

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/97006/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Robertson, Simon 2017. Rescuing Nietzsche from constitutivism. *Journal of Philosophical Research* 42 , pp. 353-377. 10.5840/jpr201739100

Publishers page: <http://dx.doi.org/10.5840/jpr201739100>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



RESCUING NIETZSCHE FROM CONSTITUTIVISM

Simon Robertson

Cardiff University

Contact Address: Philosophy, John Percival Building, Cardiff University, Colum Drive, Cardiff, CF10 3EU, UK.

Email: robertsons3@cf.ac.uk

ABSTRACT

Constitutivist theories in ethics seek to derive and justify normative ethical claims via facts about constitutive features of agency. In *Agency and the Foundations of Ethics: Nietzschean Constitutivism*, Paul Katsafanas uses Nietzsche to elucidate a version of the position he believes avoids worries besetting its competitors. This paper argues that Nietzschean constitutivism falters in many of the same places: it may remain vulnerable to ‘schmagency’ objections; it faces problems giving an account of the weights of reasons that adequately explains why we have more reason to perform some actions than others; and it is unable to generate normativity from constitutive aims. These doubts have wider import than Nietzschean constitutivism alone, though: they give good reason to think that such difficulties arise from the very structure of constitutivist approaches.

RESCUING NIETZSCHE FROM CONSTITUTIVISM

Constitutivist theories in ethics seek to derive and justify normative ethical claims via facts about constitutive features of agency. Despite numerous critics, constitutivism continues to attract adherents of various persuasions—most notably Kantian, but also, now, Nietzschean. In *Agency and the Foundations of Ethics: Nietzschean Constitutivism* (hereafter *AFE*), Paul Katsafanas uses Nietzsche to elucidate a version of the position he believes avoids worries besetting its competitors. In doing so he not only provides a wealth of insights into Nietzsche’s thought but, moreover, an innovative free-standing ethical outlook. This paper focuses on the latter. However, I’ll argue, despite some sophisticated maneuvers Nietzschean constitutivism falters in many of the same places as its more established siblings: it may remain vulnerable to familiar ‘schmagency’ objections; it faces problems giving an account of the weights of reasons that adequately explains why we have more reason to perform some actions than others; and it remains unable to generate normativity out of constitutive aims. If these doubts are sound they have wider import than Nietzschean constitutivism alone, though: they give good reason to think such difficulties arise not just from the content of particular constitutivist proposals but from the very structure of the approach. §1 introduces the basic constitutivist framework. §§2-3 present Katsafanas’ specifically Nietzschean version of it. §§4-5 critically assess the theory.

1. THE STRUCTURE OF CONSTITUTIVISM

Katsafanas distinguishes two “foundational” questions in ethics (*AFE* 1, 238):

E1. How are normative (ethical) claims justified?

E2. How are universal, authoritative normative (ethical) claims justified?

He construes *normative* claims broadly to include claims concerning what is right, wrong, good, bad, and more. Normative claims are *universal* insofar as they “apply to *all* agents”; and they are *authoritative* if they “apply to... agents, regardless of facts about their [particular] preferences, goals, and characters” (*AFE* 7, 238). A normative claim (concerning action) *applies* to an agent insofar as it entails a claim about what that agent has a normative reason to do (*AFE* 7). True universal and authoritative normative claims therefore entail true reason claims identifying reasons all agents have irrespective of their more particular

preferences, goals, and the like. Katsafanas believes that constitutivism can deliver such truths and thus answer both foundational questions.

Constitutivism is a two-stage theory. First, it gives an account of what action is via action's constitutive aims. Second, it derives normative claims from those aims. Katsafanas presents its central mechanisms via two theses (*AFE* 39):

(Constitutive Aim) Let A be a type of attitude or event. Let G be a goal. A constitutively aims at G iff

(i) each token of A aims at G, and (ii) aiming at G is part of what constitutes an attitude or event as a token of A.

(Success) If X aims at G, then G is a standard of success for X.

