AGENCY AND THE FOUNDATIONS OF ETHICS:

NIETZSCHEAN CONSTITUTIVISM

PAUL KATSAFANAS

DECEMBER 12, 2011
**TABLE OF CONTENTS**

Introduction .......................................................................................................................... 1

1. Three Challenges for Ethical Theory ............................................................................. 8

2. Normativity as Inescapability ....................................................................................... 65

3. Constitutivism and Self-Knowledge ............................................................................. 97

4. Constitutivism and Self-Constitution .......................................................................... 123

5. Action’s First Constitutive Aim: Agential Activity ..................................................... 156

6. Action’s Second Constitutive Aim: Power .................................................................. 207

7. The Structure of Nietzschean Constitutivism ................................................................. 266

8. The Normative Results Generated by Nietzschean Constitutivism ......................... 305

9. Activity, Power, and the Foundations of Ethics ............................................................. 343

References .......................................................................................................................... 351
Our experience of the world is pervaded by norms. Having promised to meet a friend for dinner, I feel *obligated* to do so. Upon entering a café on the heels of another customer, I think that he *should* have held the door for me. Rousing myself from a relaxing nap, I tell myself that I have *reason* to go to the gym. Watching the evening news, I judge that the criminal was *wrong* to murder his victim. As these examples illustrate, normative claims are ubiquitous. They inform our homeliest as well as most dire activities.

Despite their pervasiveness, normative claims are rather mysterious. They purport to have a certain authority over us: they tell us how we ought to live, or which actions we ought to perform, or which ends to pursue. But what justifies this authority? What makes it the case that we should keep promises, hold doors, or go to the gym? More momentously—what makes it the case that murder is wrong?

In its general form, this is the foundational question in ethics: how is the authority of normative claims to be justified? Recently, a great deal of attention has been directed at the idea that we might answer this foundational question by turning to the philosophy of action. According to a view that I will call *constitutivism*, action has a certain structural feature – a constitutive aim – that both constitutes events as actions and generates a standard of assessment for action. We can use this standard of assessment to derive normative claims. In short, the authority of certain normative claims arises from the bare fact that we are agents.

Thus, the great hope of constitutivism is that an investigation of the structure of agency will enable us to answer the foundational question in ethics. It will reveal why certain normative claims are justified. To see how this might work, consider an example. If you understand the nature of a game, such as chess, you thereby understand a host of normative claims that regulate chess players.
For example, part of what it is to play chess is to aim at checkmating your opponent. This aim simply must be present in order for a series of movements to count as an episode of chess playing. This aim therefore seems to generate a standard of assessment for chess players: a player is successful if she fulfills the aim. Moreover, the aim generates reasons for action: if the player sees that she can achieve checkmate by moving her rook to a certain space, then she has a reason to do so.

Constitutivists call aims of this sort constitutive aims. A constitutive aim is present in every instance of the activity-type that it regulates. It is present precisely because its presence is part of what renders the activity an instance of its kind. Or, in plainer language: if there is an activity, and participation in this activity requires having a certain aim, then all participants in that activity are going to have the aim. They are going to have the aim because otherwise, they wouldn’t be participants in the activity at all. For example, chess players have the aim of checkmate because, absent this aim, they wouldn’t be chess players at all.

Constitutivists hope to show that action itself has a constitutive aim. If we could manage this—so the hope goes—then we would be able to derive normative conclusions from that aim. Just as we can move from the fact that chess players aim at checkmate to the claim that (for example) they have reason to capture their opponents’ pieces, so too constitutivists hope that we can move from the fact that action has a constitutive aim to normative claims about what agents have reason to do. Indeed, the most ambitious versions of constitutivism attempt to show that all normative claims are ultimately derived from the constitutive aim of action.

In the following chapters, I will show that the attractions of constitutivism are considerable. If constitutivism works, then it will provide a way of justifying normative claims without positing irreducible normative truths or grounding norms merely in subjective, variable elements of human psychology. It will thereby avoid some central and longstanding problems in ethics, including the
so-called problems of practicality and queerness.

