one's state as an instance of what he calls 'general feeling', A_F, and at which one can represent oneself at the level of reflection as in a given state, A_R. Entwined into this complicated picture is his account of the psychology of self-ascription, on which certain episodes and conditions are interpreted as being particular affects. It is in relation to this issue that points A_F and A_R arise. Naturally, there is plenty more work to be done to elaborate upon the above suggestions, as well as to show that they do indeed represent Nietzsche's view. It is to this that we shall now move.

2.1

Our starting point is a *Nachlaß* passage entitled "**The belief** in **'affects'**" (hereafter 'BA'). Nietzsche writes that

> Affects are a construction of the intellect, a *fabrication of causes* which do not exist. All bodily *general feelings* that we do not understand are interpreted intellectually, i.e. a *reason* is sought for feeling one way or another, in people, experiences, etc., so something adverse, dangerous, alien, is *set* as the cause of our malaise [*Verstimmung*]: in fact, it is *sought* in *addition to* the malaise, for the sake of the *conceivability* [D e n k b a r k e i t] of our state.—Frequent blood-flows to the brain, with the feeling of suffocation, are **interpreted** as *anger*: the persons and things that rouse us to anger are releases for the physiological state.—Subsequently, after a long habituation [*Gewöhnung*], certain occurrences and general feelings are themselves so regularly connected that the sight of certain occurrences produces the state of general feeling, specifically any congestion of the blood or excitation of the semen, etc., it brings with it through close association: we then say, 'the affect is aroused'.
>
> (eKGWB 1883: 24 [20])

Nietzsche makes several interesting (if admittedly confounding) claims here, and prompts a number of questions: what, for instance, does he mean in saying that affects involve a "*fabrication*" of causes, or that a reason is sought for feeling a certain way, in "*addition to*" the state? And how is the seeking of a cause related to the interpretation and "*conceivability*" of the state? Before addressing such questions, it is worth highlighting a few pieces of technical terminology that might otherwise evade detection, and which will be important to bear in mind

as we unpack Nietzsche's view. The first appears in the suggestion that the "persons and things that rouse us to anger are *releases* for the physiological state" (emphasis added). The German word here translated as "releases" is *Auslösungen*. Nietzsche's references to *Auslösung(en)* suggest a rather particular application of the term, and almost certainly reflect his engagement with Julius Robert Mayer's "Über Auslösung" (1876), which Nietzsche had read by April 1881.[14] Mayer, a physician and physicist, conceived of *Auslösungen* as unleashings or outbursts of force (*Kraft*) under the right conditions, which is mirrored in Nietzsche's employment of the concept.[15] Indeed, Nietzsche writes explicitly of "physiological processes" as "releases of force [*Kraftauslösungen*]" (eKGWB 1884: 27 [3]). His references to *Auslösungen* (including in BA) almost always occur alongside *Reiz*, *reizen*, and so on, in relation to the thing that has stimulated, elicited, or (as above) roused the unleashed 'force'.

If we look at what Nietzsche says about *Kraftauslösungen* in the afore-mentioned note, he writes that they bring with them "a certain increase or strengthening" when they "reach the sensorium commune" (eKGWB 1884: 27 [3]). The idea of a 'sensorium commune' appeared in both philosophy and physiology, although it was more closely associated with the latter by Nietzsche's day. Roughly speaking, it denoted the hypothetical point in the brain at which there occurs a confluence of sensory information (both interoceptive and exteroceptive) from various

---

[14] Nietzsche received Julius Robert Mayer's *Die Mechanik der Wärme* (1867), a collection of papers including 'Über Auslösung', from Heinrich Köselitz. On a postcard to Köselitz, Nietzsche writes of 'Über Auslösung' as "for me the most significant and most useful [paper] in the book" (NF/BVN 1881: 103 [16/04/1881]). Capturing Mayer's use of *Auslösung* in English is difficult. Both 'trigger(ing)' and 'trip(ping)' represent possible translations, as does 'ignition' (sometimes used in work on Mayer). Walter Kaufmann's translation as "means of relieving" (WTP §670) conveys a Nietzschean sentiment (cf. HH "Man in Society", 370; see also my remarks on 'practical-psychological necessity'), but does not comport with Nietzsche's employment of *Auslösung* more broadly. 'Release' fits those passages in which Nietzsche uses the concept sufficiently well, while avoiding some of the difficulties that accompany the other translations.

[15] Mayer gives, inter alia, a gas explosion and the lighting of a match as examples of *Auslösung*. *Auslösungen*, he claims, "play not only a great and essential role" in events such as these, "but also in the living world, and especially in physiology and psychology" (Mayer, 1867: 12). If we look to Nietzsche, he of course describes *ressentiment* as "that most dangerous of blasting and explosive material" (GM III, §15). He also writes in the *Nachlaß* of the commanding affect in willing as "a *sudden explosion of force*" (eKGWB 1884: 25 [436]), and of "pleasure and displeasure, the play of affects, the feeling of discharge, explosion, freedom . . ." (eKGWB 1888: 14 [185]). One particular claim offered by Mayer was that movement was based on *Auslösungen* (which were accompanied by pleasant sensations). Not long before BA, we find Nietzsche writing that "each action builds on its affect-system" (eKGWB 1883: 7 [121]), and later of pleasure "connected with an activity itself as the release [*Auslösung*] of a bound and pent-up force" (eKGWB 1885: 1 [77]). (See also §2.2, and n. 29.)

sense-modalities.[16] This fits neatly with the second piece of terminology that I wish to flag up. If we return to BA, Nietzsche draws our attention in the second line to an unusual phrase, referring to "*general feelings* [G e m e i n g e f ü h l e]". The idea of *Gemeingefühl(e)* was also borrowed from physiology; Nietzsche writes elsewhere (in a letter to Franz Overbeck) of *Gemeingefühl*, "as the physiologists say" (NF/BVN 1887: 870 [30/06/1887]). The concept referred broadly to the interoceptive representation of the body.[17] Those bodily changes Nietzsche has in mind when talking of the viscera being "active" in the affects (eKGWB 1880: 4 [218]) come to be conceived of as 'releases', which we become aware of as states of 'general feeling'.

With these terms in place, let us move to the rest of BA. To help us understand Nietzsche's remarks therein, it is worth attending to several further passages. Although BA represents a good starting point, offering as it does a succinct sketch of the position that I take Nietzsche to hold, the claims made therein also surface elsewhere.[18] *Daybreak*, for instance, offers some nice examples. In comments deeply redolent of the physiological content of BA, Nietzsche writes that the "act of violence as a consequence of passion, of anger for example, is to be understood physiologically as an attempt to prevent a threatening attack of suffocation. Countless acts of wantonness vented on other people have been discharges [*Ableitungen*] of a sudden congestion

---

[16] Cf. Anthelme-Balthasar Richerand: "All impressions perceived by the sense organs, by the percipient extremities of nerves, are transmitted to one point in the cerebral substance. It is here that exists the *sensorium commune*" (Richerand, 1808: 284).

[17] *Gemeingefühl* can be traced in scholarly work to Johann Christian Reil's *Rhapsodieen über die Anwendung der psychischen Curmethode auf Geisteszerrüttungen* (1803), but is more widely associated with Ernst Heinrich Weber's *Der Tastsinn und das Gemeingefühl* (1846). Nietzsche's usage is consistent with the physiology of his day. Indeed, his linking of general feelings to affects was not unprecedented. See, for example, Ernst von Feuchtersleben: "Self-feeling [...] provides the ground for all other feelings and emotions. Within self-feeling, the *Gemeingefühl* takes on a human character. The pleasantness of the latter is transformed into cheerfulness, displeasure into sadness, the changing and interplay of these states is called mood" (von Feuchtersleben: 1845: 137–38). (I know of no evidence that Nietzsche was familiar with this work.)

[18] Rex Welshon has claimed that Nietzsche "never tells his readers what he thinks drives or affects actually *are*, preferring to assume, as he so often does, the categories as primitives so that he can then deploy them in his various projects" (Welshon, 2014: 108), and that beyond "generalities, little else is to be discovered in the published works to help us understand what drives and affects actually are" (Welshon, 2014: 109). I agree with Welshon that Nietzsche tends to put drives and affects to use without providing any preamble regarding how we should understand these terms. It should be noted, however, that the view I ascribe to Nietzsche here *could* be reconstructed solely from published passages.

of blood through a strong muscle-action" (D §371; cf. D §83; §86). Regarding the seeking of a cause for one's state, he writes elsewhere in *Daybreak* of the tale of a "Danish king […] wrought up to such a degree of warlike fury by the music of his minstrel" that he killed five courtiers, despite there being "no war, no enemy, if anything the opposite". The explanation proffered is that "the force which *infers from the feeling to the cause* was strong enough to overwhelm appearance and reason". The "state of feeling into which music transports us", Nietzsche goes on, "almost always contradicts the appearance of our real situation and the reasoning powers which recognize this real situation and its causes" (D §142).

There are several passages, however, of particular value to our attempt to reconstruct Nietzsche's view. In a number of published sections, entitled "*The logic of the dream*" (HH "First and Last", §13), "*Experience and invention*" (D §119), and "*The error of imaginary causes*" (TI "Errors", §4), as well as in a late *Nachlaß* passage (eKGWB 1888: 15 [90]), Nietzsche discusses the idea that dreams are interpretations of states of stimulation while asleep. The psychological picture presented in these passages changes very little across the period in which they were written, and we can draw from these discussions four key elements: (a) that our nervous system is stimulated while asleep; (b) that this is interpreted—which will involve the mind seeking, variously, 'grounds', 'reasons', 'causes', or 'motivations' (employed interchangeably here) for one's condition; (c) that this process of interpretation involves (by no means necessarily conscious) recourse to memories regarding similar states and the interpretations that were employed; and (d) that cause and effect have (in some sense) been reversed. Nietzsche is clear in these passages that it is not merely exteroceptive sensory information with which he is concerned. As he wrote of seeking a reason for one's "malaise" in BA, the dream passages discuss positing a cause for finding oneself "feeling bad" (eKGWB 1888: 15 [90]; TI "Errors", §4) or "feeling good" (TI "Errors", §4). Interoceptive sensation, on Nietzsche's view, is processed in a similar manner to the exteroceptive sensation reconstructed in episodes of dreaming.

It should come as no surprise then that the process described in the dream passages bears undeniable resemblance to the picture of affects offered in BA. If we return to that passage, features (a) and (b) are clearly present, and we find something related to (c) as well.[19] Nietzsche claims that "persons and things" rouse a physiological state, as the sounds, etc., stimulate the senses in the dream scenarios—(a). The sensations of one's bodily perturbation are then interpreted, as part of which a "*reason* is sought" (eKGWB 1883: 24 [20]) for one's state, as it is in the dream case—(b). Regarding (c), if we look to TI "Errors", §4, Nietzsche claims that memory often "swings into action in such cases without our awareness", and "calls up earlier states of a similar kind and the causal interpretations which have grown out of them".[20] He writes, in particular, of a "*habituation* [G e w ö h n u n g] to a certain cause-interpretation" (TI "Errors", §4). This is not unrelated to his use of *Gewöhnung* in BA, regarding the concomitance of incidents and feelings over lengthy periods of time. Experienced co-occurrence produces the first interpretations positing a causal link between the event or person and general feeling. Once habituated, these explanations can be retrieved from memory almost immediately (without one's being aware of any searching going on) and utilized to interpret subsequent similar feelings, and/or comparable concomitances of event and bodily change.

The remarks in BA elaborate upon the relationship between interpretations and memory. Nietzsche claims that "certain occurrences and general feelings" are "so regularly connected" that the mere sight of the event produces the state of general feeling—"specifically" the physiological features it "brings with it" (eKGWB 1883: 24 [20]). This claim resembles a fragment composed a few months earlier, in which Nietzsche writes that "Anger (and all affects), *first* a state of the body: which is interpreted". "Subsequently", he goes on, "the

---

[19] Despite no explicit reference to (d) in BA, it seems of greater relevance to Nietzsche's affect-psychology (and to BA) than this omission suggests. The gist of (d) is plausibly that interpretation usually occurs so rapidly that awareness of the putative reason for one's state is almost immediate, despite being (on Nietzsche's view) the result (or 'effect') of a complex psychological process. (We will address this process later in the paper.)

[20] Cf. from the *Nachlaß*: "When we enter into a certain physiological state, what we thought the last few times we were in that state comes to mind. There must be a release [*Auslösung*] in the brain for each state" (eKGWB 1880: 1 [115]; cf. 1881: 11 [301]).

interpretation produces the state freely" (eKGWB 1883: 9 [44]). It might appear, at first blush, that these claims—that the state of the body *precedes* interpretation, and that interpretation *produces* the state—stand in tension. When taken in conjunction with BA, however, a clearer picture emerges. Given the propinquity of the two fragments, they plausibly express a similar idea: once we have become habituated to an interpretation, this does not just result in memory deploying it automatically when we experience a familiar concomitance of putative cause and particular bodily response. In time, perceiving—or perhaps even just imagining or remembering—the interpreted, posited cause of one's previous affective states can elicit the same condition.[21]

The psychological process we have drawn out above clearly comports with the *S*, *P*, *A* order suggested by the somatic view. Nevertheless, certain facets of Nietzsche's position remain unclear. In both BA and the dream passages, feature (a) seems reasonably straightforward, and the remarks on memory and habituation are compatible and complementary, providing a fuller picture of the relation between the two, as sketched in (c). The challenge is to make sense of Nietzsche's comments regarding (b). The claim that one's state is interpreted and that this requires the positing of a cause is intriguing. I suggested earlier that an occurrent affect, for Nietzsche, comprises one's bodily state and the subsequent modulation of specific psychological processes. If we are to understand interpretation, and the seeking of a cause in particular, we must look more closely at this modulation.

2.2

The sensations of one's bodily condition trigger the search for a cause. As Nietzsche writes, in *Twilight of the Idols*, "Most of our general feelings [*Allgemeingefühle*]—every sort of restraint, pressure, tension, explosion in the play and counter-play of the organs, the condition of the

[21] Cf. Nietzsche's remark that affects "are symptoms of the formation of memory material" (eKGWB 1884: 25 [514]). Nietzsche's use of 'memory', in these passages, seems closely related to notions like 'association' and 'proceduralization'. (Thanks to Gudrun von Tevenar for raising this suggestion.)

nervus sympathicus[22] in particular—excite our cause-drive" (TI "Errors", §4; cf. "Errors", §5).[23] If we look to his remarks concerning the causes sought, Nietzsche intimates that there is something problematic about them. He writes in the *Nachlaß* of the "almost 'unconscious' distress of the viscera, the tension of the blood pressure in the abdomen, the unhealthy states of the *nervus sympathicus*" as all requiring interpretation. In those cases where one must endure "pain whose physical cause is unknown", he goes on, we place "a *false reason* behind the actual malaises of the body" (eKGWB 1885: 38 [1]). Elsewhere, he claims that habituation to certain interpretations "obstructs and even prohibits an *investigation* of the cause" (TI "Errors", §4), and that the cause "sought and imagined" is "pitiably inadequate as the real cause" (eKGWB 1888: 15 [90]).

But why, one might ask, are the putative causes (in some sense) inadequate? A line in the *Genealogy* offers a starting point. Nietzsche writes of the vengeful, suffering individual that the "*true* cause of their feeling bad, *the physiological one*, remains concealed" from them (GM III, §15, emphasis added). The true, "physiological" cause, I take it, is the sort he discusses elsewhere in regard to Luigi Cornaro. "*Cornarism*", as Nietzsche dubs it (TI "Errors", §2), involves "*confusing the consequence with the cause*" (TI "Errors", §1).[24] What Nietzsche has in mind is a *causal-explanatory* mistake: positing *X* as the cause of *Y* (and sufficient to explain *Y*) when this relation in fact fails to obtain, and both are explained by some further, deeper fact. Cornaro prescribes certain dietary habits as the *cause* of a long and healthy life, when they are in fact a *consequence* of those psycho-physiological features that Nietzsche takes to be genuinely explanatory of Cornaro's longevity.

---

[22] The term *nervus sympathicus* refers to the sympathetic nervous system.

[23] It is unclear whether Nietzsche is positing a distinct drive here. His claim in D §119 that drives "interpret nervous stimuli and, according to their requirements, posit their 'causes'" leaves open the possibility that by "cause-drive" he is referring simply to the drive (or drives) presently interpreting sensation.

[24] One should note that this is distinct from feature (d) in the dream passages (the claim that cause and effect had, in some sense, been reversed). Unlike Cornarism, (d) concerns the order of *experience*.

Nietzsche approaches affects in a similar fashion. He writes in a discussion of *ressentiment* and weakness in *Ecce Homo* that one "does not know how to get rid of anything, one does not know how to get over anything, one does not know how to repel anything—everything hurts. People and things obtrude too closely; experiences strike one too deeply; memory becomes a festering wound. Sickness itself *is* a kind of *ressentiment*" (EHα "Wise", §6). The individual suffers an enduring psycho-physiological condition that disposes them to certain intense reactive episodes and protracted moods, as well as to patterns of thought and judgement.[25] This antecedent psycho-physiology is the 'further, deeper fact' to be properly understood as the cause of particular affective reactions.[26] An object might elicit a response from one person but not another, a different response in each person, or different responses at different times in the same people. As in the case of Cornaro's diet, Nietzsche holds that the deeper considerations explain both *X* (the extracted object) and *Y* (the affective response): one's psycho-physiology is explanatorily primary regarding *why* the object in question was perceived *qua* stimulus, and why one had *that* response to it.[27]

Nevertheless, one might still ask how we are to square the idea that our thinking about affects involves a causal-explanatory error with Nietzsche's claim in BA that "persons and things" rouse

---

[25] It is important to distinguish between *ressentiment* as a brief, intense affect, and *ressentiment* as an enduring affective condition. When one is inhibited or frustrated in the discharging of 'released' states (including reactive affects triggered by an earlier frustration), this can result in an episode of vengeful resentment—*ressentiment*. Nietzsche imagines noble individuals undergoing this occasionally, but without its becoming pervasive, ingrained, or pathological. In the noble, *ressentiment* "runs its course and exhausts itself in an immediate reaction". In the weak, however—those incapable of responding, for one reason or another—continual frustration turns instances of *ressentiment* into an extensive and complex psychological disorder. It is this group that Nietzsche refers to as the "human beings of *ressentiment*" (GM I, §10). (Thanks to Ken Gemes for highlighting this distinction.)

[26] Cf. from the *Nachlaß*: a "person, a people, have suffered a physiological change, experiences this in *general feeling*, and interprets this in the language of their affects and *according to the degree* of their knowledge, without realizing that their physis is the seat of the change" (eKGWB 1881: 11 [103]).

[27] Cf. Schopenhauer, who writes that "motives never determine more than what I will at *this* time, in *this* place, and under *these* circumstances". They determine the will's "manifestation at a given point in time: they are merely the occasion [*Anlaß*] on which my will shows itself" (ZA I: 150/WWR I: 106). (Citations to the volume and page number of the complete works, edited by Arthur Hübscher [ZA], followed by the volume and page number of the English edition listed in the bibliography. Quotations are from the translation, with occasional amendments.) In a similar vein, Nietzsche writes in the *Nachlaß* that we speak of "the causes of affects, and mean their occasions [*Gelegenheiten*]" (eKGWB 1882: 3 [1], aphorism 408).

affects?[28] He does, after all, talk of interpretations eventually producing the state "freely" (eKGWB 1883: 9 [44]), and of the co-occurrence of events and feelings being so regular that the mere sight of the event elicits the physiological state (eKGWB 1883: 24 [20]). Given such remarks, Nietzsche's point in talking of "inadequate" reasons (eKGWB 1888: 15 [90]) clearly cannot be that the relevant objects fail to provoke affective responses. Instead, his claim seems to be that such things are more properly thought of as *stimuli*, in contrast to the psycho-physiological *causes* of both one's affective episode and one's having been provoked by the thing in question.[29] The causal explanations Nietzsche deems inadequate are those that treat the occasioning element as a direct mental cause,[30] sufficient to explain our reaction without reference to the individual's psycho-physiology (some features of which might well relate to their experiences, and habituation to various associations and causal explanations).

Instead of seeking causes of the explanatorily primary, 'physiological' kind in instances of perturbation, Nietzsche claims that we extract something to set as the reason for our state from the "surroundings" of our "disagreeable experiences and fears" (eKGWB 1885: 38 [1]). Later remarks suggest he understood "surroundings" sufficiently broadly as to include mental items such as thoughts, and so on (see TI "Errors", §4; §5; §6). If we look to his remarks on the function of affective states, it becomes clearer why the cause is plucked from one's environment. Affects, Nietzsche suggests, developed as a means of producing useful, beneficial behaviours (eKGWB 1881: 15 [69]; 1881: 11 [241]). The psychological mechanisms set into motion by

---

[28] Thanks to a reviewer for the *Journal of the History of Philosophy* for pressing me on this.

[29] Nietzsche writes explicitly in the *Nachlaß* that we should speak not of the "causes of willing, but the stimuli [*Reizen*] of willing" (eKGWB 1884: 25 [436]). A similar picture emerges in the published works in a discussion of the causes of action in *The Gay Science*. Nietzsche distinguishes two "*kinds of cause that are often confused*". The first is the "cause of acting", which he identifies as a "quantum of damned-up energy waiting to be used somehow". The second, the "cause of acting in a certain way, in a certain direction, with a certain goal" (including "so-called 'purposes' [and] 'vocations'" [cf. n. 7]), is "something quite insignificant", a "small accident in accordance with which this quantum 'releases' [*auslöst*] itself in one particular way". The passage has deeply Mayer-esque overtones, with Nietzsche writing of the second kind of cause as the "match in relation to the powder-keg". The error which he is concerned is that of treating as "the *driving* force" that which is "only the *directing* force"—of understanding a stimulating and directing element as the "cause of acting" (GS §360).

[30] I borrow 'mental cause' from G. E. M. Anscombe (1957); cf. Kenny (1963: 71).

bodily perturbation are driven (at least in part) by what we could call 'practical-psychological necessity'—the need to discharge one's state of released arousal, through the "proper reaction, that of the deed" (GM I, §10).

For affects to perform the beneficial function Nietzsche envisages, the cause-extracting mechanism must identify the relevant feature(s) of the environment (e.g. possible threats).[31] This process represents a central part of what I shall label 'affective interpretation'. The sensations of one's bodily state (which trigger the cause-extracting mechanism) are interpreted as related to elements of one's surroundings. As we saw in our discussion of (c), this will draw upon the stock of aetiologies stored in memory, which have been forged through experience. Once habituated, interpretations are employed in new situations (often automatically) to identify relevant features, with the extracted feature providing the intentional object for the affect (and perhaps further reactive responses) in relation to which one's behaviour can be modified, and subsequent actions directed.[32]

Although the referring of sensation back to putative cause is the feature of affect-psychology upon which Nietzsche focuses most closely, his picture of affective interpretation is plausibly richer than just this. The modulation of psychological processes consequent upon sensations reaching the 'sensorium commune' is suffusive: "even the sense-perceptions and thoughts", he

---

[31] The generation of beneficial behaviour of course requires not only that the relevant feature(s) of the environment be identified, but also that an appropriate affect be roused (e.g. fear, in the case of a threat). Nietzsche is certainly sensitive to this issue. "Pleasure and displeasure", he insists, are "never 'original facts'": instead, he countenances a form of pre-conscious appraisal of "*overall usefulness*" and "*overall harmfulness*" that has "*will-reactions (affects)*" as its output (eKGWB 1887: 11 [71]). Nietzsche often links this process to valuations, judgements, and power. If we look to BA, for instance, beneath the paragraph I have quoted are two further, separate lines in which he writes, "In 'pleasure' and 'displeasure', *judgements* are already embedded: the stimuli [*Reize*] are discriminated according to whether or not they are conducive to the feeling of power" (eKGWB 1883: 24 [20]).

[32] Nietzsche was not familiar with the language of 'intentionality'. His use of 'cause' should not mislead one, however, into thinking that he was insensitive to the possibility of cause, stimulus, and object coming apart (indeed, quite the opposite is true). Something's being a 'cause' of the sort posited in affective states does not, of course, occlude its being an intentional object: I can say both that "I am angry *about* the football", and that "the football *made* me angry". Cf. Jesse Prinz's remarks about "emotional attitudes", which he claims establish "a causal link between an emotion and the representation of an object or a state of affairs". On Prinz's view, when certain causal links between representations and emotions obtain, "we say that the emotion has the content of those representations as its intentional object" (Prinz, 2005: 20).

writes, belong among the "revelations of affect" (eKGWB 1882: 4 [126]).[33] But what does this mean? One of the dream passages is instructive here. In D §119, Nietzsche discusses the role of drives in the interpretation of sensory information. He considers how the sensations of bodily processes, bed-clothes, sounds (and so on), which remain "very similar on one night as on another", can be interpreted so differently in dreams from night to night. His conjecture is that these "*very free*, very arbitrary interpretations" occur under the influence of drives: the "inventive reasoning faculty", he claims, posits "*causes* so very different" from one night to the next "for the same nerve-stimuli", since the "prompter of the reasoning faculty" was a different drive in each instance (D §119).

Despite waking life not having this "*freedom* of interpretation", Nietzsche asks whether he has to add that there is "no *essential* difference between waking and dreaming?". Our everyday experiences, he insists, are similarly altered by occurrent states. This interpreting will plausibly involve not just the positing of causes, but also the generation of perceptual saliences (see Katsafanas, 2013: 740). Indeed, Nietzsche's remarks suggest something akin to 'affordance'. The affordances of a particular environment are "what it *offers* the animal, what it *provides* or *furnishes*, either for good or ill" (Gibson, 1979: 127; cf. 1977). Nietzsche writes in D §119 that a drive which "desires gratification—or the exercise of its strength, the discharge of this strength, the saturation of a void [...] regards every event of the day with a view to seeing how it can employ it for the attainment of its purpose". An interpreting drive highlights opportunities and interactions presented by the environment. As a result, objects, events, and individuals are ascribed a function and meaning from the perspective of the drive.[34]

---

[33] Cf. Schopenhauer: "let us picture to ourselves how much every affect or passion clouds and falsifies knowledge [...] every inclination or aversion twists, colours, and distorts not merely the judgement, but even the original intuition of things" (ZA 4: 442/WWR II: 373).

[34] See eKGWB 1885: 1 [58], where Nietzsche writes, "From each of our fundamental drives there is a different perspectival valuation of all events and experiences".

The relation between drives and affects in Nietzsche's psychology is somewhat opaque.[35] He clearly sees an intimate link between the two, and ascribes many of the same roles and capacities to both,[36] including interpreting. Indeed, he insists that one "may not ask: '*who* interprets?'—rather the interpreting itself [...] has existence (but not as a 'being', rather as a *process*, a *becoming*) as an affect" (eKGWB 1885: 2 [151]). If we consider that Nietzsche envisages affects as a means of generating beneficial behaviours, and the connection he posits between affects and evaluatively-laden perception and judgement,[37] it is overwhelmingly likely that affective interpreting bears considerable resemblance to the process described in D §119. Nietzsche writes of "sense-perceptions" as among the "revelations of affect" (eKGWB 1882: 4 [126]), I propose, as he takes the sensations of bodily arousal to both prompt the cause-extracting mechanism and modulate the interpretation of exteroceptive sensation. Under the pressure of practical-psychological necessity, affective interpretation identifies the putative cause, but also highlights features of the environment that are (directly or indirectly) conducive to the performance of the deeds that will discharge one's state. This perspectival interpreting is

---

[35] Although there is not room here for a full discussion of drives (or their relation to affects), I can offer a few remarks. It is not abundantly clear that Nietzsche's references to drives present a picture overly worked-out in its details. Drives seem to be dispositions that are individuated by a 'characteristic activity' towards which they incline the organism; an activity that proved beneficial to the organism's forebears, resulting in the drive being inherited (broadly construed). His comments in several places suggest that drives are forms of 'multi-track' affective disposition (see Julien A. Deonna and Fabrice Teroni, 2012: 7)—specifically, ones that incline the organism, either directly or indirectly, towards a characteristic, individuating activity. Nietzsche writes, for instance, of drives as the "*after-effects of long-cherished valuations* which now work instinctively, as if a *system* of pleasure- and pain-judgements. First coercion, then habituation, then need, then natural inclination (drive)" (eKGWB 1884: 25 [460]). Although I doubt that any single reading will account for *every* reference to drives in Nietzsche's writings, something along these lines would capture an awful lot of what he says, and would neatly explain the overlap between drives and affects. Cf. Paul Katsafanas' suggestion that drives are dispositions that "induce affective orientations" (Katsafanas, 2013: 740). (Thanks to a reviewer for the *Journal of the History of Philosophy* for raising this issue.)

[36] See, for example, D §109 and BGE §117 on the overcoming of a drive/affect, or GS §111 and eKGWB 1885: 1 [75] on thinking and the power-struggle of drives/affects.

[37] Nietzsche writes famously of moralities as a "sign-language of the affects" (BGE §187). If we look to the *Genealogy*, he talks of the "powerful" being "re-coloured" when seen "through the poisonous eyes of *ressentiment*" (GM I, §11). Such moral valuations are "an *interpretation*, a kind of interpreting" undertaken by the affects. Nietzsche describes the interpretation itself as a "*symptom* of certain physiological states, as well as of a certain spiritual level of judgements" (eKGWB 1885: 2 [190]; cf. 1880: 6 [445]).

evaluative, with elements of one's surroundings valenced positively or negatively as objects of inclination or aversion, attraction or repulsion.