Constitutive Aim is a *descriptive* schema elucidating what it is for something to be a token of its type. If action has a constitutive aim G, Constitutive Aim implies that: A is an instance of action only if A aims at G. Success is an *evaluative-normative* schema intended to generate normative claims. If a token action A meets G, then A is a successful and hence good instance of action. If A aims at, but fails to meet, G, A may still be an action; but it is “defective” (bad, insufficiently good, etc.) with respect to G (*AFE* 39). From this, Katsafanas believes, we can derive normative reasons. Very roughly, aiming at G entails reasons to do things satisfying G.

To illustrate with an analogy: Suppose that a constitutive aim of playing chess is to win by checkmating one's opponent. Each instance of chess-playing thus involves aiming at checkmate, where so aiming is part of what constitutes certain activities (e.g. moving chess pieces) as playing chess. If your activities are not guided by that aim at all, you are not really playing chess. Chess' constitutive aim generates a standard of success. A chess-player is successful with respect to meeting chess' constitutive aim if but only if she checkmates her opponent. Her chess-playing actions may then be evaluated as good, bad, etc., relative to achieving chess' constitutive aim. Moreover, someone playing chess has good *reasons* to do things conducive to achieving that aim—to make moves enabling her to achieve checkmate. Similarly, if action itself has constitutive aim G, something counts as an action only if it aims at G; actions can be evaluated as good or bad relative to whether they achieve G; and agents have reasons to do things satisfying G. Additionally, just as anyone playing chess has reasons to do things conducive to achieving chess' constitutive aim, if G is a feature all actions share then anyone in the business of performing

actions—i.e. *each and every agent*—has reasons to do things satisfying G. Importantly, constitutivists also emphasize a crucial *disanalogy*. Chess is an optional activity. For agents, however, action is *non-optional* or *inescapable*. At least some reasons derived from action’s constitutive aims might therefore be inescapable and authoritative: these are reasons one has simply in virtue of being an agent, where their status as reasons does not depend solely upon one’s particular goals.

This gives the template for an answer to Katsafanas’ two foundational questions. Normative claims are justified insofar as they implicate reasons to do things meeting action’s standards of success. Some such reasons are universal and authoritative because they emerge from facts about action holding for all agents independently of one’s particular goals. The next two sections explain Katsafanas’ specifically Nietzschean constitutivism: first the account of action, then the normative theory constructed from it.

2. ACTION

Action has *two* constitutive aims: *agential activity* (AFE 110-44) and *will to power* (AFE 145-82). Agential activity concerns “the agent’s contribution to the production of action” (AFE 113). An agent may be more or less active with respect to the activities she performs. Actions proper, or fully-fledged actions, are behaviors the production of which she aims to determine through reflective choice and have control over.¹ Katsafanas analyses agential activity in terms of “equilibrium”, defined via two conditions: the agent approves of her act; and further knowledge of the motives figuring in her act’s etiology would not undermine her approval (AFE 138). Call the combination of these two conditions “stable approval” (AFE 142). The connection between action and stable approval emerges from the following thoughts: insofar as an agent reflectively determines and controls her behavior, it is likely to be something she approves of and is active with respect to, whereby the behavior is attributable to her; and, insofar as she would continue to approve of it were she aware of its motivational etiology, it is something she wholeheartedly approves, thus revealing a form of psychological integration or equilibrium. We then get action’s first constitutive aim:

- (C1) Your A-ing is an action only if: you aim to determine and control the production of your A-ing—where you manifest that aim iff, and to the degree that, (i) you approve of your A-ing, and (ii)

further knowledge of the motives figuring in the etiology of your A-ing would not undermine your approval of your A-ing.²