Unfortunately, constitutivism is not well understood. Explicit defenses of the theory are relatively new, having arisen only in the past two decades.¹ Accordingly, it has been difficult to determine what the essential elements of the constitutivist framework are. The first two chapters address this point, by examining constitutivism in isolation from any particular view about the nature of action. I explain what is essential to the constitutivist approach, and I show what would be necessary in order to derive normative claims from facts about the nature of action. In addition, I address a number of objections that have recently been leveled at the very possibility of a constitutivist theory. I show that these objections depend on misunderstandings of the constitutivist project, and can therefore be rebutted.

Thus, Chapters One and Two explain what constitutivism is and show that a constitutivist theory is possible. Of course, it is one thing to show that constitutivism could succeed, and quite another to show that it actually does succeed. The hardest part of any constitutivist theory is developing a conception of action that is both independently plausible and that yields a constitutive aim. Chapters Three and Four examine David Velleman’s and Christine Korsgaard’s attempts to do so. While Velleman’s and Korsgaard’s theories are extremely insightful, I argue that they succumb to a common problem: each theory can generate substantive normative results only by alternating between an excessively strong conception of agency and a much weaker conception of agency. The weaker conception of agency is all that Velleman’s and Korsgaard’s arguments establish, while the stronger conception is necessary for their normative conclusions to follow. Absent an argument for this stronger conception of agency, then, these theories are unsuccessful.

¹ While explicit defenses of constitutivism are new, some philosophers contend that aspects of the theory—described in different terminology—are present in the work of Plato, Aristotle, and Kant (see, for example, Korsgaard 2009). I will argue that aspects of constitutivism are also present in Nietzsche’s work.
If constitutivism is to succeed, we need a new conception of agency. So far, constitutivism has been strongly associated with Kantian theories of agency. As I will explain in the following chapters, Korsgaard’s version of constitutivism is avowedly Kantian, and Velleman describes his theory as “Kinda Kantian” (Velleman 2009). But, despite a resurgence of interest in Kant, I think it is fair to say that many philosophers believe that Kant’s theory of agency fails. The Kantian theory is subject to a staggering number of objections that strike many of us as decisive: the conception of agency upon which Kant relies seems empirically implausible; the moral theory depends on the assessment of maxims, and yet the idea of a maxim is terribly obscure and seems inconsistent; the arguments attempting to show that we are committed to the categorical imperative are dubious; and even if we could solve those problems, the categorical imperative itself seems to generate no substantive results. If constitutivism were bound up with the Kantian enterprise, it would inherit all of these difficulties.

But I will show that it is not. Kant was right about this: certain rules of practical reason are constitutive of agency. We can hold on to that idea while developing it in a non-Kantian manner and grounding it in a more plausible theory of action. In pursuit of that goal, Chapters Five and Six articulate a conception of agency that is indebted both to contemporary empirical work on human psychology and to Nietzsche’s philosophical arguments. Chapter Five begins this task. There, I show that in order to incorporate certain empirical facts about the nature of human agency, we must reject elements of the dominant philosophical conception of reflective agency. The dominant account, which I trace to Locke and Kant, distinguishes activity and passivity in agency, and treats reflective or deliberative acts as paradigmatic cases of agential activity. I argue that although we need a distinction between the active and passive in action, philosophical and empirical

---

2 I will explore the first objection in Chapter Five. There is a vast literature on the other three objections. For helpful introductions to these disputes, see for example Brewer (2002) on the idea of maxims, Williams (1986) on our alleged commitment to the categorical imperative, and Wood (1990) on Hegel’s argument that the categorical imperative generates no content.
considerations show that this distinction has nothing to do with whether the action was brought about in a reflective or deliberative manner. I defend a new account of agential activity, according to which an agent is active in the production of her action iff two conditions are met: (i) the agent approves of her action, and (ii) further knowledge of the motives figuring in the etiology of this action would not undermine her approval of the action. By drawing on a psychologically realistic account of motivation and agency, we can show that agents constitutively aim at this form of agential activity.