This adds flesh to the claim that an occurrent affect, *A*, for Nietzsche, involves the modulation of specific psychological processes. As one might expect given his ascription of affects to a wide range of species, there is little in the above-described affective process that requires those cognitive skills we tend to attribute only to human beings. This is not say, however, that Nietzsche overlooks the elements of affective experience that are distinctly human. Although affects (as a psychological kind) are putatively present across a great many species, the capacity to reflect consciously upon one's state is taken by Nietzsche to be unique to human beings. We do not just form affective interpretations connecting our state and surroundings and valencing the contents of our experiences, but can also bring this connection (as well as our states and their contents) to a point of reflective awareness. Humans produce linguistically articulated aetiologies for their conditions and subsume states under affect-words. This psychology of self-ascription is an integral part of Nietzsche's writings on affects,[38] and it is this form of interpretation, which I shall call 'linguistic interpretation', that we shall now examine.

2.3

In the particular cases Nietzsche discusses, the process of interpreting one's occurrent state as this-or-that affect starts with bodily perturbation experienced as general feeling. He writes, for instance, of "feeling one way or another", and the "feeling of suffocation" (eKGWB 1883: 24 [20]). As with affective interpretation, the search for a cause is integral to linguistic interpretation. It "never satisfies us", Nietzsche writes, "simply to establish the mere fact *that* we find ourselves this way or that: we only admit this fact—become *conscious* of it—*when* we have given some kind of motivation for it" (TI "Errors", §4; cf. eKGWB 1888: 15 [90]). The point here is not that we become, say, phenomenally aware of affects only once we know what

[38] It is important, here, to distinguish 'self-ascription' from 'individuation'. Nietzsche is deeply interested in the former, but rarely addresses the latter (to my knowledge).

caused them. Nietzsche's remark plausibly expresses the same idea as his claim in BA that a "*reason*" is sought for the sake of the "*conceivability*" of our state (eKGWB 1883: 24 [20]): namely, I propose, that we become reflectively conscious of our state *as* a particular state only once a cause is posited.[39]

If correct, this rules out the possibility that Nietzsche understood the linguistic interpretation of particular general feelings as comprising identifying oneself as, say, angry, and then seeking a putative cause-object for one's anger. Such a reading would clearly involve becoming 'conscious' of one's state prior to positing the required cause. Instead, Nietzsche's position appears to be that those things to which we can attend introspectively underdetermine the self-ascription of an affect. He talks in BA of a "feeling of suffocation", and in the dream passages of someone finding themselves "feeling good" or "feeling bad" (TI "Errors", §4; cf. "Errors", §6; eKGWB 1888: 15 [90]). These are presented simply as instances of arousal with positive or negative valence. This is not to say that the sensitivity to, or recognition of, general feelings plays *no* role (or even a minor role) in the identification of affects; just that identifying a state as one for which affect-words are usually deployed requires more.

At the broadest level—that of interpreting one's condition *as an affect*—Nietzsche's remarks suggest that the minimum requirement is to refer one's state back to a certain *sort* of cause. On multiple occasions, he contrasts the "right class" of causes, to be identified by the "anatomically educated", with the causes sought by those less well-versed. The latter group, we are told, seek "a moral explanation" for their state (eKGWB 1884: 26 [92]; cf. 1885: 38 [1]; GM III, §17). Such remarks leave open the possibility that an individual sufficiently habituated to this class of explanations could (in principle) be in a position to interpret themselves as experiencing an as-

[39] I am persuaded by Mattia Riccardi's contention that Nietzsche's writings evince a 'pluralism' about consciousness (Riccardi, 2015). For our purposes, we need only distinguish phenomenal from reflective consciousness.

yet-unspecified, valenced affect, on the basis of believing the state to have been elicited by some as-yet-unidentified cause from the "psychological-moral" domain (GM III, §17).

But what of interpreting one's state as this-or-that affect (with which Nietzsche seems to be more concerned)? This will plausibly draw on those 'symptoms' Nietzsche counts among the "revelations of affect", including conscious thoughts and sense-perceptions (see eKGWB 1882: 4 [126]), and conscious feelings. His remarks in BA regarding the conjunction of "certain occurrences and general feelings" (eKGWB 1883: 24 [20]) suggest a concomitance of certain events and *particular* states—broad, *sfumato* categories of (somewhat imprecisely) recognizable general feeling. This permits the possibility that the recognition of general feeling, once paired with an appropriate cause, could provide a means through which to subsume occurrent states under affect-words.[40] This might be limited to rather coarse-grained, basic states, however, given Nietzsche's somewhat dim view of our introspective faculties. Since even simple folk-psychologies will involve complex, fine-grained affective states, this does not get us terribly far towards understanding how Nietzsche envisages the subsumption of affects under particular concepts.

If we attend to his remarks on guilt and sinfulness in the late works, a fuller picture emerges that sheds light upon the identification of states as fine-grained affects. In the *Genealogy*, Nietzsche writes of guilt as "a piece of animal psychology, no more" (GM III, §20). This likely refers to its putative origin in the process of socialization. The question of how behaviours are modified and instincts curbed under conditions of social existence was of enduring interest to Nietzsche, and elicited some subtle and insightful reflections. Social restrictions and their

---

[40] But what of moods, which are often claimed to be objectless? On my reading, Nietzsche would be committed to saying that states for which no 'psychological-moral' cause-object (including an as-yet-unspecified, hypothetical one) had been posited, could only be described in terms of rather general bodily sensations—e.g. the "feeling of suffocation" (eKGWB 1883: 24 [20]). Where particular moods are identified, Nietzsche would have to claim that an appropriate cause-object was sought and found. Suggestions made by representationalists (among others), such as that the world experienced might be the object (see Tim Crane, 1998; cf. Solomon, 1976: 172–73) or that moods involve a succession of changing objects (see Michael Tye, 2008), could also be comfortably accommodated. (To be clear, Nietzsche's view would not entail that all moods must have an object, only that *self-ascribing* particular moods would require one.)

corresponding punishments produce a range of complex psychological responses, including (perhaps most importantly) 'bad conscience'. Nietzsche writes of drives being accompanied by "either a good or a bad conscience". These complex, compound affective episodes involve the drives co-occurring with an "attendant sensation of pleasure or displeasure", acquired as "second nature" in relation to drives "already baptized good or evil", or noted as "a characteristic of beings already morally determined and evaluated" (D §38).[41]

The discussion of bad conscience in the *Genealogy* paints a subtly distinct psychological picture, placing greater emphasis on the phenomenon of aggressive self-punishment. Nietzsche writes of the "*internalizing* of man", where those instincts that "do not discharge themselves outwardly *turn themselves inwards*" (GM II, §16). In cases of frustration and/or restraint, one might not just experience a depressive affect accompanying a particular motivation, thought, or feeling, but also a *reactive* affect. If we recall the practical-psychological necessity under which the cause-extracting mechanism operates, this pressure is most intense in the strongest states, and especially in the reactive affects that so interested Nietzsche. These demand both a cause-object in relation to which deeds can be directed, and the means for one to perform those acts that would alleviate their condition. The subjugated peoples in the *Genealogy*, saddled with the burden of curbing instinctive behaviour, are presented as lacking both an obvious cause for their protracted condition, and the means to discharge their particular affects through suitable deeds. In such circumstances, almost any putative aetiology will do to facilitate the relieving of one's state.[42] Constrained by the confines of mores and custom, the disposition towards cruelty with whose satisfaction human history is perfused (according to Nietzsche) finds an outlet in punishing oneself.

---

[41] Elsewhere, in *Twilight of the Idols*, Nietzsche writes of the "liveliest drives" of the "criminal type", which "quickly grow together with the depressive affects" in circumstances where their "*virtues* have been ostracized by society" (TI "Expeditions", §45).

[42] In TI "Errors", §5, Nietzsche offers as a "First [psychological] principle" that "any explanation is better than none". Elsewhere, he speaks of sufferers seeking something on which to discharge their affect, "on some pretext or other" (GM III, §15; cf. HH "Man in Society", §370).

It is this complex "animal psychology" that is exploited by the priest through the introduction of the notion of 'sin' (GM III, §20). Nietzsche's thoughts grow from the affect-psychology hitherto discussed. Recall the *Nachlaß* note in which he writes of placing a "*false reason*" behind the "actual malaises of the body" (eKGWB 1885: 38 [1]). It is there, and in an earlier draft from the previous year, that Nietzsche talks of the "right class" of causes. His remarks in these passages (the latter of which was written at most 18 months or so after BA) bear considerable resemblance to BA. Nietzsche identifies specific states of physiological perturbation and suggests that the sensations of these states are interpreted. Those experiencing "uncertain feelings of displeasure" (eKGWB 1885: 38 [1]), as we saw earlier, seek a moral or psychological explanation. In the two passages that mention a "right class" of causes, however, a further claim is added. Nietzsche writes that, in regard to the "pain whose physical cause is unknown", the "tortured individual questions himself in such a lengthy and inquisitorial manner, until he *finds* himself or others *guilty*" (eKGWB 1885: 38 [1]; cf. 1884: 26 [92]).

By the time we reach the *Genealogy*, these thoughts are in full swing. Nietzsche talks of a "*feeling of physiological inhibition*", which, "from a lack of physiological knowledge, does not enter into consciousness as such, so that its 'cause', its remedy, can only be sought and attempted in the psychological-moral realm" (GM III, §17). Those experiencing tremendous suffering might be armed only with simple aetiologies for their particular pains and frustrations, ingrained through the concomitance of event and state. Their enduring psycho-physiological condition, however, remains unassuaged. Man, Nietzsche writes, suffers "from himself in some way or other, physiologically in any case […] uncertain why, to what end?" and is "desirous of reasons" since they "alleviate" (GM III, §20). The afflicted individual "instinctively seeks a cause for his suffering; still more precisely, a perpetrator, still more specifically, a *guilty* perpetrator" upon whom "he can discharge his affects" (GM III, §15; cf. TI "Expeditions", §34).[43] It is through

[43] This remark is related to the picture of bodily changes etc. as *Auslösungen*. Nietzsche writes of a "*Malaise* as *impeded* release [Verstimmung *als* verhinderte *Auslösung*]", claiming that "it was not the releases

facilitating this discharge of affects, which temporarily anaesthetizes the sufferer to their pain, that reasons "alleviate".

The priests of Nietzsche's *Genealogy* volunteer aetiologies that offer both object and means for the afflicted individual's reactive affects. Their message, in such abject circumstances, provides "the *first* hint concerning the 'cause'" of the sufferer's pain (GM III, §20). The priest responds to the psychological needs of the afflicted by manipulating the psychology of bad conscience, directing the frustrated drives and (most importantly) the reactive affects and *ressentiment* of the slaves back onto themselves. This is done through the provision of a complex aetiology in which suffering is *caused* by sin.[44] The concept of 'sin' is embedded in a web of broader metaphysical, moral, and psychological views, involving notions such as *causa sui* free will, claims about the nature of moral norms, and so on. Within such a framework, habituation to this interpretation facilitates the identification of fine-grained affective states. Although the object of one's state—the target of reactive deeds and attitudes—might well remain a particular individual, group, or event, etc., the putative cause is more specific. With the introduction of 'sin', one's feeling bad, previously interpreted as caused by some*one* or some *event*, is identified as moralized guilt or 'sinfulness', or moral resentment, caused by an *aspect*, namely the sin of oneself or another. This picture, it seems to me, provides the best illustration of how Nietzsche envisages the positing of (fine-grained) causes facilitating the identification of complex affects, and suggests a model that could be extended to cover the wide range of affects employed in his moral psychology.[45]

[*Auslösungen*], however violent they might have been, that did humanity the most harm, but rather the prevention of them" (eKGWB 1881: 11 [28]).

[44] Nietzsche writes that the "*entire realm of morality and religion falls under* [the] *concept of imaginary causes*". The "'Explanation'" offered for "*pleasant* general feelings [*Allgemeingefühle*]," he claims, is that they are consequences of things such as "faith, love, hope," or "the consciousness of good deeds . . . the so-called 'good conscience'" (which Nietzsche identifies as "a physiological state"), and so on. One's "*unpleasant* general feelings," by contrast, are explained by reference to (among other things) "actions we cannot approve of," our deserving to suffer, and the feeling of sin "foisted on physiological discomfort" (TI "Errors," §6; cf. D §83; A §15).

[45] This bears some resemblance to Spinoza's treatment of affects. Spinoza analyzes a wide range of states in terms of three primary affects (desire, joy/pleasure, and sadness/pain), and their relation to particular causes. Love, for instance, is "titillation accompanied by the notion of an external cause" (Spinoza, *Eth.*, IV,

3.

This leaves us with a richer picture of Nietzsche's somatic and interpretive views, including both affective and linguistic interpretation. Nevertheless, our work is not quite done. The focus in this section is A_F and A_R—the points at which one becomes aware of general feeling, and at which one can represent oneself at the level of reflection as being in a particular state, respectively. Attention to Nietzsche's thoughts on affects becoming conscious, I shall argue, affords us a means of making sense of the last remaining feature of BA to be understood—namely, the 'conceivability' claim. In addition, the complete picture of A_F and A_R suggests points of divergence from the extant view we outlined in §1, especially in regard to non-conscious affects.

I claimed earlier that Nietzsche's approach to psychological terminology was naturalistic and might issue correctives to common-sense conceptions of given states. The view we have drawn out so far presents, for various reasons, a somewhat revisionary picture. Nietzsche's complaints, however, are not restricted simply to the idea that we have overlooked or misunderstood the involvement of the body in affective states, or that we are mistaken about our access to our own mental states. He refers to affects as "constructions", "errors", or more commonly, "interpretations", of the intellect (see eKGWB 1883: 24 [20]; 1883: 24 [21]; 1881: 11 [128]). The point in such remarks is not that there are *no* introspectively salient psycho-physiological episodes that we take to be instantiations of states corresponding to affect-words. Instead, the thought is that the carvings-up under which we subsume varieties of mental experience are generated under certain pressures, and (at best) reflect significant limitations.

---

proposition 44, *proof*—quoted by Schopenhauer in "The Metaphysics of Sexual Love", ZA 4: 623/WWR II: 533). (Citations to Spinoza's *Ethics* refer to part [denoted by Roman numerals], proposition [Arabic numerals], and sub-section.) Nietzsche attributes to Spinoza the view that the *conscientiae morsus* is "a sadness, accompanied by the image of a past matter that has turned out in a manner contrary to all expectation" (GM II, §15; cf. Spinoza, *Eth.*, III, proposition 18, *Scholium 2*).

Nietzsche highlights a number of factors that purportedly contribute to psychological terminology, and our application of it, falling short of carving nature at the joints. As I suggested in §1, one thought underlying his insistence on the inadequacy of our conceptions of certain phenomena is that we derive these conceptions solely from the contents of consciousness.[46] Several distinct errors are purportedly resultant upon this shortcoming. The first I highlighted was that of mistakenly concluding that psychological phenomena consist simply in their 'symptoms', ignoring the non-conscious components of the process. A second concerned a failure to appreciate the putative ubiquity of certain phenomena. If we look to *Daybreak*, Nietzsche writes therein that "words really exist only for *superlative* degrees of [...] processes and drives": we "*are none of us*", he claims, "that which we appear to be in accordance with the states for which alone we have consciousness and words". Those "cruder outbursts of which alone we are aware make us *misunderstand* ourselves; we draw a conclusion on the basis of data in which the exceptions outweigh the rule" (D §115). The claim is that we misconceive of the states not just because we attend only to fragments of a broader process, but also because we are sensitive only to exceptional episodes.

The instances of which we become aware,[47] and which we identify as "Anger, hatred, love, pity, desire [...] joy, pain", are in fact "*extreme* states". Nietzsche writes that those "milder, middle degrees [of "processes and drives"], not to speak of the lower degrees that are continuously in play, elude us" (D §115). In contrast to the folk-psychological picture on which affects are presented as occasional, irruptive conscious states, Nietzsche maintains that they are near-constant, pervasive features of mental life. (This is not to say that *all* of one's affects are

---

[46] Nietzsche raises three further factors: i) that we carve up conscious phenomena "not *finely*, but completely coarsely" (eKGWB 1880: 6 [444]; cf. 1881: 11 [128]); ii) that this shortcoming is exacerbated by a function of psychological terminology—namely that of communicating with others—which Nietzsche takes to require coarse-grained concepts under which subtly distinct states can be subsumed; and iii) that we are not impartial theorizers regarding the mind: our folk-psychology is, in some cases, a product of motivated interpretation, and reflects moral pressures (see GM I, §13; eKGWB 1880: 6 [444]).

[47] Nietzsche's rhetorical question, a few sections later, as to whether he really need query that our "so-called consciousness is a more or less fantastic commentary on an unknown, perhaps unknowable, but *felt* text?" (D §119, emphasis added) suggests the sort of awareness he has in mind here.

constant, but that one is constantly experiencing at least one occurrent affect.) Nietzsche's alternative picture is of a struggle between affects, the constituents of which might vary between given times and fluctuate in strength from one moment to another.[48] We are encouraged to see affects not as infrequent occurrences that interrupt, say, thinking (often deleteriously), but rather as integral, permeative psycho-physiological processes, ebbing and flowing in stronger and weaker manifestations in response to stimuli.[49] Hence affective interpreting being described as "a *process*, a *becoming*" (eKGWB 1885: 2 [151]), in contrast to the picture of affects as strict, discrete states (or 'beings', in Nietzsche's terminology).[50]

The significance of its being "only for *superlative* degrees [...] of processes and drives" (D §115) that we have developed recognitional affect-words is owing to the intimate link Nietzsche posits between reflective consciousness and language.[51] If we look to *The Gay Science*, he writes that it is conscious thinking and *only* conscious thinking that "*takes place in words, that is, in communication-signs*" (GS §354; cf. eKGWB 1886: 5 [22]). The folk-psychological category- and concept- words under which states and processes are subsumed allow us to represent mental phenomena (such as superlative affective episodes) at the level of reflection. These

---

[48] The remarks on milder states resemble what Antonio Damasio calls "background feelings", which are "neither too positive nor too negative", and correspond to the "body state prevailing *between* emotions" (i.e. between the "*superlative*" instances mentioned in D §115). Background feeling "can be perceived as mostly pleasant or unpleasant" and, when enduring, "probably contributes to a mood" (Damasio, 1994: 150–51).

[49] On Nietzsche's view, the near-constant 'struggle' of drives and affects is partly constitutive of thinking (see GS §111; BGE §36; eKGWB 1885: 1 [61]; 1885: 1 [75]; 1887: 11 [310]). For discussion of Nietzsche's views on thinking, see paper two in this thesis.

[50] Nietzsche does, of course, use the word *Zustand* (state or condition) quite often in regard to affects. I take this to be reflective more of convenience than of Nietzsche's thinking the concept <STATE> to be wholly satisfactory in this context.

[51] See GS §354: "the development of language and the development of consciousness (*not* of reason, but rather only of reason's becoming conscious [*des Sich-bewusst-werdens der Vernunft*]) go hand in hand".

'signs', once employed in natural language sentences, can be taken as the object of further mental attitudes and processes.[52] Once this occurs, the state in question reaches point A_R.

This picture, I contend, underpins Nietzsche's insistence that the 'conceivability' of our state depends upon the provision of a reason for our feeling good or bad. As I suggested earlier, Nietzsche appears to be getting at the same thought when he writes in *Twilight of the Idols* that we become "*conscious*" of the state only once a "motivation" has been given (TI "Errors", §4). The idea, I submit, is that a state is 'conceivable' if one can think consciously about it.[53] Conscious thoughts, however, require content articulated *in words*. To represent an occurrent episode in natural language constructions of the sort, "I am in [state x]", "I feel [x]", and so on, linguistic interpretation must occur. As we have seen, subsuming one's state under an existing affect-word involves more than simply attending to one's state of feeling: it involves the positing of a cause for one's state. The ability to think consciously about one's state as this-or-that affect relies upon its having been interpreted *as* a particular affect, which in turn rests upon seeking, and finding, a cause. Once one can represent an affect in conscious thought, the state can reach point A_R and become conscious in the sense Nietzsche intends in TI "Errors", §4. It is, accordingly, not just practical-psychological necessity that drives the cause-extracting mechanism, but also 'cognitive-psychological necessity'—the need to be able to bring one's state to point A_R.[54]

---

[52] Nietzsche claims in the *Nachlaß* that "we summarize the capacity *to will*, *to feel*, *and to think* something of our willing, feeling, thinking *itself* with the word 'consciousness'" (NF 1885: 34 [87]; cf. 1884: 26 [49]). For discussion, see paper two in this thesis.

[53] The German word in BA is *Denkbarkeit*, which might be translated more literally as 'thinkability'. Although I do not wish to put too much weight on a single word from an unpublished passage (particularly since I take the same thought to appear in *Twilight of the Idols* without reference to 'conceivability'), Nietzsche's use of *Denkbar*, *Denkbarkeit*, etc., supports the above reading. The notion is employed in relation to conceptual carvings-up such as the "I" (eKGWB 1882: 4 [58]), or "affects, desires, willings etc." (eKGWB 1883: 24 [21]) that facilitate our thinking about given domains. (Thanks to a reviewer for the *Journal of the History of Philosophy* for encouraging me to be clearer here.)

[54] In addition to 'practical-' and 'cognitive-' psychological necessity, Nietzsche also claims that interpretation is driven by a desire to alleviate discomfort felt at being confronted with something seemingly unfamiliar, and

3.1

A corollary of this picture of A_F and A_R, on which we become conscious only of "*extreme*" instances and can represent in reflection only those states for which we have words, is that many affective episodes will be non-conscious (D §115). In this section, I want to expand on this, and to suggest that some elements of the extant view require amendment. As we saw in §1, affects, on Nietzsche's view, are commonly presented as phenomenally rich, typically conscious states, intimately related to feelings. While this arguably gets some things right, the picture we have drawn out diverges from the above characterization. As we have seen, affects will not be typically reflectively conscious, but appear in reflection only in exceptional cases. It is uncontroversial, however, that Nietzsche takes most of our mental life (affects included) to occur without our being reflectively conscious of it.[55] The claim of the extant view is instead plausibly that affects are usually experienced as rich, phenomenally conscious episodes, such that one *could* be reflectively conscious of them (though need not always be).[56] Yet, upon closer attention, Nietzsche's writings provide multiple ways in which affective episodes could be non-conscious, and depart even from extant view so construed.

If we look to D §115, talk of mildness and extremity makes clear that Nietzsche's focus is the intensity of the states in question. Since affects are, for Nietzsche, near-constant saturators of our mental lives, it would be hasty to conclude that he thought weaker affects to be phenomenally unconscious (i.e. lacking in phenomenal character). If low-level affects are near-constant, the phenomenal character of such states could be partly constitutive of the

---

that making sense of the experience in terms of something familiar has a palliative effect (see TI "Errors", §5; GS §355).

[55] From GS §354: "The whole of life would be possible without, so to speak, seeing itself in the mirror: as, in fact, even now the far greater part of this life still plays out without this mirroring."

[56] Welshon, for instance, is sensitive to the possibility of unconscious affects. He follows the claim that affective character is typically "*consciously experienced* as a particular feeling with a distinctive phenomenology" by denying "that Nietzsche's view commits him to saying that unconscious affect does not occur." As examples of states that "occurrently may be but need not be consciously experienced," Welshon offers "suppressed, repressed, dissociated, dampened or displaced" affect (Welshon, 2014: 127).

phenomenology of the homeostatic interoceptive state against which extreme affects are to be distinguished. Nietzsche seems instead to be getting at the idea that weaker affects lack phenomenal distinctness or salience. These states would still modulate one's psychological processes (as discussed in §2.2) and be causally related to conscious valuations (see BGE §187), as well as inclining or disinclining in regard to actions (see eKGWB 1883: 7 [121]; 1887: 11 [310]). The near-constant, lower-level states that Nietzsche claims "weave the web of our character and destiny" (D §115), however, do not involve severe enough bodily changes to manifest in a state of general feeling qualitatively distinct from the homeostatic background hum to which their milder manifestations contribute.[57]

Nietzsche thus holds, *contra* the extant view, that affects can (and indeed often do) generate phenomenally indistinct states of feeling. Those milder psycho-physiological responses that suffuse our mental lives fail to reach A_F, operating beneath the level at which we notice bodily arousal. Furthermore, one faces an obvious challenge in bringing those lower-level affects that fall short of A_F to point A_R. In many such instances, one will not be able to self-ascribe an occurrent affect through introspection alone (even if in possession of appropriate concepts) owing to one's being unaware of the affective episode's taking place.[58] In typical cases, these states are not only phenomenally indistinct, but also reflectively non-conscious as a consequence of being introspectively inaccessible.

---

[57] Elsewhere, Nietzsche writes, "That weak sensations are estimated to be equal, are experienced *as the same*, is the fundamental fact" (eKGWB 1884: 25 [168]). He notes in the following year "how much there is through the sensorium commune of which we have hardly a shimmer of consciousness!" (eKGWB 1885: 38 [1]). The sense of non-conscious operative here suggests the influence of Leibniz (see GS §357). It also possibly reflects the influence of the idea of a 'threshold' (*Schwelle*) of consciousness, which originated in the work of Johann Friedrich Herbart, and influenced (among others) Ernst Heinrich Weber, Gustav Theodor Fechner, and Hermann von Helmholtz. The threshold marked the point at which a stimulus (or difference between stimuli) became (or ceased to be) noticeable. It is unclear how much direct contact Nietzsche had with the relevant writings of the above figures, but it is highly unlikely that he was unfamiliar with the idea. Indeed, he uses the phrase "threshold of consciousness [*Schwelle des Bewusstseins*]" in GM I, §1.

[58] This would not, of course, prevent one from reporting affective dispositions.

There is a second way in which one might be unable to bring an occurrent affect to reflection, which follows from Nietzsche's remarks on 'conceivability'. He writes in the *Nachlaß* that "inner experience" enters consciousness "only after it has found a language that the individual *understands* . . . i.e. a translation of a state into states *familiar* to him" (eKGWB 1888: 15 [90]). This "translation" refers to linguistic interpretation, which (like affective interpretation) will usually occur automatically, utilizing material stored in memory. In some cases, one's state might defy straightforward explanation, requiring careful conscious thinking to deploy an interpretation. In certain instances, however—those where *no* cause can be found or envisaged, or where we lack applicable or adequate concepts—one's condition resists linguistic interpretation. As Nietzsche writes in *Daybreak*, "where words are lacking, we are accustomed to abandon exact observation because exact thinking there becomes uncomfortable" (D §115). Since one cannot think consciously about a given state without linguistic interpretation having occurred (even if, for example, one's bodily perturbation is sufficiently intense to have reached A_F), on those occasions where one's state cannot be subsumed under a "*familiar*" description,[59] the state remains reflectively non-conscious in virtue of being 'inconceivable'.[60]

Conclusion

I suggested at the outset that Nietzsche understood occurrent affects to be complex psycho-physiological processes, comprising one's bodily state (standardly of perturbation in response to a stimulus or stimuli) and the modulation of specific psychological processes, consequent

---

[59] The pressure of 'cognitive-psychological necessity' is such that we sometimes employ words even though they are a poor fit. Nietzsche writes, "We always express our thoughts with the words that lie to hand. Or, to express my whole suspicion: we have at any moment only the thought for which we have to hand the words able to express it approximately" (D §257).

[60] It has been suggested that states which reach A_F but not A_R might be experienced as brute bodily feeling. There will surely be instances when this is correct (e.g. where we lack a putative cause of *any* sort [see n. 40]), but such cases would not exhaust forms of inconceivability. Plausibly, episodes of what we could call 'inexact thinking'—say, where one finds oneself feeling bad about something, but cannot identify the *particular* affect one is undergoing, or where one is unsure about the underlying condition disposing one to particular outbursts—would qualify as instances of inconceivability (as one could not think consciously about being in a particular condition), but would not necessarily be experienced simply as bodily feeling. (Thanks to a reviewer for the *Journal of the History of Philosophy* for raising this suggestion.)

upon the sensation of one's state. In addition to this somatic account, I claimed, Nietzsche offers an interpretive account of self-ascription. We have now added some detail to this outline. Nietzsche's picture starts with an occasioning object, perceived *qua* stimulus owing to one's psycho-physiology. This rouses a pattern of bodily processes (physiological 'releases of force'), the sensations of which merge with exteroceptive sensory information. These bodily sensations alter features of one's psychology, modulating exteroceptive interpretation, and exciting the cause-extracting mechanism. Affective interpreting makes a causal connection between one's state and surroundings, valences these surroundings, and makes salient certain features and affordances. This process occurs partly under the pressure of 'practical-psychological necessity', so as to provide an object in relation to which one's affects can be discharged.