Will to power is a will to *encounter and overcome resistance* (AFE 157-62). Katsafanas explicates this via three ideas. The first rests on Nietzsche's conception of a *drive* (AFE 166-8). A drive is a psychological item that motivates or disposes to action. Drives have *general aims* directed at characteristic forms of activity and are expressed through more specific *goals* or *objects*. A creative drive, for instance, might aim to express itself through creative activity; the drive then takes particular goals/objects as occasions for expression, such as painting, musical composition, etc. Second, drive-motivated actions aim at power (AFE 168-71)—at encountering and overcoming resistance. A drive therefore disposes one to seek goals that enable its expression by overcoming the resistances those goals present. A drive is not sated by the attainment of its goal but reasserts itself once the goal is realized; drives aim at continual expression and are satisfied only when being expressed (only when the process they motivate is in progress) (AFE 169). Drive-motivated actions are thus behaviors expressing persistent aims to encounter and overcome resistance. Will to power, then, is not itself a drive but “*a description of the form that all drive-motivated actions take*” (AFE 170). Third, all actions are drive-motivated: all actions aim at encountering and overcoming resistance (AFE 171-6). This gives a second constitutive aim:

(C2) Your A-ing is an action only if: your A-ing aims at power (i.e. at encountering and overcoming resistance).

Four further clarifications.

First, Katsafanas emphasizes that actions *aim* at power. Even if agents do not aim at power explicitly (knowingly, consciously, etc.) whenever they act, they aim to do something that will encounter and overcome resistance. Additionally, the sub-agential drives which produce action have that aim (this follows from what it is to be an operative drive: something disposed to express itself through goals the realization of which overcomes resistance). In these extended senses, actions have the aim. Second, we are *not* motivated “to seek *all* forms of resistance” (AFE 176). That might imply, implausibly, that we continually seek enormous resistances that make everyday activities considerably more challenging—by chopping off our limbs, say. Rather, “[w]henver we act, we aim to encounter and overcome resistances that are related to the activity that we are performing” (AFE 176): resistances “that seem, to the agent, to

present appropriate objects for expression [for that] form of activity” (AFE 177). An agent typically seeks activities she thinks will present appropriate levels of resistance (e.g. challenging but achievable goals) and effective ways to overcome them (AFE 173-7). Third, actions can aim at and manifest varying degrees of power, depending on the nature of the resistances involved, how hard one tries to surmount them, etc. Thus, action’s constitutive aim of power is “differentially realizable” (AFE 200). Fourth, Katsafanas sets a *low bar* for encountering and overcoming resistance. Consider the following objection he raises to the claim that all actions aim at or manifest power. Suppose “I loaf on the couch and watch a lowbrow sitcom on television”; here, it may seem, “I am not encountering and overcoming any resistance” (AFE 181). Nonetheless, such activities could count as actions; for “there are resistances here, albeit of the most minimal sort: one must attend to the program, one must support oneself on the couch, one must resist competing desires that incline one to perform other actions, and so on (after all, loafing is marginally more demanding than non-action events such as sleeping)” (AFE 181). His strategy, then, is to set a low bar for aiming at and manifesting power. Indeed, *any* intentional activity satisfies C2 (AFE 54).

That is the nub of Katsafanas’ account of action: something counts as an action only if the consequents of both C1 and C2 are satisfied. Since the interest of constitutivism for ethics is what it delivers normatively, let’s now turn to the normative account.³

3. THE NORMATIVE ACCOUNT

The account so far suggests the following basic picture (which Katsafanas works up into something more sophisticated). There are two constitutive aims of action and two correlative standards of success: C1 and C2. So, if constitutive aims’ standards of success generate normative reasons, an agent would have reasons to do whatever satisfies the consequents of C1 and C2—anything she stably approves of that aims at power, i.e. *any action*.⁴

This is not as permissive as it may first seem. For the reasons a particular agent has are reasons for behaviors which, were they performed by *that* agent, would *count as actions*; and whether something counts as an action depends on what the *particular agent* stably approves of. Nonetheless, aside from restricting reasons to genuine actions, this basic picture allows that “anything goes”, for it supplies “no constraints on the *basis* of the agent’s approval” (AFE 145)—something Katsafanas thinks a *normative*

account should supply lest it reduce to an unfettered subjectivism.⁵ He therefore develops a more sophisticated account in which power and stable approval dynamically modify one another: stable approval circumscribes the power-directed activities an agent has reason to engage in, while will to power assesses and constrains the content of the agent's stable approval. The rest of this section presents the account's core ideas in three stages.