Thus, Chapter Five argues that action has a constitutive aim. But this aim, on its own, generates very little normative content. Its importance becomes apparent only when we link it to another aspect of agency.

To bring out this second aspect of agency, I turn to a largely untapped source of ideas about the relationship between agency and value: the work of Nietzsche. Nietzsche might seem to be an unpromising source for ideas conducive to the defense of constitutivism. After all, he is famously skeptical of ethical theorizing, and he flatly denies that there are any objective facts about what is valuable. However, as I will argue below, Nietzsche does offer ethical ideals of his own, and his critiques of traditional morality rely on the idea that a certain value – power, in particular – has a privileged normative status. I will suggest that Nietzsche claims a privileged normative status for this value precisely because we are committed to this value merely in virtue of acting. Nietzsche’s obscure claim that all actions manifest, and are to be evaluated in terms of, “will to power” is thus an attempt to move from a claim about the essential nature of action to a claim about value. Thus, surprising as it may seem, I will argue that we can use a Nietzschean claim about the constitutive features of action to derive a standard of success for action.

In short, we can use Nietzschean considerations to show that action has a second constitutive aim: power (this term is given a special technical sense, as I will explain below). In
defending this idea, I begin by considering Nietzsche’s baffling claims about revaluation. Nietzsche famously argues that we must “revalue” our values, critiquing them and in some cases replacing them with new values. I argue that Nietzsche’s revaluations are based upon the idea that power has a privileged normative status: power is the one value in terms of which all others values are to be assessed. If this is the correct interpretation of Nietzsche’s ethical theory, though, it raises a question: how could power have a privileged status, given that Nietzsche denies that there are any objective facts about what is valuable? I argue that Nietzsche’s account of psychology provides the answer: he grounds power’s privileged status in facts about the nature of human motivation. In particular, Nietzsche’s account of drives entails that human beings are ineluctably committed to valuing power. So Nietzsche’s ethical theory follows from his account of the nature of agency.

Thus, if we accept a Nietzschean account of agency—as I argue that we should—then power turns out to be a constitutive aim of action. But what exactly is “power”? We might suppose that valuing power denotes valuing conquest, mastery of others, and so forth. But this is not what Nietzsche means. Power is a term of art, for Nietzsche; he gives it a special sense. To will power is to aim at encountering and overcoming resistance in the course of pursuing other, more determinate ends. In other words, to say that we will power is to say that whenever we will an end, we aim not merely to achieve the end, but also to encounter and overcome resistances that arise in the pursuit of the end. For example, to say that I will power in the pursuit of writing this book is to say that I will not only to complete the book, but also to encounter and overcome challenges or resistances in the course of doing so.

The Nietzschean account of agency entails that all actions have this structure: whenever we will any determinate end at all, we also will to encounter and overcome resistance in the course of pursuing that end. Although this claim is paradoxical, I hope to show that it is supported both by compelling philosophical arguments and by recent empirical work on human psychology.
In sum, the Nietzschean theory of action leads to a bipartite constitutivist theory: our actions aim both at agential activity and at power. In Chapters Seven and Eight, I argue that this theory generates a range of substantive normative claims. I will show that while some of these claims conform to our ordinary thoughts about what there is reason to do, others are quite surprising. In particular, this constitutivist theory requires a reassessment of some of our most cherished values, such as the positive valuation that we place on certain forms of egalitarianism and the negative value that we place on certain forms of pain. I close by discussing the advantages that this Nietzschean version of constitutivism enjoys over competing ethical theories.

The hope, then, is that by drawing on Nietzsche’s theory of agency, we can answer the foundational question in ethics, showing how normative claims are justified. In particular, we can justify normative claims by showing that every agent aims jointly at activity and power.