Bodily perturbations span a range of intensities, with the modulation of psychological processes occurring in milder cases irrespective of whether one is aware of an occurrent affect. In "*superlative*" (D §115) instances, we become phenomenally aware of distinct states of feeling, A_F. Those insufficiently intense to be discriminated against the normal interplay of low-level states and processes usually remain non-conscious. In those instances where arousal is sufficient to reach A_F, one's state can potentially be identified as this-or-that affect. This process is linguistic interpretation, through which one's state is subsumed under an affect-word. One might also bring the putative relationship of one's condition to a cause-object extracted by affective interpretation to a point of reflection, in the form of (often intricate) linguistically articulated aetiologies. This is driven by 'cognitive-psychological necessity'. Once linguistic interpretation has occurred, natural language constructions utilizing the relevant affect-word(s) provide the content of conscious thoughts, and one becomes conscious of one's state at the level of reflection, A_R. Those states that resist interpretation (which cannot be subsumed under appropriate or adequate terms) remain reflectively non-conscious.

This fleshed-out picture provides us with something to say in response to the question of what affects, for Nietzsche, are "supposed to be" (Poellner, 2007: 229). It still leaves much work to be done, of course, including an examination of the merits and shortcomings of Nietzsche's

reflections, which I have been unable to pursue here. Nevertheless, I hope to have made some progress in making sense of his intriguing (and sometimes perplexing) remarks on the subject. Nietzsche departs from the extant view in a number of ways, and presents to us a position more substantive and less straightforward position than has often been thought. As we have seen in the course of unpacking the view, the above account underpins a number of significant claims found in Nietzsche's writings, and, I hope, provides a basis for understanding how affects can play the roles he ascribes to them in his philosophical psychology.[61]

[61] Earlier versions of this paper were presented at the University of Oxford, Birkbeck, University of London, and John Cabot University, Rome. Thanks are owed to the audiences and participants for their comments and questions. My thanks also to Stephen Mulhall, Manuel Dries, Jonathan Mitchell, Alex Prescott-Couch, James Matharu, Jay Jian, Alexander Heape, Franziska Poprawe, and Laurenz Casser, as well as to a reviewer and editor for the *Journal of the History of Philosophy*. I am particularly indebted to Peter Kail for numerous helpful discussions of the material here presented. I would also like to express my gratitude to the Arts and Humanities Research Council, who funded this research.

2

Conscious and Unconscious Thought†

Abstract: In this paper, I examine Nietzsche's views on 'thinking'. While much recent attention has been directed towards his reflections on the mind, and on consciousness in particular, considerably less attention has been paid to Nietzsche's interesting and suggestive remarks on thought. My starting point will be the claims that we "think constantly, but [do] not know it", and that only our conscious thinking "*takes place in words*" (GS §354). In the opening half of the paper, I piece together Nietzsche's understanding of unconscious thought, highlighting the role he envisages for drives and affects. Thereafter, I move on to consider the difference, for Nietzsche, between conscious and unconscious thinking. In particular, I address two important accounts. The first, put forward by Paul Katsafanas, claims that conscious states *alone* have conceptually articulated content. The second, defended most prominently by Mattia Riccardi, argues that Nietzsche's various claims evince a form of higher-order thought (HOT) theory. I argue that neither reading is quite right, and instead propose an alternative interpretation of conscious thinking "*in words*", which draws on work on inner speech.

Keywords: Nietzsche, cognition, affect, consciousness, concepts, inner speech, HOT theory

INTRODUCTION

As Nietzsche's only extended discussion of consciousness in the published works (and one of the few substantial discussions of the topic anywhere in the corpus), §354 of *The Gay Science* has attracted a fair bit of attention recently from those looking to reconstruct his views on the mind.[1] Despite interest in the passage, it is rarely noted that in addition to the claims about consciousness and communication found therein, GS §354 also contains a number of suggestive

† Material from this chapter was published by Taylor & Francis Group in *Inquiry: An Interdisciplinary Journal of Philosophy*, 62[1], pp. 1–22, on 27th September 2018, and is available online at https://www.tandfonline.com/10.1080/0020174X.2019.1527537.

[1] See, for example, Katsafanas (2005; 2016); Riccardi (2018; 2015); and Constâncio (2011).

remarks about *thinking*. Nietzsche writes that "man, like every living creature, thinks constantly, but does not know it", and that we could indeed "think, feel, will, remember, and also 'act' in every sense of the term", without any of this having "to 'enter our consciousness'". Regarding the thinking that does become conscious, we are told that it is but "the smallest part" of our mental lives. It is this thinking, and this thinking *alone*, Nietzsche insists, that "*takes place in words*" (GS §354).[2]

Nietzsche evidently had *some* considered views on thinking, albeit elusive ones. My aim here is to tease out the positions that underpin such claims and reconstruct Nietzsche's reflections on thought, as far as is possible. Such an endeavour has bearing not only on the interpretation of GS §354, but also on debates about his views on agency, deliberation, and consciousness more broadly. This enterprise is not without its challenges, however. Nietzsche's remarks on thinking are often greatly condensed, and sometimes inchoate or incomplete. In addition to this, many of the more substantial discussions are to be found in the *Nachlaß*. My intention, however, is not to ascribe to Nietzsche a fully worked-out, systematic account of cognition. Instead, I wish to argue that strands present in the published works, or traceable from the *Nachlaß* to his published writings, offer means of understanding the comments on thought in GS §354. I hope to demonstrate that despite being broad in its brushstrokes, the picture which emerges is nevertheless a philosophically interesting one.

I begin by clarifying which sorts of thinking concern us here (§1), before moving to Nietzsche's remarks on unconscious thought (§2). In particular, I highlight three facets of his picture of unconscious cognition. The first concerns the format of the thinking that occurs without our being conscious of it, which I shall claim Nietzsche takes to be imagistic (§2.1a). The second is

---

[2] Quotations from the published works are from the translations listed in the bibliography (with occasional amendments) and are cited using the abbreviations there listed, followed by Arabic numerals. Where works comprise several volumes, books, or essays (such that sections are not numbered consecutively), either Roman numerals or abbreviated section titles precede Arabic numerals. For example, GM I, §1 is *On the Genealogy of Morality*, Essay I, Section 1; and TI "Errors", §4 is *Twilight of the Idols*, "The Four Great Errors", Section 4. Citations to the *Nachlaß* refer to year, group, and section/fragment number (eKGWB year: group [note]). Translations from the *Nachlaß* are my own.

the importance of an interpretive, analogical element to much unconscious thought (§2.1b). The third feature of unconscious thinking I wish to highlight is the pervasive role of affective states in such processing (§2.1c). The subsequent section of the paper addresses conscious thought, and the relationship between conscious and unconscious thinking (§3). Therein, I consider the interpretations of Paul Katsafanas (§3.1) and Mattia Riccardi (§3.2), arguing that neither succeeds in characterizing Nietzsche's understanding of conscious thinking. In response, I propose an alternative view (§4), which draws on contemporary work on inner speech.

1.

Any attempt to unpack Nietzsche's writings on thought runs immediately into a puzzle. As well as suggesting that thought is constant in "every living thing" (GS §354), Nietzsche is also occasionally entirely dismissive of thinking, placing it among a group of "fictions that are of no use" (eKGWB 1888: 14 [122]; cf. 1887: 11 [113]). Either one of these claims would constitute a substantial philosophical commitment on its own. Their conjunction, however, raises the question of whether Nietzsche's view is even coherent. The solution to this problem is to recognize that his writings on thought convey a view that is both 'permissive' and 'pluralistic'. By this, I mean that he applies 'thinking'-language to phenomena that would not normally be considered in such terms, and that he uses 'thinking' to denote several different things. Accordingly, before we proceed, we need some clarity regarding the sorts of thinking that concern us here.

Nietzsche in fact applies the 'thinking'-terminology to four different phenomena, two of which can be put aside for our purposes. The first, hereafter labelled *function*-thinking, refers to his extension of such terminology to what would normally be considered physiological functions.

Examples of this can be found in his imputation of mindedness to basic life forms,[3] and in his explanation of the "*purposiveness* [that] rules over the smallest events" in more complex organisms—"precaution, selection, bringing-together, reparation, etc."—by reference to a "*far higher* and more comprehensive *intellect* than we know of" (eKGWB 1883: 24 [16]; cf. Z I, §4). The second form of thinking to be shelved is *subject*-thinking. This is prominent in Nietzsche's later writings, and in *Beyond Good and Evil* in particular. Consider the line of reasoning presented (rhetorically) in BGE §17, on which thinking is "an activity; for each activity there is someone who acts; therefore—", or the suggestion that common people treat thinking as "an activity for which a causal subject *must* be thought" (BGE §54). Here, thinking is understood as the activity of "an indifferent substratum" (GM I, §13); a substantive mental subject, distinct from its token mental states.

I mention *function*- and *subject*-thinking to clear the way for our discussion, as some of Nietzsche's more puzzling remarks arguably concern *these* forms of thought. If we look to the passage in which he dismisses thinking as a "fiction", for example, it is evident that Nietzsche has *subject*-thinking in his sights.[4] It is plausible, moreover, that some of his remarks about universal thought are perhaps motivated in part by the idea of *function*-thinking (though not entirely, as we shall see). Having flagged these up, we can steer clear of them. *Subject*-thinking, as a fiction, is of little relevance to Nietzsche's positive account of thought. And since our focus concerns his reflections on phenomena more commonly identified as instances of thought, we can also put aside *function*-thinking for another day. This leaves us with two remaining forms of thought, which I am going to label simply *conscious*-thinking and *unconscious*-thinking. These refer, respectively, to the thinking that alone takes place "*in words*", and the thinking that occurs

---

[3] E.g., Nietzsche claims to presuppose "*memory and a kind of mind in everything organic*" (eKGWB 1884: 25 [403]). (For a comparison of such remarks with similar claims from the physiology of the day, see Moore, 2002: 39–43.)

[4] He writes therein of the "'spirit', *something that thinks*", and claims that "*first* an act is imagined which simply does not occur, 'thinking', and *second* a subject-substratum in which every act of thinking, and nothing else, has its origin: that is to say, *both the deed and the doer are fictions*" (eKGWB 1887: 11 [113]; cf. GM I, §13; TI "Errors", §3).

"constantly" without our knowing (GS §354). We shall turn our attention to the former of these a little later on. For now, our focus is *unconscious*-thinking.

2.

If we return to GS §354, Nietzsche suggests not only that we could "think, feel, will, remember" without being conscious of it, but that "the predominant part" of our "thinking, feeling, and willing lives" takes place "without this mirroring". The passage offers little regarding the nature of this thinking, beyond the claims that it can occur unconsciously and will not take place "*in words*". Unfortunately, this doesn't get us far. Thinking "*in words*" might refer to the explicit articulation of the content of one's mental state in natural language (e.g. thinking the words to oneself). This would be compatible with the claim that *unconscious*-thinking (i.e. the thinking *not* articulated explicitly in such language) nevertheless involved language-like vehicles. Alternatively, thinking "*in words*" might refer to these vehicles, and thus to the structure of the thought-processes themselves. So understood, the claim would be that *conscious*-thinking alone involves the language-like representational format required for thought-processes to operate along logical relations between contents, making *conscious*-thinking markedly different from its unconscious counterpart.[5]

Consequently, GS §354 tells us less about *unconscious*-thinking than we might have hoped. If we look elsewhere, Nietzsche's remarks suggest a considerable difference between unconscious thought-processes and our picture of conscious thinking. This, however, requires a little unpacking. The first thing to note is that Nietzsche is clearly sceptical about the conception of thinking we derive from introspection. The "logical thinking of which logic speaks" (clarified as "a thinking where the thought itself is posited as the *cause* of new thoughts") is claimed to be "the pattern of a complete fiction: *a kind of thinking which, in reality, never occurs*" (eKGWB 1885:

[5] These construals are, of course, not mutually exclusive.

38 [2]; cf. eKGWB 1885: 34 [249]). This is because the inferences, etc., presented as articulations of the thought-process are reconstructed entirely from fragmentary evidence, viz. conscious thoughts, providing no account of the unconscious operations between said thoughts. It is, he writes elsewhere, "only apparent that a thought would directly cause another thought"; instead, the "events which are actually linked play out beneath our consciousness: the emerging sequence and one-after-the-other of feelings, thoughts, etc., are symptoms of the actual event!" (eKGWB 1885: 1 [61]).

Nietzsche's suggestions about this "actual event" are *prima facie* puzzling. In a *Nachlaß* passage in which he claims that a "*logical* process of the kind we 'find in a book' *never occurs*", Nietzsche insists, "Our thinking works fundamentally differently: between a thought and the next, an in-between world of entirely another kind prevails". The only hint provided therein regarding this "in-between world" is a mention of "the drive to contradiction, or to subjugation, etc." (eKGWB 1885: 34 [170]). Plausibly, this opaque remark relates to a theme ubiquitous in Nietzsche's published and unpublished writings regarding the role of drives/affects in cognition. In the third book of *The Gay Science*, for example, he claims that the "course of logical thoughts and inferences" actually "corresponds to a process and a struggle of drives that, taken separately, are all very illogical and unjust", and that "we usually experience only the result of the struggle" (GS §111).[6] In the fourth book of the same work, in a passage that takes a remark of Spinoza's as its starting point, Nietzsche writes of "*intelligere*" (understanding) as "*a certain behaviour of the drives towards one another*" (GS §333). If we look to the later writings, he writes of thinking in *Beyond Good and Evil* in almost identical words.[7]

---

[6] To relate this to Nietzsche's concern about the conception of thinking we derive from introspection, he writes in the subsequent section that "there is probably never such a duality [as cause and effect]; in truth a continuum faces us, from which we isolate a few pieces […] The suddenness with which many effects stand out misleads us; it is a suddenness only for us. There are an infinite number of processes that elude us in this second of suddenness" (GS §112).

[7] Thinking, Nietzsche writes, "is only a behaviour of [the] drives towards one another" (BGE §36; cf. eKGWB 1887: 11 [310]).

Similar claims abound in the *Nachlaß*. Nietzsche writes of (presumably conscious) instances of thinking as "not the inner events themselves", but a "sign-language for the power-balance of the affects" (eKGWB 1885: 1 [28]; cf. 1885: 1 [75]). In the passage where he describes thoughts, etc., as "symptoms of the actual event", he writes that beneath every thought "lies an affect": "*Each thought*, each feeling, each will", he writes, "is born *not* of a specific drive, but is a *total state*, a whole surface of the entirety of consciousness, and results from the ascertainment of the power of *all* our constituent drives in that instant [...] The next thought is a sign of how the entire power-situation has shifted in the meantime" (eKGWB 1885: 1 [61]).[8]

As we saw, our cognition is claimed to be radically different to the logical thought "of which logic speaks" (eKGWB 1885: 38 [2]) because such formalizations are simplified reconstructions of thinking-processes, abstracted from patterns of conscious thoughts. On the alternative view here presented, the "events which are actually linked" that "play out beneath our consciousness" (eKGWB 1885: 1 [61]) operate not through a succession of language-like representations, following rules of inference, etc., but (at least partly) through a struggle between affects and/or drives,[9] of which we experience only the result. It's one thing to suggest that Nietzsche held this view, however, and quite another to explain what it means, or how it could possibly represent a plausible or interesting conception of thinking. His various remarks offer a mixture of clearly-stated, even precise, claims, with some suggestive-if-rather-inchoate ones. It is to the task of reconstructing Nietzsche's picture of unconscious thought that we now move.

---

[8] Constâncio (2011) offers an interesting discussion of this passage.

[9] Nietzsche appears to use 'drive' and 'affect' interchangeably in regard to thinking. For the sake of simplicity, I will talk of 'affects' henceforth.

2.1

The first step in unpacking the idea of thinking as partly comprising an interaction of affects concerns the nature of such states themselves. Since we have insufficient room here for an extensive examination of his view of affective states,[10] I wish only to highlight two features that should be uncontroversial: first, that Nietzsche takes affects to be states that commonly occur in *response* to stimuli; and second, that he takes them to *modify* certain psychological processes. With these in mind, I'm going to make three claims which will hopefully add flesh to the bare bones laid out thus far. The first is that affect-thinking is 'imagistic'; the second, that it involves interpreting and a form of 'analogical' thinking; and the third, that it is the constant presence of affects that properly explains the putatively unabating nature of thought. Let's take these in turn.

2.1a

It is not uncommon for philosophers to claim that the format of those representations utilized in thinking-processes must be *language*-like. Recent work, however, some of it inspired by the aim of providing a compelling account of animal cognition, has emphasized the forms of thinking possible through an *image*-like representational format. While Nietzsche clearly lacked such terminology, his writings reflect a sensitivity to the idea of imagistic thinking. In several early *Nachlaß* passages on 'unconscious inferences',[11] he writes that they "arouse [his] doubts", suggesting instead that the process "will arguably be the passing from *image to image*". The "image-thinking [*Bilderdenken*]" of which Nietzsche writes is purportedly not of a "strictly *logical* nature, but is still more or less logical". Philosophers, by contrast, are claimed to have

---

[10] I have argued that occurrent affects, for Nietzsche, are complex psycho-physiological processes, comprising one's state (standardly of perturbation, in response to a stimulus or stimuli) and the modulation of specific psychological processes, resultant upon the sensation of said state (paper one, this thesis). For alternative discussions, see Leiter (2013) and Poellner (2007).

[11] The term was coined by Hermann von Helmholtz in *Handbuch der physiologischen Optik* (1867).

endeavoured "to set concept-thinking [*Begriffsdenken*] in the place of image-thinking" (eKGWB 1872: 19 [107]; cf. 1872: 19 [217]).

If we look to another note on 'image-thinking' from the same year, Nietzsche elaborates further. The process, he writes, is one on which the "intellect rapidly chooses similar images: the chosen one generates a whole plethora of images once more: quickly, however, it chooses one from this again, etc." (eKGWB 1872: 19 [78]). As an alternative to thinking understood as operating along logical relations between linguistically structured contents, Nietzsche proposes a form of *non-linguistic* thinking which works on imagistic vehicles. In this respect at least, his position hardly changes. Nine years after the above remarks about 'image-thinking', Nietzsche writes that "imagination is to be posited in the place of the '*unconscious*': they are not unconscious inferences, but rather *possibilities thrown out*, provided by the imagination" (eKGWB 1881: 11 [13]). If we move forward another two years, we find Nietzsche writing of "how thoughts are only images, how words are only signs of thoughts" (eKGWB 1883: 24 [16]).

To return briefly to the early notes, Nietzsche therein relates 'image-thinking' to drives. Having passed from "*image to image*", he writes, the final representation upon which the intellect settles "acts as a stimulus and motive": the instincts, he goes on, "also appear to be such an image-thinking, that finally becomes stimulus and motive" (eKGWB 1872: 19 [107]). We can trace this line of thought through to Nietzsche's mature work. In a passage redolent of the early notes, he conjectures that purposes, which "*usually* [occur] in the brain only after everything has been prepared for [their] execution" are "an 'inner' 'stimulus'—no more" (eKGWB 1883: 24 [34]). Elsewhere, he asks, "why could 'a purpose' not be an *epiphenomenon* in the series of changes in the forces that bring about purposive action—a pale sign-image [*Zeichenbild*] thrown into consciousness beforehand, that serves to orientate us concerning what is happening, even as a symptom of events, *not* as their cause?" (eKGWB 1886: 7 [1]).[12] By the time we reach the

[12] Remarks of this sort might be taken to support strongly the idea that Nietzsche embraced a form of epiphenomenalism about consciousness. Nietzsche's point in such passages, however, seems to be that the

publication of the fifth book of *The Gay Science* twelve months later, Nietzsche discusses two "*kinds of cause that are often confused*", distinguishing between the "cause of acting" and the "cause of acting in a certain way, in a certain direction, with a certain goal". Among the latter class, he places "so-called 'purposes'", which he countenances could be "a beautifying pretext, a self-deception of vanity after the fact" (GS §360).

This would not, of course, entail that *all* purposes must be imagistic. Nevertheless, Nietzsche likely took imagistic thinking to be at least part of the process that has purposes as its output, and very often the entirety. On his view, the 'imagination', stimulated by sensory information, produces a rapid succession of images, culminating in a final 'sign-image'—a mental representation of the state of affairs to be brought about (or avoided) through action, identifiable consciously as the 'purpose' of acting. The "represented goal of action", Nietzsche suggests, "provides an *anticipation* of release, and *thereby stimulates discharge ever further*", with the "subsequent action" providing the "actual release". In the process, the "doer *forgets* the actual *driving force* and *sees* **only** the "motive"' (eKGWB 1883: 7 [77]; cf. D §129).[13] So understood, our conception of action is distorted by the range of evidence to which we have introspective access. The 'purpose' is itself the result of a complex process, and serves to stimulate further affects and orientate subsequent action.

---

contents of consciousness shouldn't be properly thought of as the *cause* of acting, or as the initial point in the psychological process. In this respect, Nietzsche's view as expressed in such claims seems to bear greater resemblance to forms of eliminativism about (folk-psychological) accounts of conscious propositional attitudes (cf. Carruthers, 2011: 373–80). This would not entail that conscious contents lacked causal consequences, or that certain contents would have had such consequences irrespective of whether or not they had been conscious. (I make no comment here as to whether Nietzsche advocates epiphenomenalism about consciousness elsewhere.)

[13] Cf. Georg Heinrich Schneider, who claimed that, when choosing, "the representation that presently appears as the most purposive, relatively, always leads to the decision, precisely because the representation of what is most purposive relatively also causes the strongest feeling and the strongest drive, and the strongest drive also gains dominance over the others, and necessarily leads, and must lead, to action" (Schneider, 1882: 329). (Nietzsche owned a copy of this work.)

2.1b

Although such remarks fall short of constituting an 'account' of imagistic thought, they do offer a sketch of the interaction of image-like representations and affects to fit into our reconstruction of Nietzsche's *unconscious*-thinking. As it stands, however, this is a somewhat thin picture of non-linguistic cognition. Happily, there is more we can draw from Nietzsche's writings. His remarks on 'interpretation', in particular, render his view of *unconscious*-thinking somewhat more substantial. Nietzsche discusses interpreting both in relation to drives and affects. Plausibly, the process will involve, *inter alia*, the generation of perceptual saliences (see Katsafanas, 2013: 740). Indeed, Nietzsche's remarks suggest something not dissimilar to affordance perception (Gibson 1977; 1979). Objects, events, and people *afford* different behaviours, interactions, and opportunities: a handle affords the opportunity to grab; a door, the opportunity to exit. If we look to a prominent passage from *Daybreak*, Nietzsche suggests therein that a drive which "desires gratification" regards "every event of the day with a view to seeing how it can employ it for the attainment of its goal" (D §119; cf. eKGWB 1885: 1 [58]). So understood, drives make salient both features of the environment and the opportunities, interactions, etc., the environment affords, through which one's state can be discharged.

Nietzsche arguably has a similar view in mind for affects.[14] Indeed, he writes in the *Nachlaß* that "interpreting itself [...] has existence (but not as a "being", rather as a *process*, a *becoming*) as an affect" (eKGWB 1885: 2 [151]). The drives and affects can be said to have their own 'perspectives', from which sensation is interpreted and evaluated: elements of one's surroundings are valenced as objects of inclination or aversion, and ascribed a meaning and/or function from the practical perspective of the drive or affect.[15] Nietzsche sketches the sort of

[14] For an influential discussion of drives and their relation(s) to affects, see Katsafanas (2013). I offer some remarks on this relation in this thesis (see paper one, §2.2 and n. 35).

[15] Cf., from the *Nachlaß*: "From each of our fundamental drives there is a different perspectival valuation of all events and experiences" (eKGWB 1885: 1 [58]).

process he envisages in a late notebook passage. The "origin" of pleasure and displeasure, he writes, "is in the central sphere of the intellect": their "precondition is an infinitely accelerated perceiving, ordering, subsuming, recalculating, inferring". Feelings of pleasure and displeasure, described as "*will-reactions (affects)*", presuppose a "*measuring regarding overall usefulness*, *overall harmfulness*: thus, a sphere where the willing of a target (state) and a selection of the means takes place" (eKGWB 1887: 11 [71]). These remarks clearly resonate with the discussion of 'purposes' in §2.1a. On Nietzsche's picture, one's interaction with their environment is dynamic and affect-guided, with such states deployed in response to an ever-updating assessment of one's surroundings, resulting in the generation of perceptual saliences and the highlighting of affordances.

The sorts of interpretation that Nietzsche thinks affective states involve seems to instantiate a form of non-linguistic thinking called 'analogical' thinking. This refers not to an explicit, conscious connection of ideas or situations, but to the perception of analogies between them (see Bermúdez, 2003: 36). In such instances, one sees something as saliently similar to a previously experienced thing, and can bring one's 'knowhow' to bear upon it. A picture of this sort emerges from Nietzsche's remarks on the psychological mechanism that extracts putative causes for one's occurrent states from their surroundings. The most prominent discussion of this can be found in *Twilight of the Idols*.[16] Nietzsche writes therein that "Most of our general feelings—every sort of restraint, pressure, tension, explosion in the play and counter-play of our organs, the condition of the *nervus sympathicus* in particular—excite our cause-drive". Memory, which "becomes active without our being aware of it", then draws up "earlier states of a similar kind, and the causal interpretations that have grown out of them". In time, Nietzsche insists, there "arises a *habituation* to a certain causal interpretation" (TI "Errors", §4), the deployment

[16] Cf. HH "First and Last", §13; D §119; eKGWB 1883: 24 [20]; 1884: 26 [92]; 1885: 38 [1]; 1888: 15 [90]. For a discussion of the 'cause-extracting mechanism', see paper one in this thesis.

of which, as something "already *known*, experienced, inscribed in memory", has a palliative effect (TI "Errors", §5).

Many of the aetiologies utilized in such circumstances will naturally be linguistically articulated and theory-laden, such that it would be implausible to ascribe them to non-linguistic creatures. Nevertheless, Nietzsche attributes a degree of fundamentality to the referral of sensation to putative cause, sufficient to render unlikely the suggestion that it occurs *only* in language-using creatures (see eKGWB 1888: 15 [90]; 1885: 34 [54]).[17] Indeed, close attention to his remarks reveal that he took the connection of event and state to be grounded in experience (eKGWB 1883: 24 [20]; 1880: 1 [115]; TI "Errors", §5), and the process of identifying circumstances as those for which a particular explanation is apposite to rely more upon association, proceduralization, and pattern recognition than linguistically mediated reasoning. The process of identifying a putative cause for one's state, and thus of identifying which learned behaviours to pursue in order to discharge one's state, involves perceiving new situations with which one is confronted as amenable to interpretations already forged through experience.

2.1c

If this is correct, the 'struggle' of affects that Nietzsche envisages is in part a struggle between competing practical-evaluative perspectives, with each rendering salient features and affordances, and inclining towards the generation of 'purposes'. This provides some insight into how a fluctuating "power-situation" involving "*all* our constituent drives" (eKGWB 1885: 1 [61]) could be countenanced as partly constitutive of thinking. There remains, however, a further element of Nietzsche's affect-driven *unconscious*-thinking to highlight, related to his claim in GS §354 that we think "constantly, but [do] not know it". I mentioned earlier in my brief

[17] Cf. Schopenhauer, who writes, in a speculative passage on the development of the intellect and the continuity between sentient and non-sentient life, that were a plant's receptivity to its environment to intensify, it would "finally become sensation, accompanied by the capacity to refer sensation to its cause, and at last perception" (PP II, Kap. 3, §33).

remarks on *function*-thinking that I did not take it to be the whole story in terms of Nietzsche's comments about the ubiquitous and unabating nature of thought. Indeed, I would contend that the best explanation for these claims is the idea that drives and affects saturate mental life. If we turn to *Daybreak*, Nietzsche writes therein that words "exist only for *superlative* degrees [of] processes and drives", claiming that we "*are none of us* that which we appear to be in accordance with the states for which alone we have consciousness and words". We "*misunderstand* ourselves", he writes, on the basis of these "cruder outbursts", as we "draw a conclusion on the basis of data in which the exceptions outweigh the rule". Those instances of which we become conscious are actually "*extreme* states". By contrast, it is claimed that the "milder, middle degrees [of "processes and drives"], not to speak of the lower degrees that are continuously in play", manage to "elude us" (D §115).

Thought is argued to occur "constantly" (GS §354), often without our knowing, because the interaction of affects (including the stimulation and modulation of psychological processes by affects) is partly constitutive of *unconscious*-thinking, and such states are effectively constant features of mental life. The view of 'thinking' with which we are left arguably foreshadows elements of Antonio Damasio's account of practical reasoning. Damasio suggests that when making decisions, we entertain representations of the available options: the "key components", he claims, "unfold in our minds instantly, sketchily, and virtually simultaneously, too fast for the details to be clearly defined". Instead of engaging in lengthy conscious deliberation, when the "bad outcome connected with a given response option comes into mind, however fleetingly, you experience an unpleasant gut feeling" (Damasio, 1994: 173). Emotions act as a sounding-board for the contents of consciousness, marking various options with positively or negatively valenced feeling. Accordingly, for Damasio, they play an essential role in certain rational behaviours, and a far more pervasive role in human psychology than commonly assumed.