1. Katsafanas suggests we can move from agential aims to normative reason claims “provided that we grant the assumption that *aims engender standards of success*. More precisely, if you have an aim, you have a (pro tanto) reason to fulfill it” (*AFE* 184). Indeed, subject to the constraints signaled by C1 and C2 (reasons are to perform *actions*), “all aims generate standards of success... The aim itself—any aim—is reason-providing” (*AFE* 57). We thus get:

(Aim-Reason) For any (stable) aim ϕ you have and any action A: if your A-ing would fulfill ϕ , you have a reason to A.⁶

Two points about Aim-Reason. First, the reasons it implicates are *pro tanto*, a pro tanto reason understood as “a reason that has some weight, but nonetheless may be outweighed by other reasons” (*AFE* 184, n.2; also 192-3, 197-200). Second, we can distinguish two kinds of aim: constitutive aims of action; and more specific aims to realize relatively specific goals, such as winning a game, writing a book, loafing to relax, etc. It will be useful to refer to these as ‘constitutive aims’ and ‘goals’ respectively.

2. Katsafanas distinguishes two “sources” of reasons (connecting closely to action's two constitutive aims): *will to power* and *values* (*AFE* 187). Regarding will to power, “[i]f actions [constitutively] aim at encountering and overcoming resistance, and if aims are reason-providing... we have reason to seek those actions [or goals] that afford resistance” (*AFE* 184).⁷ Power is thus one source of reasons. As noted, every action constitutively aims at power; and since every action manifests at least some power, every action fulfils that aim to some degree. The claim that you have a reason to do what fulfils action's constitutive aim of power therefore fails to deliver much specific content. Nonetheless, appealing to values is intended to help. On Katsafanas' reading of Nietzsche, “values arise only through *valuing*”, where “to say that X is a value is just to say that we value X” (*AFE* 186). Agents *necessarily* value power (see below). Nonetheless, power is only one value (*AFE* 186, 188). Our other values are *not derived* from will to power (*AFE* 188-9) but emerge from our being valuing creatures embedded in a cultural setting (*AFE*

191, 196-7). What we value is closely related to C1: valuing something involves approving of it; and stably approving of an act typically signals agential activity (*AFE* 145). Importantly, what an agent values, besides power, shapes her more specific goals (*AFE* 187). And so, since reasons are aim-dependent, an agent's values and attending goals shape the content of her reasons in more determinate ways than power alone. Indeed, so long as you stably approve of your goal-related values and relevant ways to realize them, aiming at a goal the achievement of which would realize something you (stably) value gives you a reason to fulfill that goal (*AFE* 187).

Following Katsafanas, let's call reasons whose source is will to power 'power-derived reasons' (*AFE* 187). From Aim-Reason and C2, then: you have a power-derived reason to A *if* you aim at power and your A-ing would appropriately fulfill that aim. We could also talk about 'value-derived reasons': you have a value-derived reason to A *if* you stably value V and your A-ing would help fulfill some V-derived goal you have. We should nevertheless note that, since any action satisfies C1 and C2, any action an agent has a reason to perform aims both at power and at realizing the agent's value-derived goal. Thus, you have a reason to A *if* your A-ing would realize power and realize what you stably value.

3. There is a complication, though. Given that power and values present two "disparate sources" of reasons, "there is a potential for conflict" between these (*AFE* 187). Katsafanas seeks a way to resolve (or dissolve) such conflicts. (There are two reasons this is important: it might just be an adequacy-requirement on a theory that it says something about what one should do when the theory itself presents conflicting reason claims; plus, absent some way to resolve power-value conflicts, we would be left with the unconstrained injunction Katsafanas seeks to avoid—to do whatever we happen to stably value.) His strategy is to distinguish two roles will to power plays: it is not just a *source* of reasons, but also a *standard of evaluation* via which we can and should assess our other values (*AFE* 148-51, 187-9). We now need to explain the latter.