Although Nietzsche's picture is sometimes underdeveloped, and less systematic than Damasio's, the broad brushstrokes of each are remarkably congruent. As Damasio envisages processing unfolding "instantly, sketchily", etc., with emotions 'marking' represented options, Nietzsche

talks of an "infinitely accelerated perceiving, ordering, subsuming, recalculating, inferring" with valenced "*will-reactions (affects)*" as outputs (eKGWB 1887: 11 [71]), and of the "represented goal of action" providing "*anticipation* of release" (eKGWB 1883: 7 [77]). As Nietzsche talks of "milder" states, which abound between "*superlative*" instances, Damasio writes of "background feelings" that correspond to the "body state prevailing *between* emotions", which are "neither too positive nor too negative". Such feelings are "perceived as mostly pleasant or unpleasant", and, he suggests, likely contribute to moods (Damasio, 1994: 150–51). Both present affective states as ebbing and flowing, responsive to the images, 'purposes', and thoughts that become conscious. For Nietzsche, this process is near-constant. Those affects strongest at a given time dominate the intellect, stimulating and modifying psychological processes. They have at their disposal the imagination and memory, as well as the causal interpretations, etc., built up through experience. These dominant affects determine one's evaluative perspective and the affordances experienced as salient, and in doing so determine the patterns of activity pursued.

Nietzsche's suggestions regarding *unconscious*-thinking, although clearly in need of precisification and development, are far from foolish. Indeed, they anticipate elements of recent work on non-linguistic thought. In perhaps the most extensive treatment of the topic, José Luis Bermúdez discusses both imagistic and analogical reasoning as examples of forms of thought that don't rely on linguistic vehicles (see Bermúdez, 2003: 36). Bermúdez also distinguishes between what he calls 'thinking-*that*' and 'thinking-*how*' (analogous to Gilbert Ryle's distinction between 'knowing-that' and 'knowing-how'). Paradigmatic cases of thinking-*how*, for Bermúdez, have several important things in common. First, they could plausibly be described as *skills* (as opposed to bodies of knowledge). They are, more precisely, skills "that allow one to modify and adjust oneself to the environment" (Bermúdez, 2003: 36); hence Bermúdez's emphasis on the importance of affordance to any account of thinking-*how* (see Bermúdez, 2003: 47). Second, in cases of thinking-*how*, the thought processes themselves "cannot easily, if at all, be put into words". While one can naturally describe features of one's mental states or articulate conclusions, and so on, one cannot "express in language the thinking that leads up to that

conclusion" (Bermúdez, 2003: 37). Such remarks clearly chime with the role Nietzsche gives to drives and affects in an account of *unconscious*-thinking, and with his insistence that such thought "works fundamentally differently" (eKGWB 1885: 34 [170]) to the thinking of which we are conscious.

3.

It is of course true that we are capable not only of thinking-*how*, but also thinking-*that*—of which Nietzsche's *conscious*-thinking, the thought which "*takes place in words*", is presumably a paradigm. Nietzsche's remarks present us with a number of questions: how should we understand the relationship between *conscious*- and *unconscious*-thinking? And what does it mean for *conscious*-thinking alone to "*take place in words*"? In what follows, I turn to consider two influential interpretations of Nietzsche's remarks on consciousness, each of which offers a different way of answering the questions above. Neither of these approaches, I shall argue, is entirely successful.

3.1

In a number of sophisticated pieces on Nietzsche's view of the mind, Paul Katsafanas has argued that Nietzsche connects consciousness to conceptualization. More precisely, Katsafanas argues that the claim that only "conscious thinking" takes place "*in words*" (GS §354) should be understood as the claim that *all and only* conscious mental states have conceptual content (Katsafanas, 2016: 26). This conclusion is supported by four claims:

(i) Conscious thinking, and only conscious thinking, takes place in words (see GS §354).

(ii) There can be no words without concepts: Nietzsche conceives of words as "acoustic signs for concepts" (BGE §268), meaning "words express or signify concepts" (Katsafanas, 2016: 25).

(iii) To think in words, therefore, "is to think by means of concepts" (Katsafanas, 2016: 25–26)—i.e., what I have referred to as *conscious*-thinking is *conceptually articulated*.

(iv) There can be no concepts, however, "without words" (Katsafanas, 2016: 25): concepts are "possible", Nietzsche writes, "only once there are words" (eKGWB 1884: 11 [25]).

The conjunction of these claims, we are told, supports the conclusion that all and only "conscious thinking"—the states whose contents are articulated "*in words*"—is conceptually articulated. If this is correct, the relationship between the *unconscious*-thinking and *conscious*-thinking that interest us here is that the former becomes the latter through its content being conceptualized. Katsafanas' case for this conclusion is detailed and often insightful. I am not persuaded, however, that this is the correct way to understand Nietzsche's suggestion that *conscious*-thinking alone occurs "*in words*". Claims (i)–(iii), above, seem reasonably unproblematic. The key stage in Katsafanas' argument is (iv). It is *this* that binds words and concepts sufficiently closely to support his reading of GS §354, allowing him to move from the uncontroversial statement that *conscious*-thinking is conceptually articulated, to the more contentious claim that for Nietzsche, conscious content *alone* is conceptual. My concern is that there are different construals of (iv) that Katsafanas fails to separate. One relates to linguistic *ability*, the other to linguistic *articulation*. Once held apart, neither provides compelling support for reading Nietzsche's remarks about thinking "*in words*" in terms of conscious states alone involving concepts. Or so I shall argue.

3.1a

So, what might Nietzsche mean in suggesting that there are concepts "only once there are words"? On close inspection, there is more than one way to interpret the idea that there can be no concepts "without words", as Katsafanas puts it (Katsafanas, 2016: 25). One possibility is that Nietzsche means to say that a mental state can have conceptual content, only if its content is articulated in a natural language. So understood, words would be the sole vehicle for conceptual content. This is what I shall call the 'articulation' reading of (iv). Alternatively, one could read Nietzsche's remark as the claim that words are in some way necessary for concept

possession—i.e., that only once we possess language (in the relevant sense) can we possess concepts. On this 'ability' reading, only relevantly linguistically able creatures would be able to have mental states with conceptual content.[18]

Plausibly, Katsafanas must accept the 'ability' reading. For a start, the 'articulation' reading resembles the picture that emerges from Katsafanas' interpretation of Schopenhauer; a picture that Katsafanas claims Nietzsche came (at least in part) to reject (see Katsafanas, 2016: 30–31). *Contra* Schopenhauer, we are told, Nietzsche takes conscious perceptual states to have conceptual content. Katsafanas writes that perceptual content "would be conceptual if the perceived object were represented as an instance of some concept, that is, as a token of some type" (Katsafanas, 2016: 32). And this is purportedly what we find in BGE §192, where Nietzsche describes something that looks like categorical perception. Perceptual processing, as Katsafanas reads Nietzsche, putatively alters the content of perception, such that said content is adequately specifiable in terms of concepts the perceiver possesses. If this is correct, despite lacking content that is explicitly linguistically articulated, perceptual states might still be said to represent objects as "instance[s] of some concept" (Katsafanas, 2016: 32). Thus, if Katsafanas' claims here are right, this would rule out the 'articulation' reading.

What, then, of the 'ability' reading? Katsafanas argues that Nietzsche posits a close connection between conceptual capacities and linguistic abilities, with linguistic competence necessary for concept possession. For Nietzsche, mere discrimination, we are told, does not suffice: concept possession is also a matter of "understanding the concept's place in [a] system of relations" (Katsafanas, 2016: 25); of being able to employ it (alongside other concepts) in thought and communication (see Katsafanas, 2016: 36–37); and of being able to use it in abstract thought, without the thing in question being perceptually (or imaginatively) present (see Katsafanas,

[18] It should be noted that acceptance of the 'articulation' reading would entail acceptance of the 'ability' reading: if words are the only vehicles for conceptual content, only creatures in possession of a language could possess conceptual capacities.

2016: 36). Since these plausibly require linguistic competence, it is claimed that one can only possess a concept if one has mastery of a language.

If we look more closely at the 'ability' reading, however, it encounters its own difficulties. First, adequate textual support for the attribution of these views to Nietzsche is not really forthcoming. Regarding the suggestion that concept possession requires appreciation of a concept's place in a broader scheme, Katsafanas mentions that Nietzsche often links concepts to schema (see Katsafanas, 2016: 35, n. 30). In particular, Katsafanas highlights a passage in *Beyond Good and Evil* in which Nietzsche writes that "individual philosophical concepts are not arbitrary, do not grow up on their own, but in connection and relationship with each other", and remarks on "the innate systematicity and relationship of concepts" (BGE §20). This is all well and good—Nietzsche does link concepts to schemas with some regularity. But it remains unclear why such remarks should be taken to issue a *condition* for concept possession.[19]

The putative requirements on concept possession relating to both communication and abstract thought fare little better. In support of these, Katsafanas points us to BGE §268, where Nietzsche insists that using "the same words is not enough to ensure mutual understanding", stressing the need for shared experiences if people are to make themselves understood quickly and easily. Nietzsche's central point in the passage, however, seems to be that communication requires that individuals "use the same words for the same categories of inner experience" (BGE §268). The individuals in question might employ the same words, and already possess categorical abilities (with corresponding terms). The issue is the possibility of the same words corresponding to different categories in different people, the result of which is a misunderstanding between the parties involved. While this might suggest a requirement for successful communication, it is not

[19] The same can be said of some remarks from *On Truth and Lies in the Extra-Moral Sense*, which Katsafanas also mentions (see Katsafanas, 2016: 25, n. 17).

obvious that Nietzsche intends these remarks to highlight a condition on possessing the categorical, classificatory ability in the first place.

The evidence in favour of ascribing something like the 'ability' reading of (iv) thus appears to fall short of being compelling. A more decisive concern, however, is that even if such a reading could be more robustly defended, it would still fail to deliver the conclusion for which Katsafanas argues. Even if Nietzsche does tie concepts (and concept possession) to linguistic ability, the bind is plausibly not tight enough for Katsafanas' purposes. While the 'ability' reading provides a way in which conscious perceptual states could be said to have conceptual content, it does so at the cost of undercutting the motivation for the claim that unconscious states *could not*. Even if successful, Katsafanas' arguments would merely establish that, for Nietzsche, only linguistic creatures could have conceptually articulated mental states—not that only the conscious states of these creatures would have conceptual content. If Katsafanas' construal of Nietzsche's position relies upon the 'ability' reading of claim (iv), it fails to occlude the possibility that the *unconscious*-thinking of language-possessing creatures might well involve contents that represent objects as "token[s] of some type" (Katsafanas, 2016: 32).

3.1b

What, then, about the 'articulation' reading? Although incompatible with the conclusion that *all* conscious states (including, for instance, perceptual states) have conceptual content, might it not offer the correct account of the distinction between *conscious*- and *unconscious*-thinking? If words are the only vehicles for conceptual content, Nietzsche's claim that conscious thinking alone takes place in words would entail that unconscious states lacked conceptual content. As mentioned earlier, the 'articulation' reading resembles Schopenhauer's view, which surely influenced Nietzsche's thinking regarding consciousness and concepts, as Katsafanas rightly claims. The problem with this reading is that there are strong textual reasons to be cautious about ascribing the view to Nietzsche. Or, to be more precise, we should be cautious about ascribing it to the *mature* Nietzsche. There are clearly Schopenhauerian overtones to some of the younger Nietzsche's writings on thought. As we saw earlier, he contrasts "image-thinking" with

"concept-thinking", lamenting philosophers' efforts to account for everything in terms of the latter. He goes on to insist in the same passage that "Unconscious thinking must take place without concepts: thus, in *intuitions* [A n s c h a u u n g e n]" (eKGWB 1872: 19 [107]). If we look to *On Truth and Lies in the Extra-Moral Sense* (a passage to which Katsafanas draws our attention) Nietzsche again asserts the connection between concepts and linguistic articulation, claiming that "Every word immediately becomes a concept" (TL 1).

The passage from which (iv) is derived has a not-dissimilar feel. If we look at the broader quotation, Nietzsche writes therein, "First images [...] Then *words*, applied to images. Lastly, concepts, possible only when there are words—a summarizing of many images under something not clear, but audible (word)" (eKGWB 1884: 25 [168]).[20] The *structure* of Nietzsche's view changes little over time: brute sensation is reconstructed with the aid of "imagination" or "memory", utilizing "images", "forms", or "signs"—recognitional templates developed through a process of abstraction from particular experiences.[21] Linguistic ability then facilitates the articulation and communication of the content of our experiences. The issue for the 'articulation' reading is that Nietzsche's view appears to change over time regarding which of the above components should be considered 'concepts'. The greatest concentration of writings on both thinking and concepts occurs between the latter parts of 1884 and the early stages of 1886. Here, the Schopenhauerian emphasis on the essential importance of words to the development of concepts peters out. The thought expressed in the *Nachlaß* passage quoted by Katsafanas (from Spring 1884) never surfaces in the published works, nor (to my knowledge) does it appear again in the notebooks.[22]

---

[20] Cf. Nietzsche's remark in a note of Spring 1884 that concepts "emerge as auditory-images [*Hörbilder*] that condense a multiplicity of symbolic visual-images [*Seh-Bilder(n)*]" (eKGWB 1884: 24 [185]).

[21] See GS §114; BGE §192; eKGWB 1872: 19 [217]; 1881: 11 [13]; 1884: 26 [94].

[22] A note from Autumn 1884–Spring 1885 comes closest. Nietzsche mentions "primitive humans" communicating with signs: "First signs", he suggests, "then concepts, finally 'reason' in the customary sense" (eKGWB 1884: 30 [10]). This is at best ambiguous, however, since Nietzsche considered *gestures* to be signs

If we look to subsequent writings, Nietzsche's view seems gradually to shift. At a later point in 1884, his notes reflect an awareness of psychological mechanisms whose function is purportedly analogous to the process of concept development. Nietzsche writes that one "must relearn about memory", which he suggests is the "multitude of all the experiences of all organic life [...] simplifying, grouping together, and transforming into many unities". There must, he continues, be "an inner *process* that behaves like the *formation of concepts* [B e g r i f f s b i l d u n g] from many individual cases" (eKGWB 1884: 26 [94]). In mid-1885, he writes of 'signs' as essential to memory, referring in particular to *Zeichenschrift*—an image-like symbol (see eKGWB 1885: 34 [249]). It is around this time that the ideas which would eventually become the opening to §268 of *Beyond Good and Evil* first emerge. Nietzsche writes that "Words are acoustic-signs for concepts: concepts, however, are more or less secure groups of recurring sensations that appear together" (eKGWB 1885: 34 [86]), and that "concepts are signs for recognition" (eKGWB 1885: 1 [50]).

By the time these thoughts surface fully-developed in BGE §268, the line is of course that "Words are acoustic-signs for concepts; concepts, though, are more or less determinate image-signs [*Bildzeichen*] for sensations that occur together and recur frequently, for groups of sensations". Such a picture is entirely in keeping with the development of his view over the preceding 18 months. Nietzsche's mature position is one on which concept-formation is presented as the development of *image*-like signs through repeated experience, which then facilitate recognition. Conspicuous in these passages is the absence of a role for *words* in this process. As Mattia Riccardi has helpfully noted, Nietzsche's remarks in BGE §268 closely resemble David Papineau's discussion of 'perceptual concepts', which are sensory templates, very much like the *Bildzeichen* Nietzsche suggests (see Riccardi, 2018).

---

or symbols (HH "Artists and Authors", §216; eKGWB 1885: 1 [28]), and describes concepts themselves as "image-signs [*Bildzeichen*]" (BGE §268).

Indeed, Papineau in fact posits not only perceptual concepts (the templates themselves), but also perceptually-*derived* concepts (see Papineau, 2006: 118). These can be employed in thought, in the absence of perceptual or imaginative imagery of the thing to which the concept corresponds. For *this* class of concepts, Papineau countenances that language might well be necessary (see Papineau, 2006: 119). This is redolent of the form of Nietzsche's view throughout his life, where templates are employed to structure the content of experience. In Nietzsche's earlier writings, plausibly under Schopenhauer's influence, this structure is considered to be *non*-conceptual: it is only what Papineau identifies as 'perceptually-derived' concepts that count as concepts *proper* (hence Nietzsche's focus on language). Over an extended period of reflection on thinking and concepts, however, Nietzsche's view moves closer to Papineau's, with the templates *themselves*—the image-signs—considered to be (perceptual) concepts.

If this is correct, we have good reasons to resist the idea that the distinction between *conscious*- and *unconscious*-thinking maps onto the conceptual/non-conceptual distinction. As we have seen, the 'ability' reading permits the extension of conceptualization beyond states with explicitly linguistically articulated content, but in doing so fails to secure a sufficiently close bind between concepts and consciousness to support Katsafanas' reading of (iv), and thus his conclusion. The link between concepts and consciousness is tighter on the 'articulation' reading, but the view simply lacks textual warrant. Given Nietzsche's remarks, *pace* Katsafanas, the *unconscious*-thinking we discussed earlier could indeed involve conceptual content. Whatever Nietzsche is getting at in suggesting that *conscious*-thinking alone "*takes place in words*", it is not, I submit, that it alone is conceptually articulated.

3.2

As Katsafanas' account fails to offer a compelling reading of the distinction Nietzsche is making between *conscious*- and *unconscious*-thinking in GS §354, we turn our attention in this section to an alternative interpretation. A number of commentators have argued that the correct way to understand Nietzsche's various remarks on consciousness is as evincing an actualist form of

'higher-order thought' (hereafter HOT) theory.[23] On such HOT views, consciousness of a state or content is explicated in terms of the state/content being the target of an occurrent HOT (see, e.g., Rosenthal, 2005: 21–45). The most extensive discussions of the HOT theory in regard to Nietzsche can be found in the work of Mattia Riccardi, upon which I will focus here. Riccardi reads Nietzsche as a 'pluralist' about consciousness, meaning that multiple forms of 'consciousness' are present in Nietzsche's work. Riccardi discerns "qualitative" consciousness, "perceptual" consciousness, and a "dominant" sense of consciousness (^{D}consciousness) in Nietzsche's writings. It is this third form of consciousness, which he suggests is *self*-consciousness, that is claimed to operate "*according to a HOT model*" (Riccardi, 2015: 6).

Riccardi's argument for the HOT reading starts with Nietzsche's claim that consciousness is "the represented representation, the represented will, the represented feeling (*with which we alone are acquainted)*" (eKGWB 1884: 26 [49]). Far from implausibly, Riccardi takes this as evidence that Nietzsche holds some form of higher-order representation (HOR) account. The task of narrowing the field of (markedly different) HOR views down to the HOT theory brings us back to GS §354. Riccardi draws our attention to the fact that Nietzsche there equates man's need for consciousness with a need "to 'know' what distressed him, to 'know' how he felt, to 'know' what he thought" (GS §354). This, Riccardi claims, suggests that "for a certain mental state M to be ^{D}conscious it is required that M be the content of a HOT to the effect that I am in M" (Riccardi, 2015: 6). As further support, Riccardi highlights several similarities between Nietzsche's remarks and work of David Rosenthal, the most prominent (actualist) HOT theorist (see Riccardi, 2018: 6–7).

So understood, *conscious*-thinking is thinking targeted by a higher-order thought. Riccardi suggests that the comments about thought in GS §354 should be understood broadly, to include

---

[23] See Riccardi (2018; 2015); Doyle (2011). Brian Leiter has also stressed the similarities between Nietzsche's views and the HOT theory ("Katsafanas on Nietzsche on Consciousness", *Brian Leiter's Nietzsche Blog*, May 1st, 2008. Last accessed on 26th October 2017, at http://brianleiternietzsche.blogspot.co.uk/2008/05/katsafanas-on-nietzsche-on.html). For a different comparison between Nietzsche the HOT view, see Katsafanas (2016: 41–44).

all attitudinal states. Nietzsche's position, he claims, "seems to be that *all and only* ^{D}*conscious mental contents are articulated propositionally*" (Riccardi, 2015: 6). The significance of language is thus partly owing to the fact that HOTs require a state to be "ascribed to a self", which requires that one possess linguistic competence sufficient for mastery of a "self-referential pronoun" (Riccardi, 2015: 7). There is much in Riccardi's account that strikes me as persuasive. I agree that Nietzsche is a pluralist about consciousness, and that his remarks on consciousness bear some resemblance to Rosenthal's views. Moreover, the idea that *conscious*-thinking is articulated in a manner different to *unconscious*-thinking clearly comports with my earlier remarks about non-linguistic thought. Nevertheless, I have reservations about whether consciousness (as discussed in GS §354) works on a HOT model, and consequently whether the HOT reading captures what Nietzsche means by *conscious*-thinking taking place in words.

3.2a

Advocates of full-blown actualist HOT theories (like Rosenthal) present the view as a univocal account of consciousness: a mental item is conscious *iff* it is the object of a HOT, whether we are talking about phenomenal consciousness of redness or reflective awareness of one's own attitudes, etc. Given Nietzsche's comments in GS §354, Riccardi sensibly distances his reading from these more ambitious HOT views, and restricts his claim to self-consciousness (Riccardi, 2015: 5–6). On this restricted view, in order "to be conscious of one's present state of knowing *that p*, one needs to have the HOT 'I think' about *that p*" (Riccardi, 2015: 6). If we sketch out a HOT model for propositional attitudes, we start with a first-order attitude towards a proposition, e.g., *knowing*: [(*that p*)]. This is presently unconscious, becoming conscious only when targeted by a HOT. This unconscious higher-order thinking would involve a *new* proposition, i.e., *thinking*: [I KNOW (*that p*)]. Presumably, the HOT reading of Nietzsche must take this to be the stage to which Nietzsche was referring when he wrote of man's need "to 'know' what distressed him, to 'know' how he felt, to 'know' what he thought" (GS §354). Once

this unconscious higher-order thinking takes place, one can then express one's knowing *that p* (for instance) in speech or action.

There are reasons, however, to doubt that this was Nietzsche's view. The first thing to note is that, as above, extant HOT theories posit unconscious, conceptually articulated propositional thoughts at first-order and higher-order levels. If one reads Nietzsche's "*in words*" claim in terms of propositional (Riccardi) or conceptual (Katsafanas) articulation, his position would appear to diverge quite a bit from the particular HOT views to which his remarks have been compared. While putative divergence from standard actualist HOT views is hardly a decisive consideration in reconstructing Nietzsche's views, closer attention raises more pressing concerns. Reading Nietzsche's remarks as suggesting the involvement of higher-order thinking offers us two options. The first is that, on Nietzsche's picture, both first-order and higher-order thinking can be unconscious, with the former rendered conscious through being targeted by the latter. The second option is one on which unconscious first-order states are made conscious by higher-order thinking that is *intrinsically* conscious.

Let's start with the first possible reading. If we focus on Riccardi's suggestion that conscious states alone are propositionally articulated, Nietzsche's view, so understood, would be that higher-order unconscious thinking (non-propositional in structure, but possibly conceptual in content) takes as its object a similarly non-propositional unconscious first order state.[24] The result of this is a linguistically/propositionally articulated conscious thought. So construed, this seems reasonably close to some extant actualist HOT views. So, why might we demur in regard to this reading? In short, there is a genuine worry about the textual basis for attributing this complex and detailed position to Nietzsche. Although he of course writes of the "represented representation" (eKGWB 1884: 26 [49]), and makes several references in this context to Leibniz (GS §354; §357)—who certainly *did* countenance a regress of thoughts (Leibniz, 1768/1981:

[24] Riccardi allows for the possibility of unconscious conceptual content, as I do (see Riccardi, 2018).

118)—it is deeply questionable whether these constitutes a solid basis for ascribing to Nietzsche so intricate a position.

If we look to GS §354, the passage in the published works that is supposed to underpin the comparison between Nietzsche and contemporary HOT theorists, what we find therein considerably underdetermines such an ascription. If we return to Riccardi's discussion, his proposal that "for a certain mental state M to be ^{D}conscious it is required that M be the content of a HOT to the effect that I am in M" (Riccardi, 2015: 6) suggests that it is the higher-order thinking itself that is conscious and propositionally articulated, and 'knowing' how one feels, etc., involves conscious thoughts directed at first-order states: "that I am in M" does, after all, sound linguistically/propositionally structured, and thus plausibly conscious on Riccardi's reading.

On an orthodox actualist HOT view, conscious second-order thinking requires an unconscious third-order thought. Should one wish to bring their "knowing *that p*", for example, to a point of conscious introspection on the actualist HOT theory, it would require further, unconscious thinking along the lines of, *thinking*: [I THINK I KNOW (*that p*)].[25] (On a reading of Nietzsche on which only conscious thinking was propositional, similar conceptual content would have to be structured non-propositionally.) If we look to Nietzsche's writings, he certainly does envisage further cognitive operations in relation to linguistically articulated thinking (as we shall see shortly). There is no suggestion, however, that this further cognition *renders* the *conscious*-thinking conscious. Indeed, the most straightforward and intuitive way to read Nietzsche's comments, I would contend, is as presenting a picture on which linguistically articulated

[25] Cf. Rosenthal: Introspection "is the special case in which that [previously unconscious] HOT is conscious, which happens when a yet higher-order thought occurs—a third-order thought about the second-order thought" (Rosenthal, 2005: 113).

conscious-thinking about one's own mental state is conscious *without* any further HOT making it so.

This is, of course, what I referred to above as the 'second option', on which the higher-order thinking involved in Nietzsche's account is intrinsically conscious. This, however, represents a *significant* divergence from extant HOT positions. Indeed, while Nietzsche's remarks do bear resemblance to some of the things HOT theorists say, I am sceptical about whether this view is either properly or helpfully described as a 'HOT theory'. Let me expand. To start with, there seems to be a distinction that can be made, here, between being conscious of the *content* of a mental state, and being conscious of the *object* or *target* the content 'represents' or is 'about'. In the case of Nietzsche's view, the distinction would be between the content of the *conscious*-thinking, and state(s) or content(s) represented by the contents of the conscious thought.

If Nietzsche does posit intrinsically conscious thoughts, the way in which we are conscious of the content of thoughts such as 'I am in M' is clearly going to be different from the way in which we could be said to be conscious of 'M' itself (which is conscious in virtue of being the object of the higher-order thought).[26] Evidently, only the consciousness of 'M' itself could be described as working "*according to a HOT model*" (Riccardi, 2015: 6). So understood, it would be true to say that for all types of mental state, to be conscious *in that sense* will involve a lower-order mental state being targeted by a higher-order thought. (This includes 'thinking'—conscious thinking would simply be thinking targeted by an intrinsically conscious HOT.) It would be equally true, however, to say that every episode of reflective, higher-order consciousness constitutively involves (and bottoms-out in) a mental state that is conscious without itself being the target of

[26] One should note that "different" here does certainly not mean 'incompatible'—indeed, I believe Nietzsche's position involves both. We are, however, left with the problem of *in what way* the conscious thoughts are conscious: it is not clear, on first sight, which of Riccardi's various senses of 'conscious' found in Nietzsche's work ('qualitative', 'perceptual', or 'self-consciousness') intrinsically conscious thoughts would fall under. (For brief discussion of one possible answer to this, see n. 30.)

a HOT.[27] Accordingly, even on Riccardi's substantially restricted HOT reading, Nietzsche's position denies what could plausibly be considered the central tenet of HOT views (as opposed to views that simply include some role for higher-order thinking), viz. that every conscious state in the domain under question (in this case, 'self-consciousness') is conscious in virtue of being the target of a HOT. Nietzsche's account of reflective consciousness is, in short, as much a first-order view as it is a higher-order one. If this is right, Nietzsche's remark about *conscious*-thinking alone taking place "*in words*" is surely not pointing to his endorsement of a HOT theory.[28]

4.

Nietzsche's picture clearly involves meta-representation; specifically, it involves thoughts about first-order mental states and/or their contents. This alone, however, underdetermines the ascription of a HOT theory. Accordingly, I wish instead to sketch an alternative; one which I hope retains the virtues of the HOT theory as a reading of Nietzsche, without some of its baggage. His position, I content, more closely resembles contemporary work that links consciousness to *inner speech* than it does standard HOT views. The 'inner speech' claim, in short, is that conscious thinking occurs in imaged natural language sentences.[29] (Understandably, theorists advocating such a position focus most commonly upon articulation in *inner* speech, although, strictly speaking, there is no deep distinction for our purposes between thinking to oneself and talking to oneself or others.)

Nietzsche is remarkably consistent in linking consciousness explicitly to language. In addition to the remarks in *The Gay Science* with which we started, he writes in *Daybreak* that "where words

---

[27] Riccardi does not, I believe, properly separate the consciousness of content and object/target, talking sometimes of conscious states as conscious interpretations, and of consciousness of the objects of said interpretations (see Riccardi, unpublished manuscript).

[28] Thanks to Mattia Riccardi and Reid Blackman for comments that helped make this section clearer.

[29] See Ryle (1949: 176); Gauker (1994: 221); Carruthers (1996: 51); Bermúdez (2003: 159–60).

are lacking, we are accustomed to abandon exact observation because exact thinking there becomes painful; indeed, in earlier times one involuntarily concluded that where the realm of words ceased the realm of existence ceased also" (D §115). If we look to the notebooks from the previous year, Nietzsche writes there that the "thought-image consists of *words*" (eKGWB 1880: 6 [361]; cf. eKGWB 1880: 6 [362]). A later passage on consciousness and communication, some of which would surface in GS §354, touches upon the question of how consciousness is "possible". Therein, Nietzsche concedes that he is "far from coming up with answers", offering only, "words, and not more!" (eKGWB 1884: 30 [10]). If we move forward to the period in which he was composing the fifth book of *The Gay Science*, we find Nietzsche insisting that "we *think only* in the form of language": we "*cease to think*", he claims, "*when we refuse to do so under linguistic constraint*; we barely reach the doubt that sees this limitation as a limitation" (eKGWB 1886: 5 [22]).

Such passages, I contend, are most naturally read as linking conscious thinking to sentences articulated internally in natural language. As a reading of Nietzsche, this has several strengths beyond simply fitting well with remarks such as those presented above. First of all, it allows us to take him at his word: *conscious*-thinking "*takes place in words*", and an inner speech reading offers a pleasingly straightforward sense in which that is so.[30] The meta-representational element

[30] Paul Katsafanas has suggested to me that the differences between our views might be merely terminological. Upon reflection, this does not seem to me to be right. First, it is arguably the case that quite a lot will hang on which parts of Nietzsche's picture is 'conceptual'. For instance, one could reply to Katsafanas' suggestion with a version of the claim he offers in response to Mattia Riccardi's work—viz. that how we understand 'concepts' on Nietzsche's view has consequences for how we understand his remarks about falsification (see Katsafanas, 2016: 72–73). Katsafanas is surely right that Nietzsche thinks concepts contribute to falsification, and his suggestions as to how this occurs are insightful. But accounting for falsification purely in terms of conceptualization seems to overlook the close attention Nietzsche pays to other factors, such as explicitly 'grammatical' falsification, for example (see, e.g., BGE §20; §54). A further concern is that the view of concepts (and their role) that Katsafanas ascribes to Nietzsche requires us to ascribe to him a quite egregious error. (While this might not be a decisive reason to reject a reading on its own, there seem to me to be good independent reasons to think that Nietzsche did not in fact make this error.) The mistake I have in mind is the claim that consciousness necessarily falsifies more than the unconscious (see Katsafanas, 2016: 73), where the necessity involved is of a reasonably weighty sort. Katsafanas claims that connecting consciousness to language (as both Riccardi and I do) raises a problem as such views fail to account for this claim. Riccardi's response—that conscious states falsify necessarily in virtue of necessarily employing language—is purportedly insufficient, as Katsafanas claims this would make conscious falsification a merely contingent matter, resulting simply from one's having a less-than-perfect language (see Katsafanas, 2016: 73, n. 26). This claim is entailed by views Katsafanas attributes to Nietzsche: viz. (a) that all and only conscious states are conceptually-articulated (Katsafanas, 2016: 26); and (b) that falsification proper is a matter of conceptualization (Katsafanas, 2016: 48–63). Katsafanas articulates his claim in terms of consciousness being necessarily *more* falsifying than the

of Nietzsche's view is also comfortably accommodated by the inner speech reading; the fact that we can represent our mental states and/or their contents in natural language sentences is clearly beyond doubt. Indeed, Nietzsche appears to take this to be a particular form of a broader phenomenon. He writes in the *Nachlaß* that "the capacity *to will*, *to feel*, *and to think* something of our willing, feeling, thinking *itself*" is summarized "with the word 'consciousness'" (eKGWB 1885: 34 [87]). This passage suggests strongly that our capacity to *think* about our other mental states is, on Nietzsche's view, continuous with other types of meta-psychology, all of which represent forms of 'consciousness'.

If we turn to the published works, we find further support for this reading of Nietzsche's view of consciousness. Starting from GS §354, Nietzsche writes therein that "it is only as a social animal that man has learned to become conscious of himself". This is plausibly a reference not just to his conviction as to the importance of language and communication to the development

---

unconscious as he acknowledges Riccardi's point that there will be a low-level form of falsification that occurs in unconscious states when brute sensation is structured according to sensory templates (see Katsafanas, 2016: 73, n. 25). Once Katsafanas accepts this however, it is unclear how his own account avoids an analogous concern to the one he raises for Riccardi. Indeed, one need not even countenance a *perfect* conceptual scheme. It need only be possible that a conceptual scheme could falsify *no more than* the low-level falsification that occurs in unconscious states for the view Katsafanas ascribes to Nietzsche to be false. It is unclear how Katsafanas could avoid this conclusion without simply availing himself of sort of developmental accounts of language and communication that I (and I assume Riccardi) would use to explain falsification, which hardly seem sufficient for the sense of 'necessity' operative in Katsafanas' rejection of Riccardi's view. That these bottom-out in 'contingent' facts about human psychology, etc., seems to me entirely in line with Nietzsche's broader method. The reading I advance in this paper allows us to diverge from Katsafanas' interpretation on both (a) and (b), and in doing so, avoid having to make Katsafanas' problematic claim.

Further to this, however, it is by no means obvious to me that our views are separated by merely terminological differences—or, at least, one would need to commit to more than I have here for that to be unavoidably the case. One way of fleshing out a view of the sort that Nietzsche advances would be to think of inner speech as a distinct *mode*. Peter Carruthers, for instance, argues that there is no global amodal workspace: in order to become (access) conscious, information has to represented through a sense-modality (see Carruthers, 2011: ch. 3). On views such as this, the role of inner speech is to represent the content of thoughts through a form of mental representation—an auditory equivalent to (visual) mental imagery—that is appropriately sensory. (Indeed, Nietzsche in fact refers to spoken communication as a *Tonzeichensprache* or "auditory sign-language" [HH "Artists and Authors", §216].) As one can utilize imagination to represent and manipulate visual information drawn from memory, one can do the same with the auditory imagery of sounds and words. For our purposes, the point here is that it is not (*contra* Katsafanas) the *structure* of the mental content that makes something conscious/unconscious. It is a feature of the mode that makes its content the kind of thing we can be conscious of. It is just that, in the case of conscious thinking, the mode necessarily requires content that is articulated in language (conceptual content that was not or could not be articulated in language would still be insufficient on this sort of picture).

of reflective consciousness, but also to the purported significance of socially-necessitated self-control. He writes in *Daybreak* that social animals monitor the effects they "exert upon the perceptions of other animals"; in so doing, the social creature, we are told, "learns to look back upon itself, to take itself 'objectively'" (D §26). Social existence requires that one be able to curb instinctive behaviours in order to avoid conflict and reprisal. As Nietzsche makes clear later in the same work, this is achieved through the development of compound affective episodes, in which negative or positive affects attach to first-order motivations either encouraged or prohibited within the society.

If we look to *On the Genealogy of Morality*, Nietzsche offers an account of the "origin of 'bad conscience' [...] a serious illness to which man was forced to succumb by the pressure of the most fundamental of all changes which he experienced", viz. that of finding "himself imprisoned within the confines of society and peace". Under such circumstances, individuals were "reduced to relying on thinking, inference, calculation, and the connection of cause and effect; that is, to relying on their 'consciousness'".[31] Nietzsche suggests that with the directing inwards of those instincts now frustrated in outward expression, the "whole inner world, originally stretched thinly as though between two layers of skin, was expanded and extended itself, and gained depth, breadth, and height in proportion to the degree that the external discharge of man's instincts was *obstructed*". Nietzsche writes of the discharging of one's frustrated instincts internally—the "*internalization* of man", as Nietzsche refers to it—as the event with which "first develops in humans that which man later calls his 'soul'" (GM II, §16).

Nietzsche speculates (not implausibly) that vulnerability and the concomitant necessity of social co-operation required animals to control certain impulses and behaviours within the confines of

[31] Cf. a number of late notes in the *Nachlaß*, where Nietzsche writes, "In all becoming-conscious there is expressed a discomfort of the organism; it has to try something new, nothing is sufficiently adapted for it, there is toil, tension, strain—all this *is* becoming-conscious" (eKGWB 1888: 15 [25]); and that "All perfect acts are unconscious and no longer subject to will; consciousness is the expression of an imperfect and often morbid state in a person. *Personal perfection as conditioned by will*, as *consciousness*, as reasoning with dialectics, is a caricature, a kind of self-contradiction—a degree of consciousness makes perfection *impossible*" (eKGWB 1888: 14 [128]).

a social group, and to become aware of themselves and their effect on others. This is achieved through the development of mental states and processes that take first-order states, etc., as their object. This, I would suggest, is the "animal consciousness" that Nietzsche refers to in GS §354.[32] A significant subset of such meta-psychological episodes comprises instances of linguistically articulated thinking about one's states. This should not, of course, be taken to suggest that there is nothing of particular importance about the advent of language (as GS §354 surely makes clear). Complex communication, going beyond "the look, the expression, the gesture" (GS §354), enables reflective consciousness, which in turn facilitates further cognition. If we look closely, there are significant similarities between Nietzsche's writings and those of theorists that have focussed on the role of public language in cognition. Natural language has been widely implicated in the facilitation of complex instances of meta-cognition, rendering possible what Andy Clark calls "*second-order cognitive dynamics*": once we "formulate a thought in words", Clark writes, "it becomes an object both for ourselves and for others. As an object it is the kind of thing we can have thoughts about". In contrast to the HOT theory, in "creating the object we need have no thoughts about thoughts—but once it is there, the opportunity immediately exists" (Clark, 1996: 177).

Something along these lines seems to be very much what Nietzsche has in mind. For Nietzsche, as for Clark, natural language sentences act as stimuli for further unconscious processing. Nietzsche writes that in conscious thought, "everything expresses something of our total-state in signs", and that "for consciousness, each thought acts namely as a stimulant" (eKGWB 1885: 38 [1]). Thoughts articulated in natural language provoke a further struggle of affects, with one's subsequent thoughts reflecting how the "entire power situation has shifted in the meantime" (eKGWB 1885: 1 [61]). In contrast to the HOT reading of Nietzsche, on which conscious thoughts require the ascription of a particular mental state or content to an 'I', an inner speech

[32] For a similar claim, and an insightful (and more extensive) discussion of some of these passages, see Riccardi (2015).

view need not be so restrictive. Although Nietzsche connects the *development* of reflective consciousness to 'knowing' how one feels and what one thinks, etc., *conscious*-thinking need not concern itself only with mental states. Higher-order, linguistically articulated thinking could comprise broken phrases, for example, and concern features of the external world as much as one's inner life. So understood, the self-ascription of mental states constitutes a part (albeit, for Nietzsche, a significant one) of one's total set of instances of *conscious*-thinking, and said states are only one type of thing of which we can be reflectively conscious.

Crucially, given my remarks in the preceding sections, the inner speech reading also avoids the difficulties of the HOT theory *qua* interpretation of Nietzsche. This is because Nietzsche, so understood, needn't commit to any particular account of what makes linguistically articulated sentences like [I KNOW (*that p*)] conscious. Although he offers few suggestions regarding how non-linguistic thinking is pressed into public language,[33] it is unclear that he would be obliged to. Nietzsche is engaged in speculative theorizing, and inner speech is an undeniable part of mental life. He can posit natural language sentences "silently uttered or entertained in thought" (Bermúdez, 2005: 281) *without* having to posit accompanying iterations of HOTs. If correct, reading Nietzsche as anticipating elements of modern inner speech views provides a means of understanding his various claims regarding consciousness and language which retains the virtues of the HOT view without its complications.

Conclusion

Nietzsche's remarks on thinking are apt to confuse. His notebooks reveal an enduring interest in the topic, and a wealth of interesting (if sometimes underdeveloped) positions. As was not

---

[33] In keeping with much I have said above, Nietzsche's comments suggest that we struggle to articulate the content of unconscious thinking in language. He writes (in a passage entitled "*Thoughts and words*") that one "cannot even reproduce one's thoughts in words entirely" (GS §244); and asks elsewhere, "If there were an 'in itself' at all, what, then, would be the 'in itself' of a *thought*?" (eKGWB 1885: 1 [74]). His view appears to be that conscious linguistic articulation *fixes* the content of a thought (for other cognitive processes to respond to). In *Daybreak*, for instance, he writes that we "always express our thoughts with the words that lie to hand. Or, to express my whole suspicion: we have at any moment only the thought for which we have to hand the words" (D §257).

uncommon for Nietzsche, this lengthy period of often-fruitful reflection surfaces in the published works as handful of greatly condensed and highly suggestive comments, open to a variety of interpretations. Nietzsche's picture of *unconscious*-thinking, I have argued, grows from three distinct contentions: that *unconscious*-thinking utilizes imagistic vehicles; that it involves interpretation and analogical thinking; and that it is partly constituted by a "struggle" of affects. So understood, *unconscious*-thinking does not operate along relations between syntactic features of language-like constructions (as might be suggested by logical reconstructions of the thought-process). I have also argued that in claiming *conscious*-thinking "*takes place in words*", Nietzsche has in mind neither conceptualization, nor something akin to the HOT theory. Instead, his position is closer to contemporary work on inner speech. We become conscious of thinking by representing it in imaged natural language sentences, which become the stimuli for further *unconscious*-thinking.

There is, of course, more to be said about Nietzsche's writings on thought. The above interpretation, if correct, brings Nietzsche's writings into contact with areas of contemporary research (e.g. non-linguistic thought) and bears in interesting and important ways on his views of motivation and agency, epiphenomenalism, eliminativism, and reflective choice, as well as on his thoughts about the role of consciousness generally. These discussions will have to wait for another day. I hope, however, to have provided here an adequate basis from which such topics can be approached.[34]

---

[34] Sincere thanks are owed to the participants at the 3rd ISNS workshop for their helpful comments and suggestions, which I have done my best to address or incorporate. Thanks also to Luke J. Davies for comments on an earlier version of this paper.

The Social Structure of Drives

Abstract: In §12 of *Beyond Good and Evil*, Nietzsche claims that the untenable status of the "atomistic" soul leaves open the path for "refinements" of the hypothesis. Perhaps the most suggestive of the refinements Nietzsche offers is the idea of the soul as a "social structure [*Gesellschaftsbau*] of drives and affects". It is this idea—or, more accurately, the image found therein—that will serve as the primary focus of our attention here. Nietzsche writes on more than one occasion of the self (or mind) as being in some sense like a "social" structure of many "souls", or like an "*oligarchy*" or "aristocracy". So, what significance, if any, should we attach to the idea of this structure being a "social" one? Indeed, what is it for drives and affects to stand in a 'structure'? Over the course of this paper, I will develop answers to these questions. The social metaphor is a rich one that grows out of different sources and develops over Nietzsche's working life, present in both his earliest philosophical efforts and in his mature work. Properly understood, the notion illuminates what Nietzsche means when talking of drives, etc., in structures or hierarchies. In §1, I address the sole extant engagement with this metaphor, in the work of Maudemarie Clark and David Dudrick. I then move to examine the areas of divergence between this reading, and more common causal construals of the idea of a 'rank order' of drives (§2). In the third section, I discuss Nietzsche's reflections on organic unity, with particular reference to his doctoral notes on Kant and teleology (§3). The fourth section examines Nietzsche's discussion of the state in the *Genealogy* (§4). I then use this picture to explore the range of structural relations in which drives can stand to one another (§5).

Keywords: Nietzsche, the self, the origin of the state, drives, Kant, teleology

INTRODUCTION

In a prominent section from the first book of *Beyond Good and Evil*, Nietzsche claims that "*who* [*one*] *is*" is a matter of the "rank order [*Rangordnung*]" in which "the innermost drives of [one's] nature are placed in relation to each other" (BGE §6).[1] Although the passage offers little by way

[1] Nietzsche's published works are cited using the abbreviations listed in the bibliography, followed by Arabic numerals. Where works comprise several volumes, books, or essays (such that sections are not numbered consecutively), either Roman numerals or abbreviated section titles precede Arabic numerals. For example, GM I, §1 is *On the Genealogy of Morality*, Essay I, Section 1; and TI "Errors", §4 is *Twilight of the Idols*, "The Four Great Errors", Section 4. Citations to the *Nachlaß* refer to year, group, and section/fragment number (eKGWB year: group [note]). Citations to Nietzsche's proposed doctoral dissertation are to KGW section,

of elaboration on this remark, a section a bit further on plausibly provides a hint. In BGE §12, Nietzsche addresses the "*atomism of the soul*", an idea of which he insists science must rid itself. The purging of the notion of the soul as something "indestructible, eternal, indivisible", we are told, leaves open the path for "new versions and refinements of the soul-hypothesis". Nietzsche offers up the "mortal soul", "the soul as subject-multiplicity", and the soul as a "social structure [*Gesellschaftsbau*] of drives and affects", as ideas that "want henceforth to have citizens' rights in science".

It is the last of these three "refinements" in particular—or, more precisely, Nietzsche's choice of image therein—that will occupy our attention here. A "social structure" is plausibly the sort of thing he had in mind when talking of the 'rank order' in which one's drives purportedly stand. The image of the self as some sort of community appears on more than one occasion. If we look to BGE §19, Nietzsche likens what occurs in the event of willing to the "social structure [*Gesellschaftsbau*]" of many "souls" acting like a *commonwealth* in which the ruling class identifies with the success of the community. Elsewhere, he writes of the organism as an *oligarchy* (GM II, §1), and, in the *Nachlaß*, of the body as an *aristocracy* (eKGWB 1884: 25 [134]; 1885: 2 [76]; 40 [42]; 37 [4]). The metaphor is certainly a suggestive one, and presents to us a number of questions. What is it, for instance, for drives to stand in a "structure"? And what significance (if any) should we attach to the idea that said structure is, in some sense, *social*? The task of this paper is to draw out answers to these questions. At its most developed, the social metaphor is a rich one. Nietzsche's understanding of a *Gesellschaftsbau*, I contend, developed over the course of his working life, and drew upon a range of different sources. To stand in a 'social structure', we shall see, is to instantiate a set of quite specific features. Once fleshed out, the image reveals much about the nature of the relations Nietzsche envisages between one's constituent drives and affects.

---

volume, page number. Quotations from the published works are from the translations listed in the bibliography (with occasional amendments). Translations from the *Nachlaß* are my own.

This paper has five parts. In §1, I examine the extant discussions of the idea of a 'rank order' of drives, with a particular focus on Maudemarie Clark and David Dudrick's recent idea of a *normative* order. I then move to consider more closely the areas of disagreement between their view, and existing causal readings, suggesting that their interpretation faces serious difficulties (§2). Sections §3–§5 lay out my alternative reading. I first address Nietzsche's interest in purposiveness in nature, looking specifically at the notes he prepared for his (uncompleted) doctoral thesis (§3). In §4, I look closely at the account of the origin of the state in *On the Genealogy of Morality*, which represents Nietzsche's most extensive discussion of social structures. In the final section, §5, I use the account in the *Genealogy* to flesh-out the variety of relations in which drives and affects can stand to one another, presenting a dispositional causal reading of Nietzsche's 'social structure'.

1.

One way of making sense of Nietzsche's remarks is to understand the self as a *causally-determined* structure of drives and affects—i.e. a stable order of strength or power. A good example of such a reading can be found in the work of Brian Leiter. On Leiter's view, Nietzsche's comments about 'who one is' and the *Rangordnung* of drives refer to one's "*type-facts*" (Leiter, 2002: 9). Nietzsche, it is argued, has a "Doctrine of the Types", according to which "Each person has a fixed psycho-physical constitution, which defines him as a particular *type* of person" (Leiter, 1998: 230). Type-facts are physiological and psychological facts which, when taken in conjunction with environmental factors, (causally) explain the attitudes, beliefs, and behaviour of the individual in question. Said facts are necessary, but not sufficient, determinants of the "trajectory of that person's life" (Leiter, 1998: 224).

On Leiter's picture, psycho-physiological facts determine 'who one is'—i.e. the *type* of person one happens to be. Nietzsche's works are replete with what appear to be types; from priests and slaves to masters and nobles, from free spirits, *übermenschen*, and sovereign individuals to Jesuits and the Last Man. Given that 'who one is' corresponds to the rank order of one's "innermost

drives" (BGE §6), the upshot of Leiter-esque readings is that the range of 'types' is (at least partly) a range of structural relationships in which such drives can stand to one another. These structures are a matter of certain drives dominating and others being dominated—a 'social' rank-order determined causally, resultant upon the strength of the constituent drives.

Notable dissent from this view has come in the form of Maudemarie Clark and David Dudrick's *The Soul of Nietzsche's Beyond Good and Evil* (2012). In light of Nietzsche's ready employment of the metaphors of a 'commonwealth', 'oligarchy', or 'aristocracy', Clark and Dudrick suggest that the sort of social structure Nietzsche has in mind is an explicitly *political* one. They argue that the image of a *Gesellschaftsbau*, correctly understood, presents a considerable challenge to the sort of causal picture outlined above. Taking Leiter's reading of BGE §6 as their foil, Clark and Dudrick claim that such accounts fail to appreciate that a social structure of the sort Nietzsche points to is "not just a causal order, but is also a normative one" (Clark and Dudrick, 2012: 150). The significance of this choice of metaphor is that such a hierarchy is purportedly an order of "legitimacy and authority, not simply strength" (Clark and Dudrick, 2012: 188). In explicitly political structures, the "ruler or ruling class" is able to rule "in part, because it is recognized as having the authority to rule, to speak for the whole society" (Clark and Dudrick, 2012: 150).

The significance of this distinction, for Clark and Dudrick, is owing to the putative role it plays in their argument that Nietzsche's account of the self involves an irreducibly normative dimension—one that causal readings simply cannot accommodate. To condense Clark and Dudrick's often fastidious case for this conclusion, a political order of legitimacy is argued to be necessary for Nietzsche to avoid 'losing the soul' (see Clark and Dudrick, 2012: 175). In BGE §12, the passage in which the idea of a *Gesellschaftsbau* is introduced, Nietzsche writes of the "clumsiness of the naturalists", who "barely touch upon 'the soul' without also losing it". Clark and Dudrick parse this remark as the claim that reducing the "soul or psyche to forces—that is, to drives" would be to "debunk the whole idea of the soul" (Clark and Dudrick, 2012:

158), much as Boscovich's reduction of the atom to extensionless centres of force is claimed to have effectively eliminated it from serious science.

The upshot of this is that, whatever the soul is, if Nietzsche wishes to retain it (as he clearly does), it cannot be reducible to a causally-determined structure of drives. At this point, defenders of such views might insist that this amounts to a *reductio* of Clark and Dudrick's interpretation. After all, the various remarks in §6 and §12 of *Beyond Good and Evil* upon which we have focussed seem to equate the self with a structure or rank of drives. But this, Clark and Dudrick insist, is only apparent. Indeed, they argue that close attention to BGE §6 raises a serious problem for anyone looking to make such an equation. Therein, Nietzsche describes "every great philosophy to date" as a "personal *confession* of its author", and claims that the evaluative intentions "in every philosophy" represented "the actual seed of life from which the whole plant had grown in each instance". The philosopher's morality, we are told, provides "decided and decisive *testimony*" as to who they are (BGE §6, emphases added). On Leiter's view, one's morality gives "testimony" since it is caused/explained by the relevant type-facts concerning the structure of one's drives and affects. The problem for such interpretations, Clark and Dudrick contend, is that a "confession" or "testimony" is most naturally understood as *revealing* something (see Clark and Dudrick, 2012: 152). Yet, Nietzsche expresses a rather full-throated scepticism about our capacity to acquire knowledge of the drives (as Leiter notes). If this is right, how could a philosopher's morality bear witness to anything regarding drives?

Clark and Dudrick propose a way around this problem: Instead of one's morality revealing something about the antecedent psychology that causes/explains it, they propose that one's morality *is* 'who one is'. Nietzsche's view, they contend, is that a philosophy "expresses and exhibits one's morality, the structure of one's commitments, and therefore the kind of person one is". The testimony provided by the philosopher's morality thus concerns "not the particular drives, but *who* the person is", i.e. "her values, her commitments", understood at the "level of the person" and expressed in attitudes and actions. *Contra* Leiter, Clark and Dudrick claim that the relationship between such personal-level values and sub-personal psychology is not one of

causation, but *constitution* (Clark and Dudrick, 2012: 152). The morality that testifies decisively to one's being this-or-that type of person is constituted by the normative order of one's drives as observable, medium-sized dry goods are constituted by unobservable "microproperties". Although being a particular type of person will be a matter of having one's drives stand in a certain rank order, a philosophy's making known the structure of one's commitments "does not thereby make known the drives that inspired the philosophy" (Clark and Dudrick, 2012: 152) any more than knowledge of the macro-level properties of my desk entails knowledge of its micro-properties.

So understood, Nietzsche avoids losing the soul by identifying it with something irreducibly situated at the personal-level (viz. the individual's values). This, however, requires a means of "distinguishing values from mere desires and of explaining how values gain their foothold in human motivation" (Clark and Dudrick, 2012: 175). Acting "according to values" requires that we seek out what we judge to be *good*, and not simply act from our strongest impulse (Clark and Dudrick, 2012: 176). Such a distinction, it is claimed, stands beyond the reach of causal readings. In treating "drives or appetites as sufficient to explain human action", such views leave "no role for values in human motivation" (Clark and Dudrick, 2012: 175). Clark and Dudrick argue that the idea of a political order, however, provides just the basis for distinguishing values and desires that Nietzsche requires. In those instances in which one acts from values, the drives with authority effectively pull rank. In the "well-structured, happy commonwealth" (BGE §19), a recalcitrant drive is able to recognize the legitimacy of those higher in standing and will fall into line accordingly, even if stronger than the commanding instincts.

The natural question, at this point, is how something like a drive could be said to 'recognize' anything, let alone things like authority or legitimacy. If we return to §6 of *Beyond Good and Evil*, Nietzsche's remarks suggest a possible answer. He writes that "all of the fundamental human drives have done philosophy, and that each one of them would like only too well to represent itself as the ultimate purpose of existence and legitimate *master* [*berechtigten* H e r r n] of all the other drives". Each, he insists, is "desirous of ruling: and it is as *such* that it tries to philosophize"

(BGE §6). Although peculiar, such language is far from anomalous. In a number of passages in which he discusses drives, Nietzsche talks as if they are *homunculi*—i.e. proto-persons engaging in activity as agents *proper*.[2] He writes in *Daybreak*, for instance, of a drive that "wants to have its victim" (D §140), and on several occasions of drives "commanding" and "obeying" (BGE §19; cf. 1885: 36 [22]). Elsewhere, he suggests that drives "*interpret the world*", and (in a line redolent of BGE §6) that each drive "has its perspective that it would like to force all the other drives to accept as a norm" (eKGWB 1886: 7 [60]).

Clark and Dudrick contend that this language is no mere metaphor: Nietzsche talks of drives as if they were homunculi precisely because they *are* homunculi. He is claimed to be engaged in what Daniel Dennett called a "*top-down*" approach to psychological explanation—i.e., one that looks to explain complex behaviour at higher levels of psychological functioning by analysing them in terms of "more and more detailed smaller systems or processes" (Dennett, 1978: 110). All that is required of this strategy is that the drives are "*simpler* than the activities of the person that they are [quoting Dennett] 'rung in to explain'" (Clark and Dudrick, 2012: 198), and Clark and Dudrick are confident that, in this respect, homuncular drives fit the bill. The drives' "agency of a sort" consists simply in *commanding* and *obeying* (Clark and Dudrick, 2012: 196). These, it is claimed, are simpler than the psychological phenomena Nietzsche looks to account for, while still being sufficiently complex to capture the 'political' nature of the drives' behaviour.

So understood, Nietzsche's references to various social structures are intended to communicate to his most attentive and perspicacious readers a number of positions and philosophical concerns hidden beneath the surface meanings of the text.[3] This imagery, we are told, conveys

---

[2] Cf. Peter Poellner's remark that Nietzsche "treats drives not as attributes of agents (like desires), but as agents themselves" (Poellner, 1995: 174).

[3] Those commentators who claim there to be an esoteric meaning to *Beyond Good and Evil* beneath its exterior offer subtly different explanations as to *why* Nietzsche would employ such a method. Laurence Lampert suggests that Nietzsche knew that the surface text would appeal to his (presumably future) readers, and that this would offer a way in, through which the actual argument of BGE could appear as readers views were changed by the content of the work (see Lampert, 2001: 1). Clark and Dudrick, by comparison, take the challenge of extracting the esoteric (normatively-focussed) thrust of BGE from the "crudely naturalistic"

that drives' interactions are not merely matters of strength, but are characterized by normative relations. In the wider context of the first book of *Beyond Good and Evil*, the *Gesellschaftsbau* is also supposed to underpin the irreducible normativity of the self, and thus Nietzsche's pervasive interest in normative questions. The Nietzsche of *Beyond Good and Evil*, we are told, by no means turns his back on the "normative aspirations of traditional philosophy" (Clark and Dudrick, 2012: 9). Indeed, Clark and Dudrick suggest that he intends to establish himself as "Kant's true heir" (Clark and Dudrick, 2012: 85) and "carry out Kant's normative project" (Clark and Dudrick, 2012: 75). If this is right, Clark and Dudrick's reading would apply considerable pressure to existing (predominantly dispositional) accounts of the drives, and to standard causal interpretations of Nietzsche's remarks. Such views risk not only overlooking the irreducibly normative dimension to his psychology, but also mischaracterizing much of his broader project.

2.

So, how compelling should we find this interpretation of Nietzsche's political metaphors? As my remarks thus far suggest, there seem, plausibly, to be two central areas of disagreement between the normative reading and standard causal accounts, viz. the specific nature of the drives' interactions, and the issue of what Nietzsche refers to in talking of 'who one is'. Since it is the claim that one *is* one's morality that provides the central argumentative thrust in favour of the normative reading, it is to this that we shall turn our attention first.

2.a

It is perhaps not immediately clear what it means to say that you *are* your morality. As an example of the sort of thing they mean by philosopher's 'values', 'commitments', and so on, Clark and Dudrick offer "the relative strength of her commitment to truth" (Clark and Dudrick,

---

exoteric text (Clark and Dudrick, 2012: 9) to be motivated by the intention of teaching future philosophers *how* to read his works well. (Given the infrequency of esoteric readings of BGE, one might be tempted to conclude that Nietzsche adopted what has proven—so far—to be an impressively ineffective method of achieving either of these aims.)

2012: 152). This suggests something like an evaluative orientation to the world, e.g. that of *valuing* truth, *valuing* creativity, etc. So construed, one's 'morality' would be an enduring evaluative psychology, manifest in patterns of occurrent valuing, motivation, and the seeing of things in particular (evaluative) lights. If we read BGE §6 with this in mind, Nietzsche's view would be that the theoretical and evaluative commitments of "every great philosophy" are underpinned by a stable evaluative orientation—a standpoint from which some things are experienced as salient, desirable, admirable, etc., and others unpleasant, dull, or repugnant, and so on—and that this enduring, personal-level psychological orientation is constituted by drives standing in a certain rank-order.[4]

As we saw earlier, in order to retain an intuitive sense in which one's morality could offer a testimony, Clark and Dudrick read Nietzsche as picking out this enduring orientation when speaking of 'who one is'. An initial concern is whether shifting attention away from the structure of drives to a personal-level psychological orientation does in fact offer a satisfying sense of testifying to something. Clark and Dudrick do not consistently separate Nietzsche's remarks about each great *philosophy* being a confession, from those about the philosopher's *morality* providing testimony as to who they are. The distinction is made neatly when recounting Leiter's view, on which a philosophy reveals one's morality (the 'seed' from which it grew), and one's morality, in turn, testifies regarding 'who one is' (the structure of one's innermost drives). Clark and Dudrick's description of their own account further down the same page, however, claims that "a philosophy and a morality" bears witness to one's values and commitments (Clark and Dudrick, 2012: 152). This conflation of philosophy and morality masks the fact that, although a philosophy might intuitively be thought to 'confess' something of one's underlying evaluative attitudes, it is far less intuitive to talk of a thing (e.g. a morality) providing decisive testimony/evidence (*Zeugniss*) as to itself (i.e. 'who one is').[5] Indeed, it is hardly obvious that

[4] This fits neatly with remarks Clark and Dudrick make elsewhere about value being "recognizable in the world only from a *perspective* constituted by affective dispositions" (Clark and Dudrick, 2012: 128).

[5] Although I think most people would be happy to say, somewhat casually, that valuing certain things is evidence of your being a certain sort of person (cf. Clark and Dudrick, 2012: 152), this is at best only

this is any more common-sensical than claiming that one's morality offers a testimony from which we are unable to derive any knowledge.[6] If Clark and Dudrick are right in their presentation of the options before us, Nietzsche's choice of metaphor at the end of BGE §6 seems infelicitous, to say the least.

There is reason, however, to think that they are mistaken. For even if we were to allow that one's having particular values and commitments could be said to be 'evidence' for one's being a person with those values, etc., there are significant problems with the argument Clark and Dudrick adduce to motivate their reading. If we recall, they argue that Nietzsche could not possibly intend to suggest that morality reveals something about the drives, owing to his scepticism about self-knowledge more broadly and drive-knowledge in particular. At first glance, this seems to stand on solid textual ground. Nietzsche is clearly deeply critical of our capacity for self-knowledge (see, e.g., HH "Man Alone", §491; D §115; §116; §119; GS §335; §354), and of the very desirability of our achieving it anyway (see EHα "Clever", §9). Yet, Clark and Dudrick's argument relies upon Nietzsche adopting a scepticism on which drive-knowledge *simpliciter* is beyond our reach. If we look closely at Nietzsche's remarks on the topic, however, his views appear to be somewhat more complex than this, and suggest a number of important distinctions and qualifications.

To start with, it matters *why* Nietzsche thinks we routinely, perhaps even unavoidably, fail in our attempts to achieve knowledge of the drives. One factor is, undoubtedly, that he takes the majority of our mental lives to unfold away from the scrutiny of conscious attention; with many of these processes being such that we could not become reflectively aware of them, even if we

---

superficially plausible in the context of Clark and Dudrick's view. The claim that one's morality *makes* one a certain sort of person (see Clark and Dudrick, 2012: 151) naturally prompts the question, "well, what sort of person is that?"—to which the answer can only be, "the sort of person that has that morality".

[6] It is perhaps worth noting, in relation to this suggestion, that Nietzsche says of himself at the beginning of *Ecce Homo* that he had "not neglected to 'bear witness' about [him]self", but that he had been "neither heard, not even so much as seen", owing to his task being so much greater than those of his contemporaries (EHα "Foreword", §1).

wished to (see, e.g., D §129). Another concerns the putative extent to which the folk-psychological and linguistic frameworks we inherit (and that philosophy draws upon) are inadequate and/or error-strewn, which then distorts subsequent attempts to think about the domains in question (see, e.g., GS §354). A third concern, stated memorably at the opening of the *Genealogy*, is that we have purportedly "never sought ourselves" (GM P, §1). The important thing to note is that none of these worries get us to the all-encompassing scepticism about drive-knowledge that Clark and Dudrick require. Such criticisms raise questions about our capacities for introspection, the concepts and categories with which we think about psychology, and the appropriate means through which to understand ourselves. These combined, however, by no means occlude all forms of drive-knowledge.

Although several of Nietzsche's sceptical remarks are ambiguous regarding the particular sort of knowledge he has in his sights,[7] he clearly reserves special scorn for *self*-knowledge. He writes in *Daybreak*, for example, that however "far one might like to go in self-knowledge, nothing, to the contrary, can be more incomplete than the image of the total *drives* that constitute his being" (D §119). Putting aside the fact that Nietzsche here addresses knowledge of the *total* (*gesammten*) drives one comprises, comments of this sort would clearly not rule out third-personal claims about drive-psychology. Indeed, the nature of Nietzsche's apprehensions regarding human self-understanding leave reasonably untouched the possibility that a sufficiently insightful psychologist, armed with a preferable framework of concepts and an appreciation of the limits of psychological speculation, might well make some true claims about human psychology—even about drives.[8] This should frankly come as no surprise. Nietzsche's sceptical remarks seem often intended to provoke us into examining our concepts, questioning our assumptions, and

---

[7] Nietzsche's remarks, e.g., that what "men have found so hard to understand is their ignorance of themselves!" (D §116), could be meant to convey either that humanity's understanding of itself (i.e. its psychology, etc.) is inadequate, or that an individual's understanding of themselves is inadequate.

[8] This is of course compatible with Nietzsche's believing that *all* psychological terminology is a kind of *Bilderrede* that should not be reified (see D §119; eKGWB 1881: 11 [128]; HH "Assorted Opinions", §5). Nietzsche's view seems plausibly to be simply that claims to knowledge should be made with this qualification (see eKGWB 1886: 7 [60]).

'seeking' ourselves, and his preferred alternative accounts are commonly third-personal in nature.

But even this refinement of Nietzsche's putative scepticism about drive-knowledge fails to get it quite right. For instance, he does in fact make some claims to self-knowledge in *Ecce Homo* (see, e.g., EHα "Clever", §8; §9), which suggests the possibility of speculative, *retrospective* knowledge about one's own psychology, acquired through skilled reflection on extended patterns of past behaviour, conscious thought, valuing, and so forth (as one would in the case of another person). He also seems open to utilizing certain sorts of introspectively acquired phenomenal, interoceptive, and proprioceptive evidence, as we see in BGE §19 for example.[9] So what is Nietzsche getting at, then, with his repeated dismissals of our capacities for self-knowledge? Upon closer inspection, Nietzsche's scepticism in fact appears to concern two different, rather specific things. The first is the idea of knowledge of one's drives, motives, etc., acquired through something like *inner sense*. This he takes simply to be a nonsense, spoken of only by those that have failed to appreciate the interpretive nature of all self-knowledge.[10] His second target is knowledge acquired *solely* through introspection. One thought here (as mentioned earlier) is that the contents of consciousness are purportedly a mere sliver of our mental life. Nietzsche seems also to think, however, that those things introspectively accessible

[9] He writes there that "in all willing, there is first a plurality of sensations, namely the sensation of the condition '*away from which* we go', the sensation of the condition '*towards which* we go', the sensation of this '*from*" and '*towards*' itself, and then besides, an accompanying muscular sensation, which, even without our putting in motion '*arms and legs*', commences its action by force of habit". He refers to this approach as being "more cautious for once […] 'unphilosophical'" (BGE §19). Plausibly this reflects Nietzsche's concern, voiced in several notes from the period in which *Beyond Good and Evil* was being written, regarding the "danger of the direct questioning of the subject *about* the subject and of all self-reflection of the spirit", which lies in the fact that "it could be useful and important for one's activity to interpret oneself *falsely*". The answer to this, he suggests, is to "question the body and reject the evidence of the sharpened senses […] to see whether the inferior parts themselves cannot enter into communication with us" (eKGWB 1885: 40 [21]; cf. 1886: 5 [56]).

[10] Regarding "the famous 'inner sense'" and introspective inner certainties, see eKGWB 1888: 14 [152]; 1885 38 [3]; BGE §16; TI "Errors", §3. For a remarkably similar picture of self-knowledge—in its broad strokes, at least—see Carruthers (2011).

in a given instance radically underdetermine a single, correct interpretation.[11] (Indeed, this particular concern seems to extend also to much third-personal psychology.)

This picture leaves open the possibility of something like a morality bearing witness to those facts about the drives than can be inferred from the conjunction of relevant historical information and psychological and/or physiological 'signs', contextualized in broader patterns of behaviour, thought, and valuing. While establishing the identity of particular drives operative in given episodes seems an unlikely result for such a method, it might be possible to discern the influence of certain drives in extended periods of human activity—such as, say, an "instinct of cruelty" in various social practices (see EHα "Excellent Books", "The Genealogy"; GM II, §6). It is also by no means clear that the sorts of problems that Nietzsche thinks arise in relation to knowledge of drive identity in given episodes would beset speculative theorizing about drive *structure*. Indeed, this is precisely the sort of thing we find in much of Nietzsche's psychology. His remarks on great individuals, for example, stress the diversity of their capacities, and the degree of stability and integration among their drives, expressed in enduring strength of will and feeling, as well as self-affirmation (see, e.g., EHα "Clever", §9; TI "Expedition", §49; cf. eKGWB 1888: 14 [157]).[12]

---

[11] See BGE §32: "intention is only a sign and symptom, something which still needs interpretation, and furthermore a sign which carries too many meanings, and thus, by itself alone, means almost nothing" (cf. eKGWB 1884: 26 [92]; 1885: 38 [1]). If we look to some of Nietzsche's various sceptical remarks in *Daybreak*—often held up as the decisive evidence of his scepticism about drive knowledge—they are comfortably compatible with the construal of his views here offered (see D §115; §116; §119; §129)

[12] If we consider some of Nietzsche's discussions of morality in *Twilight of the Idols*, we encounter a similar focus. He writes of *décadence* morality as reflecting a "Disgregation of the instincts!" (TI "Expeditions", §35), and of decadent moderns as possessing drives that "contradict, disturb, and destroy one another"—a form of "physiological self-contradiction"—which requires that "at least one of these instinct-systems should be *paralysed* beneath an iron pressure" to produce an individual (TI "Expeditions", §41; cf. CW 2nd Postscript). Consider also the contrast we find in the presentation of Goethe and Schopenhauer. The former is described as a "realist" that "*affirmed* everything related to him in this respect", and who allowed himself "the whole compass and wealth of naturalness" (TI "Expeditions", §49, emphasis added). By contrast, Schopenhauer, described as "for a psychologist a case of the first order", is claimed to have engaged in a "nihilistic total devaluation of life" (TI "Expeditions", §21), and is presented as wearied by the vigour of his own instincts (see, e.g., TI "Expeditions", §22). If we look what Nietzsche says about such cases in *Beyond Good and Evil*, he focuses upon structural relations between drives. Nietzsche writes of individuals with "conflicting—and not merely conflicting—drives and value standards that fight with each other and rarely leave each other in peace" that their "most basic desire is for an end to that war that he *is* […] a notion of happiness as primarily rest, lack of disruption, satiety, unity at last" (BGE §200). Plausibly, Schopenhauer's evaluative views—his pessimism

If we return to BGE §6, these are clearly significant. Even if we were to agree that a 'confession' or 'testimony' had to present something from which we could derive knowledge, it is far from obvious that Nietzsche thinks moralities need to confess something *to their bearers*. Nothing in the passage commits Nietzsche to the idea that whatever is conveyed by the morality must be accessible via inner sense or derivable from single episodes; nor is there anything that suggests that 'who one is' a matter exclusively—or even primarily—of identifying *particular* drives.[13] Accordingly, there is no reason to think that BGE §6 poses a problem for the causal reading of Nietzsche's *Gesellschaftsbau*. Such views can indeed provide an account of how moral or philosophical positions might be said to betray something of one's psychology—even if perhaps only to a "psychologist that has no equal" (EHα "Excellent Books", §5).

If correct, this undercuts much of Clark and Dudrick's motivation for insisting that Nietzsche must locate the soul at the personal level to avoid it meeting the same fate as the atom. They could of course continue to claim that their reading of BGE §12 is the correct one, but Nietzsche's remarks on Boscovich are far from transparent. Absent the problem raised by BGE §6, defenders of the causal reading would seem to be on solid ground in responding that Nietzsche's writings plausibly have the 'exoteric', surface meaning they do because—in this case at least—that is what he is genuinely trying to say.

2.b

This brings us to the second area of disagreement between the causal and normative readings. Clark and Dudrick's case for reading the *Gesellschaftsbau* as integral to a broader argument regarding the self and personal-level valuing clearly falls short. Nevertheless, it might still be

---

and valorizing of abnegation—reflect this struggle in the suppression of recalcitrant instincts. As Nietzsche claims in the *Nachlaß*, "*Conscience* is the feeling in which the rank order [*Rangordnung*] of our drives comes to our consciousness" (eKGWB 1883: 15 [51]).

[13] Nietzsche makes this particularly explicit later in *Beyond Good and Evil*, writing that "The valuations of a man betray something of the *structure* [A u f b a u] of his soul" (BGE §268).

true that Nietzsche intended such language to convey something about the nature of the drives' relation(s) to one another—viz. that the order he has in mind is normative, and that the drives' resultant interactions are characterized by commanding and obeying. As we have seen, a normative order is purportedly a hierarchy of "legitimacy and authority, not simply strength" (Clark and Dudrick, 2012: 188). Although Clark and Dudrick do not really approach the question of what this means head on, they do connect the idea of a political order to Nietzsche's remarks on the formation of the state. The initial story is thus plausibly very close to the one advocates of the causal reading would advance. As the structured relations between individuals develop through "brutish, forcible interaction" (Clark and Dudrick, 2012: 203), the political order among drives comes about because of each instinct's will to power—it's lust to rule over the other drives—which results in the strong subjecting weak (see Clark and Dudrick, 2012: 206; 209).

On Clark and Dudrick's reading, however, this "brutish, forcible interaction" results in behaviour that is "no longer simply brute and forced", but has "become political" (Clark and Dudrick, 2012: 203). The dominant activity of the strong generates a social order, into which the weak must then integrate themselves if they wish to dominate others and/or benefit in some way from the power accrued to those above them. In a discussion of military hierarchies, Clark and Dudrick make clear what this integration involves: to be "part of such an organization", they write, "one must accept the *authority* of those above one in the military hierarchy, which means acknowledging their right to give orders, and obeying these orders" (Clark and Dudrick, 2012: 202).[14] In exchange for accepting the legitimacy of the strongest drives' dominance, and obeying when commanded, subordinate drives acquire a position in the community that affords

[14] This interpretation of their view is supported by their claim, elsewhere, that it is a "conceptual necessity" that commanded drives usually obey: "If the commands of the politically high-ranking drives *were not* usually obeyed", they claim, "these drives would not be politically high-ranking" (Clark and Dudrick, 2012: 194). If Clark and Dudrick are to avoid the accusation of having conflated 'being legitimate' with 'being seen to be so', or of confusing *authority* with *power*, then the conception of legitimacy they have in mind must be connected to consent or agreement.

them the opportunity to highlight "certain aspects of reality, ones that would engender in the person a commitment to the pursuit of the drive's object" (Clark and Dudrick, 2012: 209).

This is an intriguing suggestion. Nietzsche certainly thinks that human social arrangements involve jostling for positions of relative power within existing structures. He writes in *Thus Spake Zarathustra*, for example, that "even in the will of those that serve I found the will to be master […] the lesser gives itself to the greater, that it may have pleasure and power over the least" (Z "Of Self-Overcoming"). Nevertheless, Clark and Dudrick's reading faces some challenges. To start with, their position, I shall argue, is philosophically problematic. This is, of course, by no means an objection in itself to ascribing the view to Nietzsche. But such an ascription would require compelling textual support, if simpler, well-supported, and less problematic reading are to hand. (In the coming sections, I am going to be arguing that there is indeed a less problematic view available that is at least as plausible as Clark and Dudrick's.) As will come out in the course of the discussion, the homuncular account they attribute to Nietzsche also appears to sit rather uncomfortably with several of his other views.

To see the difficulties with Clark and Dudrick's reading, we need first to address the standard criticism of homuncular accounts. As they themselves correctly note, such views are commonly thought to confront the horns of a dilemma. In short, if drives are complex enough to possess the capacities attributed to them (e.g. those required to command or obey), do they have any explanatory power? If they are simple enough to explain in the manner required, however, can we accurately characterize their behaviour as, say, commanding and obeying, in a *non-metaphorical* sense? This is often taken to be a forceful, perhaps even decisive, critique of homuncular views. Clark and Dudrick contend, however, that in relation to their account, these criticisms miss their target.

Regarding the first horn of the dilemma, they argue that, in contrast to what occurs at the level of the person, their account "does not have *the drives* overcoming resistance and acting on values—it has them simply commanding and obeying" (Clark and Dudrick, 2012: 198).

Accordingly, we are told, the behaviour of the drives *is* simpler than the personal-level events they are brought in to explain. In response to the second horn of the dilemma, Clark and Dudrick point to wolf packs and chimpanzee societies as examples of groups that can be helpfully described as engaging in political activity,[15] declaring that "In saying that the drives form a political order [...] Nietzsche need not take them to be *conscious* of their political situation—he need not take them to be conscious *at all*. He must only think that their behaviour is illuminated by interpreting it in political terms" (Clark and Dudrick, 2012: 199). Nietzsche's homunculi, it is claimed, are thus simple enough, but not too simple. The horns of the dilemma can be avoided.

This is a novel response to a long-standing problem for homuncular accounts. Yet, I am not persuaded that it is wholly successful. Let's start with the first horn of the dilemma—viz. that homunculi end up being too complex to properly explain anything. As it stands, this is a little vague. Simply pointing out that an explanation is made in terms of resources more complex than the thing explained is not sufficient to demonstrate that it is problematic.[16] In acceptable cases of explanation by reference to more complex entities, however, the entities themselves are drawn from existing frameworks that are independently motivated (and, ideally, well-understood). The requirement that homunculi be simpler than the *explananda* plausibly arises since they are *not* borrowed from some independent, existing set of explanatory resources. The

[15] Nietzsche was familiar with work on complex animal societies. Mattia Riccardi has suggested, for instance, that Nietzsche's understanding of social structures might we have been influenced by Alfred Espinas' *Des Sociétés Animales* (Riccardi, 2017: 32).

[16] This is true of 'horizontal' explanation, where the explanandum is accounted for by reference to an antecedent cause (perhaps in conjunction with certain background conditions), and 'vertical' explanation, which involves accounting for something by (re)describing it, often in terms of a different framework. Someone's playing a note-perfect rendition of the Emperor Concerto might explain a nearby bird flying off, despite people being more complex than birds, and playing the piano being more complex than startled flapping (horizontal explanation). It is also fine to account for simple group behaviour be reference to the complex individuals that constitute the group (vertical). Homuncular explanations generally take the form of vertical explanations. I owe the 'horizontal'/'vertical' terminology to Martin Davies.

motivation for positing homunculi is, instead, that they offer a means of breaking complex phenomena down into something more tractable.[17]

There are different ways in which a sub-personal system can be simpler than the thing it is brought in to explain. Along the first dimension of complexity, a system is simpler than a target phenomenon if it has *fewer* functions/capacities. Along the second, a system is simpler than another if its distinct functions, etc., are less *sophisticated* than those of the thing it is brought in to explain. By the first standard, a drive with one complex functional capacity would be simpler than a drive with multiple basic capacities. By the standard of the second, the drive with a multitude of very basic functional abilities would be considered simpler. For each of these dimensions, there is a form of macro-level complexity that a correspondingly simple homunculus would be well-placed to explain. If one is analysing the *copiousness* of capacities at the macro-level, an acceptable homuncular explanation would account for this by positing numerically simple systems at the micro-level. If attempting to account for the *sophistication* of macro-level capacities, an appropriate homuncular explanation would posit less functionally complex sub-personal systems (even if multitudinous in the number of capacities).

This concern raised by the first horn of the dilemma is thus properly understood as being, at least in part, about whether complex homunculi are simpler than the target phenomenon *in the appropriate way*. There seems to be something deeply unsatisfying about positing copious functions to explain copiousness, or functional sophistication to explain sophisticated functioning. If all I am looking to account for is something's mere *presence* at the macro-level, this sort of explanation can, sometimes, be permissible. We might, for instance, explain something's colour at the macro-level by noting the colour of its constituent parts. If, however, we are looking to explain not something's presence, but its very *possibility*—how it is that this

[17] In Dennett's original discussion of homuncular explanation, for instance, homunculi represent a bridge (or part of a bridge) to sub-personal entities sufficiently simple as to be accounted for causally (see Dennett, 1978: 80–81). Plausibly, it is important to Clark and Dudrick's view that homuncular drives *not* play this role, for risk of 'losing' the soul. This raises serious problems regarding the independent plausibility of the picture they ascribe to Nietzsche.

phenomenon could occur at all—positing it at the micro-level is explanatorily otiose.[18] Nietzsche was, of course, greatly sensitive to this. His complaint in *Beyond Good and Evil* about Kant and the "young theologians of the Tübinger Stift" is precisely that their enthusiastic positing of faculties was less an explanation than a "repetition of the question" (BGE §11). The first horn of the dilemma posed by the standard criticism applies to homuncular accounts with precisely this failing, the most clear-cut and egregious cases of which will of course be those where *exactly the same* functional capacity is posited at the sub-personal level as is to be explained at the personal-level.

If we examine Clark and Dudrick's account, it seems to run into difficulty along both dimensions of complexity. Let's take them in turn. Since Nietzsche utilizes drives to explain a wide range of psychological phenomena, homunculi that only command and obey would offer an unobjectionable homuncular explanation of the *copiousness* of personal-level capacities. The suggestion, however, that the drives do nothing but command and obey doesn't appear to be an adequate characterization of Clark and Dudrick's own remarks, let alone Nietzsche's. In explicating given phenomena, they routinely appeal to a far wider range of capacities than just commanding and obeying. Consider, for example, Nietzsche's contention in BGE §6 that each drive seeks to present itself as the "legitimate *master*" of the others, which Clark and Dudrick take as support for their normative reading (see Clark and Dudrick, 2012: 150–51). It is to this end, Nietzsche suggests, that each of the fundamental drives have practised philosophy. Clark and Dudrick parse this as the claim that a given drive gains control of both the cognitive apparatus and the drive to knowledge, "systematically develop[ing] and defend[ing] *an account* of reality from its point of view" (Clark and Dudrick, 2012: 146). Not only are these further capacities (in addition to commanding and obeying), but they are also plausibly complex

[18] Cf. Clark and Dudrick, who, in laying out the standard criticism, state that explanation by reference to drives that "engage in the activities characteristic of persons […] is not an *explanation* in any recognizable sense" (Clark and Dudrick, 2012: 197).

abilities that require the possession of a whole range of smaller skills/habits.[19] This also highlights that the agential language employed by Nietzsche is *far* richer than merely commanding and obeying, even if we attend only to the first book of *Beyond Good and Evil* (as Clark and Dudrick do).

It is also clear, if we turn to the second dimension of complexity, that remarks of the sort Nietzsche makes in BGE §6 cause trouble here too. Even if we were to accept that drives that develop and defend systematic accounts of reality, and reflect upon the persuasiveness of these accounts, were simpler *vis-à-vis* the number of capacities possessed, it is far from obvious that these are simpler than, say, willing in terms of *sophistication*. Indeed, even if commanding and obeying exhausted the drives' repertoire, Clark and Dudrick's position would still plausibly flirt with the most egregious form of homunculus fallacy. Contrary to their claim that, on their account, the drives are not "overcoming resistance and acting on values" (Clark and Dudrick, 2012: 198), this seems to be precisely what their conception of obeying involves. Subordinate drives in the well-structured society acknowledge normative considerations regarding legitimacy and rank, and desist—even when they are the *stronger* force—despite possessing their own will to power.

What is required to avoid these issues is, of course, a genuine sense of commanding/obeying that is less sophisticated than things like willing. This brings us to the second horn of the dilemma—viz., the concern that, *if* homunculi are simple enough to explain the target phenomena, the agential descriptions employed will be little more than metaphor. Pre-empting this worry, Clark and Dudrick suggest that there is a perfectly good, non-metaphorical sense of commanding and obeying that we find in the behaviour of creatures such as wolves and chimpanzees. Such animals' interactions can be helpfully understood as 'political' in virtue of

[19] Cf. Nietzsche's claim in *The Gay Science* that the development of scientific thought first required the cultivation of "the doubting drive, the denying drive, the waiting drive, the collecting drive, the dissolving drive", which eventually come to "feel like functions of one organizing force in one human being" (GS §113).

their societies being structured into clear hierarchies. They are also plausibly simpler than human beings along both of the dimensions laid out earlier.

I have, however, a number of reservations about this. First, Clark and Dudrick borrow the notion of a "space of reasons" from Wilfrid Sellars when arguing for an irreducibly normative element to Nietzsche's psychology. Sellars, however, understood this as a matter of "justifying and being able to justify what one says" (Sellars, 1956: 294)—of integrating a mental state/content into a network of articulatable normative relations. Yet, this commonly involves reason-responsiveness and reasons discourse in quite a robust sense, requiring the representation of reasons *qua* reasons. If commanding and obeying take place within the "space of reasons" in the sense its advocates generally mean, the application of those terms to animal behaviour will still be metaphorical. On a genuinely Sellarsian picture, wolves and chimpanzees, while certainly some of the more complex inhabitants of the 'space of causes', reside there all the same.

Nietzsche's remarks in a number of places suggest a picture quite sympathetic to this suggestion. In a section entitled "*Animals and Morality*", he claims that the "practices demanded in polite society", including "the setting aside of one's virtues as well as one's most vehement desires, presenting oneself as equal, falling into line, diminishing oneself", can all be found "*in a crude form* everywhere, even in the depths of the animal world" (emphasis added). When we look more closely at what Nietzsche says about these behaviours, he suggests that their purpose is "to elude one's pursuers and be favoured in the pursuit of one's prey. For this reason, animals learn to master themselves […] a consequence of each drive that teaches us to seek food and elude enemies" (D §26). These comments bear resemblance to discussions, in *Daybreak* and elsewhere, in which Nietzsche explains the curbing of instinct and desire in the context of social existence by reference to the development of further affective dispositions, resulting in complex compound affective episodes comprising a first-order drive and an acquired "depressive" affect (TI "Expeditions", §45; cf. D §38). So understood, animals' morality "in a crude form" (D §26)

looks an awful lot more like affective conditioning and complex learning than the recognition of legitimacy.[20]

Upon reflection, these remarks suggest two further problems for Clark and Dudrick. First, although chimpanzees, etc., might be simpler than human beings regarding both the number and sophistication of their capacities, they are still remarkably complex things, possessing their own drives, habits, affects, perceptions, imagination, and so forth. To return to D §26, for example, Nietzsche therein suggests that the social animal "assesses the effect it produces upon the perceptions of other animals and from this learns to look back upon itself, to take itself 'objectively'". While such processing might not require the uniquely human cognitive capacities (that Nietzsche so often derides), it certainly would seem to require the intellect, of which drives, on Clark and Dudrick's reading, are wholly independent.

Second, were they to resist this, and to maintain *either* that Chimpanzees (*sans* human cognitive capacities) or drives (*sans* personal-level perceptions, imagination, affects, etc.) were capable of commanding/obeying, it becomes unclear why we should stop there. If neither the recognition of normative considerations, nor the corresponding self-regulation, require these capacities, why should we think that the target personal-level phenomena that drives are supposed to explain are that complex either? In short, the better a job Clark and Dudrick do in persuading us that animal obeying is obeying *proper*, the less obvious it is that phenomena like willing should be considered more sophisticated. And with that, Clark and Dudrick have simply lurched from the second horn of the dilemma, back to the first.

[20] One way of putting the thought is that recognizing a dominant individual *as* a dominant individual does not suffice for recognizing the *legitimacy* of their status. Nietzsche's characterization of animal behaviour seems only to require the former.

3.

As should now be evident, Clark and Dudrick's reading of Nietzsche is problematic, both philosophically and in terms of how it aligns with a range of other things he says. As earlier stated, a view's being false certainly does not mean that it wasn't also Nietzsche's view. Neither, for that matter, does Clark and Dudrick's running into a number of philosophical difficulties to which Nietzsche was sensitive occlude our ascribing their view to him (although it might gently encourage us to think again). The burden placed upon such a reading, however, increases if there is a viable, well-supported, and philosophically less problematic alternative available. Clark and Dudrick are right to highlight that work to date has failed to pay sufficient attention to Nietzsche's choice of metaphor. My primary intention in what remains of this paper is to provide an alternative interpretation of what Nietzsche intends the idea of a *Gesellschaftsbau* to convey.

There are several factors that have likely contributed to the lack of ink spilled on the idea of a "social structure of drives and affects" (BGE §6). The first is that one might reasonably be hesitant about reading *too much* into this particular image. As Mattia Riccardi notes, "concrete illustrations" of what such societies might look like are "frustratingly rare in [Nietzsche's] published work" (Riccardi, 2017: 26–27). Indeed, it is plausibly for this reason that Bernard Williams observed, somewhat sardonically, that although there have been attempts to extract from Nietzsche's work something resembling political thought, his writing appears often to function as a blank surface onto which interpreters have been able to project what they wish to find (see Williams, 2001: xi-xii).[21] Nietzsche of course hardly hid his distaste for politics,

[21] For interesting subsequent treatments of Nietzsche and political thought, see Shaw (2010) and Drochon (2016).

describing himself in the 1908 printing of *Ecce Homo* as "the last *anti-political* German" (EHβ "Wise", §3).[22]

Another factor that plausibly contributes to the limited attention paid to what Nietzsche meant by a 'social structure'—one very much in the spirit of not reading too much into it—is that there might seem to be a rather straightforward interpretation of the metaphor available. Societies are clearly constituted by a great many distinct people, institutions, etc. An intuitive (and far-from-implausible) explanation of the metaphor would be that selves are like social structures in that they are composite things that we routinely represent as single, unified things. There is clearly something right about this, but I am going to argue that there is much more to the metaphor than the idea that the self is a fragmentary thing.[23] Nietzsche's selection of imagery was rarely, if ever, careless—let alone repeatedly so. Its purpose, I contend, is not simply to convey that the self is a multiplicity we represent as a single entity, but that organic unities are to be understood as organized multiplicities *in the same way* that a society is an organized multiplicity. This is what I take him to be suggesting in a note written between 1885–86, in which he states that "All unity is *only* unity as *organization* and *interaction*: not unlike the way in which a human community is a unity: so, a contrast to atomistic *anarchy*; thus a *dominance-construct* [Herrschafts-Gebilde] that *signifies* One, but *is* not One" (eKGWB 1885: 2 [87]). As this rather condensed note suggests, there is a fair deal going on in this comparison of self and state that stands in need of unpacking. To this end, our first port of call will be providing a little background to Nietzsche's interest in the idea of unity "as organization".

---

[22] It is widely accepted that this version was edited, and perhaps trimmed-down, by Heinrich Köselitz ('Peter Gast'). §3 of "Why I Am So Wise" in the versions used by scholars now contains some rather choice remarks about Nietzsche's mother and sister.

[23] Nietzsche's suggestion that we might conceive of the soul as the "multiplicity of the subject" (BGE §12; cf. eKGWB 1885: 40 [42]) would render the social metaphor somewhat redundant if the multiplicity of constituents were all Nietzsche was looking to convey.

3.1

In 1868, seventeen or so years prior to the *Nachlaß* note just mentioned, Nietzsche started preparing a doctoral dissertation on a more explicitly philosophical theme than his prior work. His notes indicate that he intended to write about teleology and the organic "since Kant", examining certain features of Kant's discussion of teleology in light subsequent German materialism and developments in the life sciences. In the third critique (*The Critique of the Power of Judgement*, hereafter CPJ), Kant considers the nature and necessity of teleological judgement. Things are called *purposive*, he states, as long as their possibility "can only be explained and understood by us, so far as we assume, as their basis, a causality according to purposes, i.e. a will that would have so arranged them, according to a representation of a certain rule" (CPJ §10).[24] A subset of those things for which this is true are organisms, or "organized beings" as Kant refers to them, which he identifies as *Naturzwecke* or *natural purposes* (see CPJ §64; §65).[25]

To illustrate the idea of a *Naturzweck*, Kant provides the contrasting cases of a tree (a natural purpose) and a watch (a purposive artefact). The tree, like the watch, is an organization of parts understood only in relation to the whole. We make sense of both by appeals to teleological notions—e.g. "this part is *for* that" or "this part has *that* purpose". Unlike the watch, however, the tree is described as being "*both cause and effect of itself*". Kant has three things in mind here: first, that organisms reproduce, and in so doing, 'cause' the form "with regard to its *species* [G a t t u n g]"; second, that organisms take in nourishment, and sustain, grow, and repair themselves; and third, that the parts of an organism stand in relationships of mutual support and/or dependence (CPJ §64).[26] Kant's central claim, at least as it is presented in sections 64

---

[24] This is the case even if the possibility of an "object, state of mind, or also action" does not require the representation of a purpose (CPJ §10).

[25] It is common, in Kant scholarship, to translate *Zweck* as "end" (often so as to contrast it, in the moral philosophy, with "means"). I use "purpose", here, because that is how the word is more commonly translated in Nietzsche's works.

[26] While it is correct that the parts of a watch stand in relationships of dependence relative to the end of the watch functioning, Kant's point is that the *parts* of the organism depend upon each other for their own

and 65 of CPJ,[27] is that we experience organisms as being such that the nature of their constituent parts are, in some sense, determined by their overall form. It is for this reason, he thinks, that we can *only* conceive of natural purposes (and their parts) when imagined as the result of an instance of willing, with the overall form as the represented purpose or end. Kant's famous remark concerning the less-than-rosy prospects for a Newton for the blade of grass rests upon the idea that we cannot achieve "sufficient knowledge of organized beings and their inner possibility, much less an explanation of them, by looking merely to mechanical principles of nature" (CPJ §75).[28]

Although it goes without saying that there is considerably more to Kant's arguments than I have been able to convey in the short paragraph above, it suffices for our purposes to have sketched out this brief summary. Nietzsche's remarks in response evince a characteristically idiosyncratic mixture of sources, reflecting both classical influences and an interest in the sciences of his day. The first strand I wish to draw out concerns what Nietzsche refers to as the "Empedoclean view" (KGW 1.4: 550; 568). He stresses on more than one occasion that such a view stands very much contrary to Kant's insistence that we cannot but conceive of organisms as though willed. This almost certainly refers to Empedocles' *cosmogony*, a (*very*) early precursor to evolutionary views. Empedocles claimed that as the elements entered into combination, all manner of weird and wonderful forms resulted—lone limbs, limbless torsos, heads without necks, and so forth, all abounded. These then formed unions based upon chance bumpings into one another, brought about by cosmic cycles of love and strife. The eventual result of said

---

functioning. Without leaves the tree's roots, etc., will die; likewise without the roots, the leaves and branches will die.

[27] It is not clear that Kant's various statements of his position all claim the same thing. For an excellent discussion, see Guyer (2006: ch. 10).

[28] Kant qualifies this remark somewhat, claiming straight after that it would be "presumptuous" to conclude that an explanation of the possibility of organisms that didn't involve a design of sorts could not be found. His position, instead, is that we should refrain from passing judgement one way or the other (i.e., that a mechanical explanation can account for organisms, or that a designer must exist). We should instead acknowledge that, given the restrictions of our cognitive capacities, *we* are incapable of explaining such things without recourse to an intelligent designer (§75).

bumpings were creatures with a variety of combinations of body parts (including parts from different species, combined in the same creature). Most of these combinations, Empedocles speculates, didn't fare too well, and met rather hasty ends. Some forms endured, however, producing the rich panoply of living things that now surrounds us.

Despite its abundant shortcomings, Nietzsche contends that the Empedoclean view gets something importantly right. He insists that "we know the method of nature as to how such a 'purposive' body emerges, a senseless method"; one on which "purposiveness proves itself only as viability" (KGW 1.4: 553). The purposive, he goes on, "originated as a special case of the possible", in which "a welter of forms originates, i.e. mechanical compositions: among these innumerable [forms] there can be viable ones" (KGW 1.4: 559). These are "configured according to a chain of failed and half-successful trials" (KGW 1.4: 562). Nietzsche's first complaint thus concerns the idea that we cannot avoid conceiving of purposive natural things as instantiating a form which determines the nature of the thing's parts. Although he might have winced somewhat at the idea that his view amounted to claiming that we could fully explain an organism by reference to the "mechanical principles of nature" (CPJ §75), particularly given his understanding of 'mechanism',[29] his position certainly does amount to saying that while we can, and often do, think of organisms in the way Kant outlines, it is not the case that we *must* in order to provide an explanation of them (or their various parts).[30]

Nietzsche's remarks suggest that part of his intention was to supplant Kant's *form*-first conception of purposive entities with a *part*-first conception. Instead of understanding natural purposes as each possessing a form that seems to determine the nature of its parts, Nietzsche's view is that organisms comprise a multitude of parts, whose interactions determine its form(s).

---

[29] At the time of the preparation of his dissertation, Nietzsche takes mechanism to be "explaining by outer causes" (KGW 1.4: 557). What Kant means by "mechanical" remains a contested issue. His remarks, however, suggest a slightly broader conception of the mechanical than Nietzsche's.

[30] This idea, that we could replace a more robust teleology of intention, with a naturalized notion of 'purpose', understood as a product of "viability", is of considerable importance to Nietzsche's mature work. See, e.g., GS §1; §4; BGE §4; §188; eKGWB 1887: 9 [173]; 1884: 26 [137]; 1880: 6 [366].

This becomes somewhat clearer if we move to the second set of critical comments that can be extracted from Nietzsche's doctoral notes. At first blush, these seem to come rather from left field. Nietzsche's notes shift from engagement with Kant to the work of Goethe. Strange though this might sound, on close inspection, what Nietzsche looks to draw from Goethe's writings is of relevance both to his thoughts on Kant, and to our enquiry more broadly. Nietzsche focuses specifically on Goethe's work on morphology, the study of the *form* of living things. Goethe had initially broached this topic in his work on metamorphosis (see Goethe, 2012/1790), and coined the term 'morphology' six years later. Organic life, as Goethe presents it, undergoes a continual process of formation and transformation. The notion of 'form' we find in Goethe's work is wholly dynamic, understood as an abstraction from particular stages of a process of ongoing flux, development, and mutation.[31]

Nietzsche was clearly impressed by this idea. He writes that the "eternally becoming is life; by the nature of our intellect we conceive forms: our intellect is too obtuse to apprehend the perpetual metamorphosis: that which is cognizable to it, it calls form". Nietzsche takes this thought to touch on a number of issues, one of which is of particular interest to us here.[32] He follows the above remarks with the line, "Truly, there can be no form because in each point sits an infinitude" (KGW 1.4: 570). This likely represents a slightly hyperbolic statement of an idea again derived from Goethe. If we return to Nietzsche's doctoral notes, he cites Goethe directly, quoting a line from *Bildung und Umbildung organischer Naturen*, one of Goethe's most prominent papers on morphology. Nietzsche writes, "'Everything that lives', Goethe says, 'is no unity, but a plurality: even insofar as it appears as a unity to us, it is the gathering of living independent beings'" (KGW 1.4: 556; cf. Goethe, 1807/1950: 16). The dynamic conception of form we find

---

[31] See Ernst Cassirer: Goethe "did not think geometrically or statically, but dynamically throughout. He did not reject permanence, but recognized no other than that which displays itself in the midst of change" (Cassirer: 1950: 138).

[32] The other claim Nietzsche makes in connection to this is that Kant has "foist[ed] the concept of a *thing*" (KGW 1.4: 566) on what is in fact a dynamic process. The idea that certain sorts of cognition require the presumption of identical objects that instantiate fixed forms, and that this is in some sense problematic, pops up at numerous points over the rest of Nietzsche's productive life (particularly in his notebooks).

in Goethe's writings is a product of the interaction of the constituent parts, understood as living individuals. Echoing this, Nietzsche follows the claim that "there can be no form because in each point sits an infinitude" with the suggestion that a "similar concept [to form] is the individual. So one calls organisms units, centres, or ends. But there are only units for our intellect. Each individual has an infinitude of living individuals within itself" (KGW 1.4: 570; cf. BGE §19).

As is to be expected of a collection of notes for a project that Nietzsche never actually completed (or even, for that matter, properly commenced), the above remarks are broad and somewhat inchoate. These preparatory writings, however, offer a glimpse into several areas of admittedly underdeveloped, but nevertheless interesting, engagement with the issue of purposiveness in the organic world. In what follows, I wish to suggest that the thoughts, here loosely expressed, remain with Nietzsche into his mature work, and are sharpened into something far more substantive. Indeed, it is thinking along these lines, I contend, that underpins the idea of unity "as *organization* and *interaction*: not unlike the way in which a human community is a unity […] a *dominance-construct* that *signifies* One, but *is* not One" (eKGWB 1885: 2 [87]). It is precisely to characterize the nature of this organization and interaction that Nietzsche employs the metaphor of the 'social structure'.

4.

Establishing Nietzsche's interest in the purposiveness of organisms, and his sympathy for the idea that their forms are generated by the interaction of constituent parts, offers us a starting point from which to flesh out the idea of a *Gesellschaftsbau*. While Nietzsche says little in his dissertation notes regarding what the "living individuals" might be, when he offers an example of a "power unconsciously creating the purposive", his suggestion is "the instinct of the animals". This, we are told, "is the standpoint of natural philosophy" (KGW 1.4: 570). Such comments plausibly reflect Nietzsche's earliest steps in the direction of what would eventually become the picture we encounter in BGE §12. As it stands, however, this is not greatly different

to the soul-as-composite idea that I suggested earlier was incomplete as an interpretation of Nietzsche's 'social structure' metaphor. In this section, I want to draw out a more precise picture of what he took a social or political structure to involve. The fullest articulation of Nietzsche's thinking on the matter is of course to be found in *On the Genealogy of Morality*, and I agree with Clark and Dudrick that what we encounter therein is of considerable relevance to how we understand the *Gesellschaftsbau*. I want, however, to suggest a quite different reading to theirs, which requires us first to attend to some particular ideas that seem to have contributed to the development of Nietzsche's views.

4.1.

Our first port of call is the work of Wilhelm Roux. Roux, who later became better known as an embryologist, was focused on a quite different topic at the time of Nietzsche's engagement with his writing. The particular piece with which Nietzsche was well-acquainted was Roux's *Der Kampf der Theile im Organismus*, which Nietzsche read in 1881.[33] Its significance, for our purposes, is that *Der Kampf* was written to address precisely the issue that had so interested Nietzsche in his proposed dissertation. Roux aimed to provide a naturalistic account of what he referred to as the "self-formation of the purposive [*Selbstgestaltung des Zweckmässigen*]" (Roux, 1881: 27). Natural purposes, Roux contended, emerge through a struggle between the parts of the organism. Cells, tissues, organ systems, etc., compete for nutriment and space in cycles of excitation and over-compensatory response, with the victorious parties in these struggles assimilating and appropriating available nourishment (Roux, 1881: 160–61). Perhaps more importantly for us, Roux proposed that self-regulation in organic entities is a matter of weaker parts being assimilated to stronger ones, and being re-directed in their functions, under the pressure of the more powerful entities (Roux, 1881: 163–64).

[33] The work was published that year. Nietzsche requested a copy from Ernst Schmeitzner on 21st June, 1881 (eKGWB/BVN 1881: 118).

If we turn to the *Nachlaß*, the impact of Roux on Nietzsche's thought is clearly discernible. In a series of notes from 1883, Nietzsche writes that the "struggle of the tissues becomes a *regulating principle*: the principle of the *functional self-formation* [S e l b s t g e s t a l t u n g] *of the most purposive proportions*" (eKGWB 1883: 7 [190]); and talks of those qualities of cells for which "assimilation exceeds disintegration" and that engage in "overcompensation" regarding consumed material, which "justifies the *dominance* of this important property over the other qualities" in the "struggle for food and space" (eKGWB 1883: 7 [95]).[34] The *Nachlaß* from this period not only evinces the impact of the physiology Nietzsche was reading, however, but also makes clear his connecting these themes with both psychology and social imagery. He writes, for instance, "As cell stands beside cell physiologically, so drive beside drive. The general image of our being is *a socialization* [V e r g e s e l l s c h a f t u n g] *of drives*, with ongoing rivalries and individual pacts between each other. The intellect [is] the object of the contests" (eKGWB 1883: 7 [94]).

A later note, entitled "Of Rank Order", written around the time of the passage concerning unity "as *organization* and *interaction*" (eKGWB 1885: 2 [87]), reflects Nietzsche's continued interest in Rouxian ideas.[35] He writes, still, of the "struggle of the tissues", but, by this point, is utilizing explicitly political metaphors to highlight some very particular relations. Nietzsche talks of the "aristocracy in the body" and of "slavery and the division of labour", claiming that "the higher type [is] only possible through the *subjugation* of [the] lower to a function". He goes on to suggest that "disintegration" occurs with the "impotence of the dominant cells to organize the appropriated. The *formative* force is that which wants always to have new 'material' (even more so, 'force') available" (eKGWB 1885: 2 [76]).

---

[34] Nietzsche, in this passage, in fact connects the struggle between cells, etc., to the "viability [*Lebensfähigkeit*]" of processes (eKGWB 1883: 7 [95]).

[35] To be clear, I am not claiming here that Roux was the only influence on the thoughts expressed in these passages, or, correspondingly, in his eventual view of social structures. In these particular sections, the influence of William Rolph's *Biologische Probleme* (1882) is also evident.

Roux's work presented to Nietzsche a view from the contemporary sciences of purposive organisms as communities of 'living individuals'. Central to this view, however, was the claim that functional organization was possible only through relations of dominance and assimilation. This would undoubtedly have resonated with the picture of human communities that Nietzsche had encountered in other works. For example, if we look to Walter Bagehot's *Physics and Politics* (published in German as *Der Ursprung der Nationen*, of which Nietzsche owned a copy), we find a discussion of the formative stages of human societies.[36] Bagehot argued that early communities were necessarily restrictive, oppressive, and illiberal, so as to mould people into something that could exist in a civilized society (Bagehot, 1873: 27; cf. BGE §262).

4.2.

While Nietzsche left us little that could be readily construed as systematic political thought, he very clearly shared Bagehot's interest in the *origin* of the state. If we turn to the *Genealogy*, Nietzsche offers therein a discussion of the forces that resulted in the formation of early social hierarchies. At the root of "noble races" and societies, Nietzsche writes, we find the "beast of prey". He offers as examples, "Roman, Arab, Germanic, Japanese nobility, Homeric heroes, Scandinavian Vikings" (GM I, §11), each a group of "conquerors and lords, which, organized in a warlike manner and with the power to organize, unhesitatingly lay its terrible paws on a population enormously superior in numbers perhaps, but still formless, still roaming about" (GM II, §17). These groups of nobles, we are told, were kept "within limits" among their equals by "mores, worship, custom, gratitude, still more by mutual surveillance, by jealousy", developing restraint, tact, loyalty, and friendship. In such circumstances, however, the "hidden base needs to discharge itself, the animal must get out, must go back into the wilderness". The

[36] *Physics and Politics* was first published in1873. The German translation came out in 1874.

"blond beasts of prey", as Nietzsche christens them, discharge the tension created by restraint *inter pares* by directing their frustrated drives towards groups external to their own (GM I, §11).[37]

It was this, we are told, that produced the earliest "states", which appeared first as "a terrible tyranny, as a crushing and ruthless machinery". The state as a "fixed form" had its genesis in "an act of force", and indeed "could be brought to its completion only by acts of force" (GM II, §17). The order imposed by the blond beasts was enforced through painful punishments for recalcitrance, which resulted in "an increase of fear, a sharpening of prudence, [and] mastery of the appetites" (GM II, §15). The need to master—to *counter* or *control*—those problematic instincts on which the peoples had previously relied resulted in a proliferation of *bad conscience*,[38] which made "a raw material of people and half-animals" into something "kneaded and pliable" (GM II, §17; cf. Bagehot, 1873: 27).

The discharging of those frustrated drives on the part of the beasts made this "raw material" not just into something malleable, but something "*formed*". Nietzsche states explicitly that the beasts of prey are "the most involuntary, unconscious artists there are", and that the forging of the state was "an instinctive creating of forms, impressing of forms" (GM II, §17). This is not to say that the dominant groups were unaware of the actions they were performing, but instead that they did not engage in such behaviour with the conscious purpose or intention of producing the resultant state of affairs. The fruit of this unconscious creation is a "a ruling structure that *lives*, in which parts and functions are delimited and related to one another, in which nothing at all finds a place that has not first had placed in to it a 'meaning' with respect to the whole" (GM II, §17).[39]

---

[37] Cf. Nietzsche's claim, in *Twilight of the Idols*, that Greek institutions developed as a protection against their "will to power". The result of this was a "tremendous inward tension that discharged itself in terrible and ruthless hostility toward the outside world" (TI "Ancients", §3).

[38] For a more detailed discussion of the psychology of bad conscience, see paper one in this thesis. The "bad conscience" in play here is that second form mentioned in that paper (§2.3), explicitly involving aggressive self-punishment.

[39] One might ask, at this point, about the relation between this remark and what I referred to earlier as Kant's *form*-first conception of purposive entities. Nietzsche does, after all, talk of things finding a place, and having

The picture of the human community with which we are confronted in the *Genealogy* is rich in its detail, and betrays the influence of Nietzsche's engagement with both political and physiological thought.[40] The form of the state is generated initially by the oppressive force of a dominant group, "under the pressure of the blows of their hammers, of the artists' violence" (GM II, §17; cf. BGE §225). This social structure is not only forged but also maintained under this pressure, requiring relations of domination and subjugation between components to mould the subordinate groups into functionally contributing parts that then perpetuate the "ruling structure that *lives*" (GM II, §17). If *this* is the sort social structure Nietzsche has in mind, and we surely have good reason to think that it is, then as the beasts overpower peoples and, from the vantage point of their ends, ascribe functions and interpret meanings in relation to the whole, so will dominant drives and affects. Selves, so understood, would be the result of the strongest instincts forcing and shaping subordinate states and dispositions into a form or structure—i.e. an organized, functional psychological constitution.

This presents a picture of the social order quite different to the one we find in the work of Clark and Dudrick. On their view, a crucial part of what Nietzsche intends to convey concerns the drives' recognition of normative relations, and their corresponding behaviour. Yet, if we look to his remarks about political structures, Nietzsche's eye is very much trained on the part of the process that Clark and Dudrick acknowledge consists in "brutish, forcible interaction" (Clark and Dudrick, 2012: 203). While he does think that this pressure alters the behaviour of the subjugated groups, we find little in the *Genealogy* about the slaves recognizing the legitimacy of

---

meaning, in relation to the whole. A *part*-first conception, however, does not commit Nietzsche to the idea that parts do not have functions in relation to the whole. He is committed only to the claims that the (fluid) form of the whole is a product of the interaction of its parts, and that we can explain/conceive of the thing in question without having to think of it as akin to something designed.

[40] On a number of occasions in the *Genealogy*, Nietzsche is quite explicit in stressing the continuity of his claims about the state and its practices with the organic realm more broadly. In the course of a discussion of punishment, for instance, he writes that the "form is fluid, but the 'meaning' even more so . . . Even in the individual organism things are no different: with each essential growth of the whole, the 'meaning' of the individual organs shifts as well [...] all happening in the organic world is an *over-powering*, *a becoming-lord-over*; and that, in turn, all overpowering and becoming-lord-over is a new interpreting, an arranging by means of which the previous 'meaning' and 'purpose' must of necessity become obscured or entirely extinguished" (GM II, §12). For remarks regarding individual psychology that are redolent of this, see §290 of *The Gay Science*.

the masters, or the nomads consenting to the enduring tyranny of the blond beasts. The reason for this, I suggest, is the same as explains Nietzsche's generous extension of political metaphors not just to drives, but across the organic domain *in toto*, to cells, tissues, and organs: namely, he simply does not intend such imagery to highlight the constituent parts' capacity to recognize normative considerations.

5.

We have, so far, delved a little deeper into the origin of Nietzsche's interest in the structure of organisms and into his understanding of the state as evincing its own organic unity, tracing the thread of these ideas from his doctoral dissertation through to the *Genealogy*. In this section, I wish to flesh-out what it would mean for a soul to be a "social structure of drives and affects" (BGE §12), and to employ Nietzsche's social metaphor to reconstruct the possible relationships between drives. Much as Roux's account of the behaviour of cells, etc., required them to possess some quite basic formal features regarding stimuli-sensitivity and over-compensatory response, Nietzsche's account of the relations between drives/affects also plausibly presumes some shared features. The first of these is that the 'behaviour' of these psychological tokens is characterized by *will to power*. On more than one occasion, Nietzsche talks of both drives and affects as being either reducible to, or forms/manifestations of, will to power (see, e.g., eKGWB 1888: 14 [121]; 1885: 40 [61]). Drives, for instance, are characterized by striving for expression through particular activities or patterns of activity. This compulsion to discharge brings the drives/affects into conflict with external obstacles (e.g. features of one's environment, other people, etc.) and internal resistance (i.e. *other* drives and affects, equally characterized by striving towards *their* discharging activity). It is for this reason that Nietzsche describes the drives as being each a "kind of lust to rule" (eKGWB 1886: 7 [60]).

The second feature of drives that I wish to highlight is their *plasticity*. If we return to the *Genealogy*, in the case of the 'beasts', they are constrained by the mores of their society, and vent their frustrated drives on others. In the case of the subjugated peoples, the restrictions the

encounter are those imposed by the beasts, and their frustrated drives eventually discharge on, *inter alia*, themselves (for complex reasons). As Nietzsche makes clear, drives that are frustrated in *all* externally directed attempts to discharge will turn inward, and become *internalized* (GM II, §16). Where the drives cannot overcome the resistance faced—when the standard object for the drive is unavailable, or the characteristic activity to which the drive inclines the organism is unachievable— they are sufficiently plastic as to seek alternative means of discharging.[41]

If we take these two features of drives in conjunction, it presents to us a number of different, complex structural relations in which such states can stand to one another. The first of these we can label *dominance*. One drive *dominates* another in a given instance if the two drives compel the organism towards incompatible activities, and one wins out as the stronger in a direct 'struggle'. As we saw earlier, Nietzsche takes the "intellect" to be the "object of the contests" (eKGWB 1883: 7 [94]). In such instances, the stronger drive modulates cognition and monopolizes cognitive resources (cf. Gemes, 2009: 51).[42] That drive would be in a *position* of dominance if it were consistently more powerful than its rival, and won out in most (or perhaps all) of the situations in which the two drives were in direct competition. As Nietzsche makes clear in *Daybreak*, the lasting effect of this dominance will be for the "vehemence" of the dominated drive to be weakened (D §109; cf. §119), thus cementing the dominance of one over the other. Regarding the analogy with the state, *dominance* is the relationship in which the blond beasts stand to the tribes and nomads in their earliest encounters: the beasts win out and impose their will on the weaker peoples, irrespective of resistance.

The *dominance* relation requires only that drives be characterized by will to power. The plasticity of drives, however, makes possible a second, more complex, relation—namely, that of

---

[41] To be slightly more precise, the most plausible way to think of plasticity within a Nietzschean framework would be variable ranges of input sensitivity. In cases where dispositions are overwhelmed with opportunities to discharge, the range of stimulating input contracts. In cases where drives are starved of opportunities, the range of possible stimuli becomes *much* broader, sometimes extending to entirely new objects and activities (cf. discussion of 'sublimation' further on in this section). If these new objects/activities provide a means of discharging, the dispositions' sensitivity to such things is reinforced.

[42] I have argued elsewhere that this modulation is part of the affective process (see paper one, this thesis).

sublimation.[43] This is the relation Nietzsche describes in the following *Nachlaß* passage, where he writes, "*Overcoming of the affects*?—No, not if that means weakening and annihilating them. *Instead*, *drawing them into service*, which might include exercising a long tyranny over them (not just as an individual but even earlier, as a community, race, etc.)". This 'tyrannizing' frustrates the drive or affect in its compelling of the organism towards certain activities, thus forcing it to seek alternative means of discharging. Owing to their plasticity and will to power, the drives (initially dominated and weakened) find avenues through which to discharge, internal to the ends towards which the dominant drive is compelling the organism.[44] Nietzsche follows the above lines in the *Nachlaß* by saying, "In the end they [the affects] are trustingly given back some freedom: they love us like good servants and voluntarily go where our best interests want to go" (eKGWB 1885: 1 [122]). As the beasts do to the subjugated peoples, the tyrannical drives change (or condition) the dominated dispositions. Certain pathways to discharging become burned into the drives and affects, such that the activity within the end of the dominant power to which they have turned becomes the primary activity to which the now *sublimated* drive aims to compel the organism.[45]

5.1.

This, however, is still not quite the full picture, in terms of the relationships in which drives can stand to one another. As we have seen, one more than one occasion, Nietzsche selects the metaphor of an aristocracy (see eKGWB: 1885, 2 [76]) or oligarchy (GM II, §1) to illustrate

---

[43] For insightful discussions of Nietzsche's account of sublimation, see Richardson (1996) and Gemes (2009). My description here adds little to their accounts.

[44] As an example of this, imagine someone with a drive to competitive activity that is constantly frustrated by their more powerful affective disposition of hunger. In such circumstances, the competitive instinct might come to find expression in the activities towards which my hunger propels me, such as preparing and eating food. I might attempt to eat faster than adjacent diners in a restaurant, for instance; or to sample a wider range of cuisines than this-or-that person, etc.

[45] To be more precise, we should perhaps distinguish between a *fully* sublimated drive and a drive *in the process* of being sublimated. Whereas the former might well have been re-directed towards a new activity to the extent that it can be "given back some freedom" (eKGWB 1885: 1 [122]), the latter, by contrast, seeks expression in the activities to which stronger drives incline the organism, but would likely still push towards its original individuating activity if not for this coercive tyrannizing. (Thanks to Ken Gemes for this suggestion.)

the 'social' structure of organic things. In the *Genealogy*, Nietzsche of course posits not a single dominant individual, but a class or caste that impose form on the (eventually malleable) raw material; and in *Beyond Good and Evil*, he likens willing to the "ruling class" of a happy commonwealth identifying with the success of the entire "social structure of many 'souls'" (BGE §19). There is textual support for the idea that Nietzsche felt that a single, dominant drive could perform the task of shaping the various drives into a whole. He writes in the *Nachlaß*, for instance, that a "Multiplicity and disaggregation of the impulses, lack of system among them, results as 'weak will'; their coordination under the dominance of a single one results as 'strong will'" (eKGWB 1888: 14 [219]; cf. 1888: 14 [157]), which could be read as stating that one drive will form the self. Yet, as I mentioned earlier, Nietzsche was rarely careless in his choice of metaphor, often trying them out for long periods of time in the notebooks before employing them in published passages. So, in what sense might organic life be like an aristocracy?

If we attend closely to the passages in which Nietzsche touches on the formation of the self, he in fact talks reasonably infrequently about a single drive or instinct as the formative force. The explanation for this, I believe, can be found by attending to his discussions of great individuals. Nietzsche expresses scepticism on more than one occasion about whether the average human being is much beyond a jumbled assortment of poorly-aligned impulses (see, e.g., TI "Expeditions", §35; §41), so it should perhaps come as no surprise that discussions of *Selbstgestaltung* are skewed rather in the direction of 'higher types'. He suggests in multiple passages that great individuals will have a range of strong and diverse instincts and abilities. Nietzsche describes Goethe, for instance, as "strong enough" to permit himself "the whole compass and wealth of naturalness", and as someone that "aspired to […] *totality*; he strove against the separation of reason, sensuality, feeling, will" (TI "Expeditions", §49). The greatest, he writes elsewhere, "possess perhaps also the great virtues, but at that moment still their opposites […] out of the presence of opposites, and out of their feelings, the great man in particular, *the bow with the great tension*, emerges" (eKGWB 1885: 35 [18]).

Perhaps most instructive of all is Nietzsche's discussion of *his own* development in *Ecce Homo*, in which we find the same sentiment resurfacing: "For the task of a *revaluation of values*", he writes, "more capacities perhaps were required than have dwelt together in one individual; above all antithetical capacities which however are not allowed to disturb or destroy one another. Order of rank among capacities; distance; the art of dividing without making inimical; mixing up nothing, 'reconciling' nothing; a tremendous multiplicity which is nonetheless the opposite of chaos—this has been the precondition, the protracted secret labour and artistic working of my instinct" (EHα "Clever", §9).[46] The relationship between Nietzsche's various drives and dispositions expressed in this passage is clearly not simply one of dominance, or even sublimation, the latter of which would represent the *reconciling* of different instincts—just what he claims did not occur.

Although Nietzsche does refer to an instinct in the final line quoted above, it is plausibly not a dominant drive, but the "instinct for *self-defence*" to which he draws our attention in the preceding section (EHα "Clever", §8).[47] If we look for the locus of this formation process, Nietzsche makes it clear that it is in fact the "organizing 'idea'", i.e. the idea of the revaluation of values. If we look to the 1886 preface of *Human, All Too Human*, some of which bears

---

[46] Some of Nietzsche's ideas in regard to great individuals seem to have been influenced, once again, by his engagements with physiology. If we return to the passage of Goethe's that Nietzsche quoted approvingly in the preparatory notes for his dissertation, after the point at which Nietzsche's quotation ends, the passage continues as follows: "The *more imperfect* a being is, the more do its individual parts *resemble each other*, and the more do these parts *resemble the whole*. The *more perfect* the being is, the more *dissimilar are its parts*. In the former case the parts are more or less a repetition of the whole; in the latter case they are totally unlike the whole […] The more the parts resemble each other, the less subordination is there of one to the other. *Subordination of parts indicates a high grade of organization*" (Goethe, 1807: 16). (This passage was quoted in George Henry Lewes' *Life of Goethe*, where, in order to illustrate Goethe's point, the author offers a human society as an example, stating that in a simple society, "every man is his own tailor" [Lewes, 1855: 107]. Nietzsche read Lewes' *History of Modern Philosophy* in the 1870s. Whether he read *Life of Goethe* is unclear, but it is an intriguing thought.) A similar picture is presented in the work of Carl Nägeli, in his *Mechanisch-physiologische Theorie der Abstammungslehre* (1884). In a passage underlined by Nietzsche in his own copy, Nägeli writes that "Perfection in my sense is therefore nothing other than the progression towards a more complex structure and to greater division of labour" (Nägeli, 1884: 13). If we look to the *Nachlaß*, Nietzsche writes therein that "greater complexity, sharp differentiation, the contiguity of developed organs and functions with the intermediate members disappearing—if that is *perfection*, then a will to power manifests itself in the organic process" (eKGWB 1886: 7 [9]).

[47] This instinct putatively facilitated Nietzsche's flourishing, but by drawing him away from environments inimical to his well-being, not by tyrannizing over his others drives (see EHα "Clever", §8).

significant resemblance to EH "Clever", §9, we find Nietzsche addressing free spirits—people "in whom a *task* [A u s g a b e] wants to become incarnate". This task, he writes, comes to "rule among and in the individual facets of [their] destiny like an unconscious pregnancy" (HAH P, §7). Given Nietzsche's drive-psychology, it is surely correct that 'tasks' and 'ideas' coming to dominate will correspond to the interaction of drives and affects. Nevertheless, these represent peculiar word choices if Nietzsche is simply trying to express the idea that a stable self is forming around a single, dominant drive.

The reason for these particular choices, I submit, is that Nietzsche simply *is* referring (rather straightforwardly) to an idea or task because he takes activity to play a crucial role in the formation of the self. The phrase "organizing 'idea'" arises only in Nietzsche's discussion of himself, as it is only there that he is talking about an individual whose psychology developed around an *intellectual* task that gradually became clearer over time. John Richardson's work highlights a final drive relationship that seems to me to be relevant here. Richardson writes of *synthesis*, which he uses as a broad term to cover the variety of ways in which drives can form stable combinations tending towards particular activities, the way in drives combine to form an enduring self, and the way in which people form structured relations within a society (see Richardson, 1996: 44–52). One possible form of *synthetic* will that Richardson mentions in passing is a drive complex in which "the drives are more evenly balanced, with one perhaps dominant in some respects or contexts but the other dominant in other respects". In these instances, he suggests, we "won't still attribute the activity of the resulting whole to either of its members", which now "form a unit with a distinctive activity of its own" (Richardson, 1996: 47). After a sufficiently protracted period of time, the organism can be said to have a new, synthetic drive to the activity in question.[48]

The idea that there might be activities that facilitate the discharge of more than one drive, without sublimation having to take place, hits me as a tremendously important one. I am

[48] To return to the example offered in n. 44, one's competitive drive and hunger disposition could be fully discharged in an *eating contest* without either having to sublimate or weaken.

sceptical about the textual warrant for referring to the subsequent stable construct as a *drive*, as, to my eye, Nietzsche tends to reserve 'drive', etc., for the simpler units into which he breaks down our capacities for complex activities/willings, even once these capacities have become quite deeply-ingrained dispositions (see, e.g., GS §113; cf. BGE §229; §230). Nevertheless, Richardson's insight here seems to offer a way of making sense of the claim that something like an "organizing 'idea'" could shape the self. The tasks Nietzsche has in mind when discussing great individuals facilitate a maximally broad and powerful expression of their strongest instincts, which, in time, form an integrated composite of diverse dispositions, where few are compromised, reconciled or mixed up. This, of course, only strengthens the strongest drives, and renders more robust the hierarchical relations between drives and affects.[49]

This general picture provides a means of understanding why, in the case of the great individuals at the very least (and perhaps universally), Nietzsche selected the metaphor of an aristocracy or oligarchy. As the beasts did in the *Genealogy*, the great individual's dominant, aggressive drives form a stable and (in one sense) cooperative class, which compels the individual towards the particular activities or tasks that facilitate their maximally powerful mutual expression (cf. eKGWB 1883: 7 [94]).[50] Nietzsche rarely talks explicitly of great individuals in terms of their having a single tyrannical drive, but more commonly addresses their characteristic 'tasks', and speaks of them having strong or enduring *wills*.[51] The structure of drives is plausibly not to be understood as the order in which drives stand at all times, in every particular instance of willing

---

[49] As Nietzsche writes in the *Nachlaß*, "Our actions *re-form us*: in each action, certain forces are exercised, and others *not* exercised, [and] so temporarily neglected: an affect always affirms itself at the expense of other affects, whose power it takes away. The actions that we *do the most* are finally like a *fixed enclosure* around us […] [similarly] a regular neglect shapes people" (eKGWB 1883: 7 [120]).

[50] Individuals with integrated groups of dispositions but no unifying task end up as "chameleons", the "most interesting men". Nietzsche writes that such types are healthy "as soon as an [internal] order is found", and are described as "happy and sure", but "have no development": their "states lie beside one another, even when they have been separated seven times. They change, they do not *become*…" (eKGWB 1888: 14 [157]).

[51] See, e.g., GM II, §2; BGE §21; eKGWB 1888: 14 [157]. (For arguments *against* the idea that the sovereign individual represents Nietzsche's ideal, see Hatab [2008: 75–78], Davis Acampora [2006: 147–62] and Rukgaber [2012: 231–39].)

or activity, but is instead a complex dispositional order of strength-relations.[52] The individual's 'will' is discernible in *patterns* of occurrent willing that reflect drive activity directed towards a given end over an extended period of time.[53] The position with which we end up that involves a contrast, for Nietzsche, between the enduring *self* and the *occurrent mind*.

An individual like Goethe engages, as Nietzsche stresses, in a wide variety of activities and has a diverse set of apparently antithetical instincts and capacities. In different circumstances, different dominant drives might well take the lead. Additionally, as we saw earlier, the successfully sublimated affects can also be "given back some freedom" (eKGWB 1885: 1 [122]; cf. Z II, §6). This suggests the possibility of episodes where the most powerful occurrent instinct or affect at a given moment is one sublimated within the overall hierarchy. These instances, in the well-structured individual, will perhaps occur when stronger drives are sated, and/or the activity to which the sublimated drive compels the organism is not one inimical to those things valuable from the perspective of the dominant instincts. In the well-structured self, however, awareness of purposes and desires that *would* undermine the tasks to which the dominant drives incline will stimulate in them a strong response, consuming the cognitive machinery and excluding the recalcitrant disposition from directing the organism.[54] It is in this sense, I take it, that the mind is, for Nietzsche, an organic unity—a "*dominance-construct* that *signifies* One, but *is* not One" (eKGWB 1885: 2 [87]).

---

[52] Subsequent to writing this paper, I have read Mattia Riccardi's account of the structure of drives (2017), to which my account here bears significant resemblance.

[53] One can thus have a 'will to *X*' even if one's instances of willing are produced by different drives at different times. Correspondingly, different individuals can both manifest the same will, but with different drives producing the actions.

[54] Although I do not have space to pursue the idea further, it seems to me that this provides the most promising way of understanding the idea of "having one's pro and contra *in one's power*" (GM III, §12), viz. as a capacity to indulge various affects (from the perspective of the dominant instincts) without eroding the stability of the social structure. This is, as Nietzsche notes, the greatest example of "self-control" (eKGWB 1887: 10 [143]).

CONCLUSION

As is so often the case with Nietzsche, the swift simplicity with which he presents his reflections rather masks the quantity of thought that he had put into the issue. What appears in BGE §12 and §19 as a single suggestive word—*Gesellschaftsbau*—in fact conveys almost two decades of increasingly refined thought on an important issue. As we have seen, although Clark and Dudrick are wholly right to stress the importance of the metaphor of a social structure, their account of Nietzsche's view runs into considerable trouble. I have attempted, in this paper, to present a viable alternative; a form of dispositional causal reading, that engages in depth with what Nietzsche meant when comparing the self to an aristocracy or oligarchy. His early reflections of teleology and the organic were complemented over time by his engagement with physiology and social thought. At both the level of the organism and the level of the state, Nietzsche sees dynamic form produced by the interaction of constituent parts, conceived of as 'living individuals'. This interaction, however, is specifically one of dominance, sublimation, and (sometimes) synthesis, driving a diversification of capacities and more complex forms of integration. Nietzsche's social metaphor, as we have seen, provides a means of connecting his thoughts on organic structure, social psychology, and great individuals into a coherent whole.

Conclusion

Over the course of this thesis, I have ascribed to Nietzsche a number of substantial positions. I have argued that occurrent affects, on his account, are complex psycho-physiological episodes, comprising patterns of bodily states and the subsequent modulation of specific psychological mechanisms and processes (paper 1). His view anticipates contemporary work on both somatic theories of emotion and interpretive accounts of self-knowledge. Yet, it is also a distinct and original contribution to each area, diverging in interesting and fruitful ways from contemporary positions. The idea that affects partly comprise a modification and directing of the intellect allows Nietzsche to develop a remarkably prescient picture of non-linguistic thought, which he then pairs with an account of conscious thinking that precedes contemporary work on inner speech (paper 2). Nietzsche's account of the interaction of unconscious affects and conscious thoughts (states with linguistically articulated contents that stimulate further unconscious processing) provides a surprisingly sophisticated picture of the occurrent mind, given how pioneering his thoughts on these topics were. This is then contrasted with the enduring self, which Nietzsche conceives of as a 'social structure', necessarily involving relations of dominance and submission (paper 3).

I suggested in the opening paper that the absence of a precise account of what emotions were supposed to be was a serious problem for those hoping to find a compelling psychology among Nietzsche's writings. The pervasiveness of emotion and mood in his picture of mind means that, without such an account, any attempt to reconstruct his positions on various questions stands on fragile foundations. In this thesis, I have attempted to demonstrate that Nietzsche had a range of sophisticated and well-motivated views on emotion, and that he then employed these sensibly and intelligently in the construction of various further positions. Far from being "dangerous" or "half-animal" (eKGWB 1887: 11 [310]), as his interlocutors were accused of thinking,

Nietzsche presents to us a view on which the passions are essential and ineliminable parts of our mental lives—of our*selves*—a view that finds considerable support in contemporary work on affects.

There remains much to be done in this area. I have not been able, here, to critically assess the various components of Nietzsche's views, for instance; or to approach the full complement of positions he develops on which the account of affect has bearing. Nietzsche's discussions of knowing and believing, willing and valuing, and the creation of values itself, all draw on emotions and moods. He offers novel characterizations of particular affects, passions, and disorders, all of which are fertile philosophical ground for further study. There is also, as I have hinted, much to be gained from examining contemporary work on emotion through the lens of Nietzsche's writings, and vice versa. I hope, however, that this thesis has put forward a persuasive case for Nietzsche's being a profound philosopher of emotion, and this being an area of considerable import and interest.

Bibliography

Nietzsche's writings (in translation)

A: *The Antichrist*, translated by R. J. Hollingdale. (London: Penguin, 1968.)

BGE: *Beyond Good and Evil*, (eds. Rolf-Peter Horstmann and Judith Norman), translated by Judith Norman. (Cambridge: Cambridge University Press, 2002.)

CW: *The Case of Wagner: A Musician's Problem*, in *The Antichrist*, *Ecce Homo*, *Twilight of the Idols*, *and other writings*, (eds. Aaron Ridley and Judith Norman), translated by Judith Norman. (Cambridge: Cambridge University Press, 2005.)

D: *Daybreak*, (eds. Maudemarie Clark and Brian Leiter), translated by R. J. Hollingdale. (Cambridge: Cambridge University Press, 1997.)

EHα: *Ecce Homo*, in *The Antichrist*, *Ecce Homo*, *Twilight of the Idols*, *and other writings*, (eds. Aaron Ridley and Judith Norman), translated by Judith Norman. (Cambridge: Cambridge University Press, 2005.)

EHβ: *Ecce Homo* [1908 printing], translated by R. J. Hollingdale. (London: Penguin, 1979.)

GM: *On the Genealogy of Morality*, translated by Maudemarie Clark and Alan Swensen. (Indianapolis: Hackett, 1998.)

GS: *The Gay Science*, (ed. Bernard Williams), translated by Josephine Nauckhoff and Adrian Del Caro. (Cambridge: Cambridge University Press, 2001.)

HH: *Human*, *All Too Human*, translated by R. J. Hollingdale. (Cambridge: Cambridge University Press, 1986.)

TI: *Twilight of the Idols*, translated by R. J. Hollingdale. (London: Penguin, 1968.)

TL: *On Truth and Lies in the Extra-Moral Sense*. In *Writings from the Early Notebooks.* Edited by Raymond Geuss & Alexander Nehemas. Translated by Ladislaus Löb. (Cambridge: Cambridge University Press, 2009).

WTP: *The Will to Power*, (ed. Walter Kaufmann), translated by Walter Kaufmann and R. J. Hollingdale. (New York: Vintage Books, 1967.)

Z: *Thus Spake Zarathustra*. Translated by A. Tille. (London: Everyman, 1950).

German Editions of Nietzsche

eKGWB: *Digitale Kritische Gesamtausgabe Werke und Briefe*. Edited by Giorgio Colli & Mazzino Montinari. Accessed at http://www.nietzschesource.org/#eKGWB.

eKGWB/BVN: *Nachgelassene Fragmente*, *Briefe von Nietzsche*.

KGW: *Kritische Gesamtausgabe Werke*. Edited by Giorgio Colli & Mazzino Montinari (Berlin and New York: Walter de Gruyter, 1967)

Other works cited:

Acampora, Christa D. "On Sovereignty and Overhumanity: Why It Matters How We Read Nietzsche's Genealogy II: 2". In *Nietzsche's On the Genealogy of Morality: Critical Essays*, edited by Christa Davis Acampora, 147–62. New York: Rowman and Littlefield, 2006.

Anderson, R. Lanier. "What is a Nietzschean Self?" In *Nietzsche, Naturalism, and Normativity*, edited by Christopher Janaway and Simon Robertson, 202–35. Cambridge: Cambridge University Press, 2012.

Anscombe, G. E. M. *Intention*. Oxford: Blackwell, 1957.

Asma, Stephen, and Gabriel, Remi. *The Emotional Mind*. Harvard: Harvard University Press, 2019

Bagehot, Walter. *Physics and Politics*. New York: Cosimo, 1873.

Barlassina, Luca, and Albert Newen. "The role of bodily perception in emotion: In defense of an impure somatic theory". *Philosophy and Phenomenological Research* 89 (2014): 637–78.

Bermúdez, José Luis. *Thinking without Words*. Oxford: Oxford University Press, 2003.

———. *Philosophy of Psychology: A Contemporary Introduction*. New York: Routledge, 2005.

Carruthers, Peter. *Language, Thought, and Consciousness*. Cambridge: Cambridge University Press, 1996.

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

Cassirer, Ernst. *The Problem of Knowledge: Philosophy, Science, and History since Hegel*. New Haven: Yale University Press, 1950.

Clark, Andy. "Dealing in Futures: Folk Psychology and the Role of Representations in Cognitive Science". In *The Churchlands and Their Critics*, edited by Robert N. McCauley, 86–103. Oxford: Blackwell, 1996.

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

Constâncio, João. "On Consciousness: Nietzsche's Departure from Schopenhauer". *Nietzsche-Studien* 40: 1–42, 2011.

Crane, Tim. "Intentionality as the Mark of the Mental". In *Contemporary Issues in the Philosophy of Mind*, edited by Anthony O'Hear, 229–52. Cambridge: Cambridge University Press, 1998.

Damasio, Antonio. *Descartes' Error*. London: Vintage, 1994.

Dennett, Daniel. *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: Bradford Books, 1978.

Deonna, Julien, and Fabrice Teroni. *The Emotions: A Philosophical Introduction*. Oxford: Routledge, 2012.

Doyle, Tsarina. "Nietzsche, Consciousness, and Human Agency". *Idealistic Studies* 41: 11–30, 2011.

Drochon, Hugo. *Nietzsche's Great Politics*. Princeton: Princeton University Press, 2016.

Espinas, Alfred. *Die thierischen Gesellschaften: Eine vergleichendpsychologie Untersuchung*. Vieweg: Braunschweig, 1880.

Feuchtersleben, Ernst von. *Lehrbuch der ärztlichen Seelenkunde: als Skizze zu Vorträgen*. Vienna: Carl Gerold, 1845.

Gauker, Christopher. *Thinking Out Loud: An Essay on the Relation between Thought and Language*. Princeton: Princeton University Press, 1994.

Gemes, Ken. "Freud and Nietzsche on Sublimation". *Journal of Nietzsche Studies* 38 (1): 38–59.

Gibson, James J. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin, 1979.

———. "The Theory of Affordances". In *Perceiving, Acting, and Knowing: Towards an Ecological Psychology*, edited by John Bransford and Robert Shaw, 67–82. Hillsdale: Lawrence Erlbaum, 1977.

Goethe, Johann Wolfgang. *Schriften zur Morphologie* II, vol. XIX in *Goethe*, Goethe, J. W. von, & Liselotte Bäuerle Lohrer. Wisconsin: University of Wisconsin, Madison, 1807/1950.

———. *Versuch die Metamorphose der Pflanzen zu erklären*. Tredition Press, 1790/2012.

Griffiths, Paul E. *What Emotions Really Are*. Chicago: University of Chicago Press, 1997.

Guyer, Paul. *Kant*. Oxford: Routledge, 2006.

Hatab, Lawrence J. *Nietzsche's On the Genealogy of Morality: An Introduction*. Cambridge: Cambridge University Press, 2008.

Helmholtz, Hermann von. *Handbuch der physiologischen Optik*. Leipzig: Leopold Voss, 1867.

Hume, David. *A Treatise of Human Nature*, (ed. L.A. Selby-Bigge). Oxford: Clarendon Press, 1739/1967.

James, William. "The Physical Basis of Emotion". *Psychological Review* 1 (1894): 516–29

———. *The Principles of Psychology, Volume II*. New York: Henry Holy and Co., 1890.

———. "What is an Emotion?" *Mind* 9 (1884): 188–205

Janaway, Christopher. *Beyond Selflessness: Reading Nietzsche's Genealogy*. Oxford: Oxford University Press, 2007.

Kant, Immanuel. *Critique of the Power of Judgement*. Translated by Werner Pluhar. Indianapolis: Hackett Publishing Co., 1790/1987.

Katsafanas, Paul. "Nietzsche's Philosophical Psychology". In *Oxford Handbook of Nietzsche*, edited by Ken Gemes and John Richardson, 727–52. Oxford: Oxford University Press, 2013.

———. "Nietzsche's Theory of Mind: Consciousness and Conceptualization". *European Journal of Philosophy* 13 (1): 1–31, 2005.

———. *The Nietzschean Self: Moral Psychology, Agency, and the Unconscious*. Oxford: Oxford University Press, 2016.

———. "Value, Affect, Drive". In *Nietzsche on Mind and Nature*, edited by Manuel Dries and P. J. E. Kail, 163–88. Oxford: Oxford University Press, 2015.

Kenny, Anthony. *Action, Emotion, and Will*. London: Routledge and Keegan Paul, 1963.

Lampert, Laurence. *Nietzsche's Task: An Interpretation of Beyond Good and Evil*. New Haven: Yale University Press, 2001.

Leibniz, Gottfried Wilhelm. *New Essays on Human Understanding*. Translated by Peter Remnant and Jonathan Bennett. Cambridge: Cambridge University Press, 1768/1981.

Leiter, Brian. "Moralities are a Sign-Language of the Affects". *Social Philosophy and Policy*, 30 (2013): 237–58.

———. *Nietzsche on Morality*. New York: Routledge, 2002.

———. "The Paradox of Free Will and Self-Creation in Nietzsche". In *Willing and Nothingness: Schopenhauer as Nietzsche's Educator*, edited by Christopher Janaway, 217–57. Oxford: Oxford University Press, 1998.

Lewes, George Henry. *Life of Goethe*. London: Smith, Elder, and Co., 1855.

Mayer, Julius Robert von. *Die Mechanik der Wärme in gesammelten Schriften*. Stuttgart: J. G. Cotta'schen, 1867.

Moore, Gregory. *Nietzsche, Biology and Metaphor*. Cambridge: Cambridge University Press, 2002.

Papineau, David. "Phenomenal and Perceptual Concepts". In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, edited by Torin Andrew Alter & Sven Walter, 111–44. Oxford: Oxford University Press, 2006.

Poellner, Peter. "Affect, Value, and Objectivity". In *Nietzsche and Morality*, edited by Brian Leiter and Neil Sinhababu, 227–61. Oxford: Clarendon Press, 2007.

———. *Nietzsche and Metaphysics*. Oxford: Clarendon Press, 1995.

Prinz, Jesse. "Are Emotions Feelings?" *Journal of Consciousness Studies* 12 (2005): 9–25

———. *Gut Reactions: A Perceptual Theory of Emotion*. Oxford: Oxford University Press, 2004.

Reil, Johann Christian. *Rhapsodieen über die Anwendung der psychischen Curmethode auf Geisteszerrüttungen*. Halle: Curt, 1803.

Riccardi, Mattia. "Inner Opacity. Nietzsche on Introspection and Agency". *Inquiry: An Interdisciplinary Journal of Philosophy* 58 (3): 221–43, 2015.

———. "Nietzsche on the Superficiality of Consciousness". In *Nietzsche on Consciousness and the Embodied Mind*, edited by Manuel Dries, 93–112. Berlin: De Gruyter, 2018.

———. "Nietzsche's Pluralism about Consciousness". *British Journal for the History of Philosophy* 24 (2015): 132–54.

———. "Virtuous Homunculi: Nietzsche on the Order of the Drives". *Inquiry: An Interdisciplinary Journal of Philosophy* 61 (2017): 21–41.

Richardson, John. *Nietzsche's System*. Oxford: Oxford University Press, 1996.

Richerand, Balthasar Anthelme. *The Elements of Physiology*. Translated by Robert Kerrison. Philadelphia: Hopkins and Earle, 1808.

Rolph, William Henry. *Biologische Probleme. Zugleich als Versuch einer rationellen Ethik*. Leipzig: Engelmann, 1882.

Rosenthal, David. *Consciousness and Mind*. Oxford: Clarendon Press, Oxford University Press, 2005.

Roux, Wilhelm. *Der Kampf der Theile im Organismus*. Leipzig: Engelmann, 1881.

Rukgaber, Matthew. "The "Sovereign Individual" and the "Ascetic Ideal": On a Perennial Misreading of the Second Essay of Nietzsche's On the Genealogy of Morality". *Journal of Nietzsche Studies*, 43 (2012): 213–39.

Ryle, Gilbert. *The Concept of Mind.* London: Hutchinson, 1949.

Schneider, Georg Heinrich. *Der menschliche Wille vom Standpunkte der neueren Entwickelungstheorien (des Darwinismus)*. Berlin. Dümmlers, 1882.

Schopenhauer, A. *Parerga and Paralipomena*, 2 vols. Zürich: Diogenes, 1851/1977. [Abbreviated as PP].

——. *The World as Will and Representation*. 2 vols. Translated by E. F. Payne. New York: Dover, 1969. [Abbreviated as WWR]

———. *Zürcher Ausgabe*, 10 vols. Edited by Arthur Hübscher. Zürich: Diogenes Publishing Co., 1977. [Abbreviated as ZA]

Sellars, Wilfrid. "Empiricism and the Philosophy of Mind". In Herbert Feigl and Michael Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, *Volume I: The Foundations of Science and the Concepts of Psychology and Psychoanalysis*, 253–329. Minnesota: University of Minnesota Press, 1956.

Shaw, Tamsin. *Nietzsche's Political Scepticism*. Princeton: Princeton University Press, 2007.

Solomon, Robert. "Getting Angry: The Jamesian Theory of Emotion in Anthropology". In *Culture Theory: Essays on Mind*, *Self*, *and Emotion*, edited by Richard Schweder and Robert Le Vine, 238–54. Cambridge: Cambridge University Press, 1984.

———. *The Passions*, New York: Anchor Press, 1976.

Spinoza, Baruch. *Ethics*, *Treatise on the Emendation of the Intellect*, *and Selected Letters*. Edited by Seymour Feldman. Translated by Samuel Shirley. Indianapolis: Hackett Publishing Co., 1991.

Tye, Michael. "The Experience of Emotion: An Intentionalist Theory". *Revue Internationale de Philosophie* 62 (2008): 25–50.

Weber, Ernst Heinrich. *Der Tastsinn und das Gemeingefühl*. Leipzig: Engelmann, 1846.

Welshon, Rex. *Nietzsche's Dynamic Metapsychology: This Uncanny Animal*. Basingstoke: Palgrave Macmillan, 2014.