Katsafanas explicates conflicts between power and values in terms of conflicts amongst the content of reason *claims*: such claims conflict when they recommend conflicting courses of action. He allows for various types of conflict (*AFE* 191-9), which we can summarize as follows:

(Power-Value-Conflict) Value V conflicts with power if: (the content of the attitude involved in) your valuing V implies your acceptance of the claim that (a) there is a (value-derived) reason for you to A;

but will to power entails that either (b) there is not a (power-derived) reason for you to A, or (c) there is a (power-derived) reason for you not to A, or (d) there is a (power-derived reason) for you to B where you cannot both A and B.⁸

What should you do if your values conflict with power? “Usually”, Katsafanas notes, “when we discover inconsistencies between two values A and B, we have three options: rejecting A, rejecting B, or striking a compromise” (*AFE* 152). On the Nietzschean theory, however, “the only possible response is to reject the value that conflicts with power. In this sense, power has a privileged normative status” (*AFE* 152; see also 147-56, 187-197). Indeed, “any value that conflicts with will to power should be rejected” (*AFE* 187). Let’s mark this up:

(Privilege) Power has a privileged normative status—whereby if there is a conflict between your value V and power (and hence between value-derived and power-derived reason claims), you should reject V.

Power-value conflicts are thus resolved by rejecting power-conflicting values. The role given to power in Privilege also gives a way to critically assess our values and goals more generally. As Katsafanas puts it:

[P]otential acts are assessed along [three] dimensions: (1) The extent to which the act would present opportunities for encountering and overcoming resistance. (2) Whether the act is permitted, recommended, or forbidden by the other values we embrace... (3) The extent to which [our values] are compatible with will to power... So the Nietzschean ethical theory functions in the following way. First, an agent [assesses] her values in light of will to power. She sheds some values and embraces others. She then uses this new set of values, together with the valuation of power, to determine what she should do. (*AFE* 190-1)

The basic idea is that you should periodically assess your values and modify your evaluative set by selecting value-generated goals more expressive of power. This yields the following model of reasons:

You have a reason to A *iff*: your A-ing would realize power and realize what you stably value, and your A-ing does not conflict with will to power.

The thesis Privilege is therefore crucial to the overall account: it resolves apparent conflicts and supplies a constraint on which values we should embrace and act in light of.

With the main exposition complete, the next two sections argue that Nietzschean constitutivism faces serious difficulties. We turn in §5 to Katsafanas' meta-normative account: how constitutive aims generate normativity. §4 examines the first-order theory.

4. ASSESSING THE NORMATIVE ACCOUNT

Given its centrality to the theory, Privilege will be our predominant focus. The key issue is why we should accept it. In examining this, I'll argue that the theory falls prey to at least one of two objections commonly levied against other constitutivisms—"Schmagency" and "Why Bother?" objections—and, as a result, is unable to vindicate Privilege. To see what is at stake, Subsection A outlines these two objections. B then explains Katsafanas' argument for Privilege. C-E distinguish three ways to interpret that argument. The first two interpretations face both objections, however; while the third, even if it avoids the Schmagency Objection, remains subject to the Why Bother? Objection. F summarizes how these problems impact on Privilege and the theory as a whole. G raises a wider substantive concern emerging from these difficulties.

A. Two objections

First the Schmagency Objection (see *AFE* 53-6). Suppose that a constitutive aim of chess is checkmating your opponent while abiding by chess' rules. Ann and Bob start playing a game. Several moves in, Ann moves her knight diagonally, contra chess rules. Bob complains: "That's not a legal chess move". Ann responds: "I know but I don't care about that chess rule—I like to move knights diagonally". Bob counters: "Well you're not really playing chess". Ann says: "Fair enough: call it 'chess' or 'schmess'—this is how I'm playing". Furthermore, she claims, because she is playing schmess she has no reason to abide by the chess rule prohibiting moving knights diagonally—and, she adds, she has good reasons to do the things schmess constitutively involves.

Analogous thoughts generate a worry for constitutivism. Suppose that action constitutively aims at G and that two options, A-ing and B-ing, look similar in all respects *except* that A-ing aims at G whereas B-ing does not. So B-ing is not an action. Yet if you could (decide to) B, it looks as though you could (decide to) perform a non-action behavior instead of a genuine action. The question for constitutivism is

why you should (or have any reason to) perform the genuine action. Constitutivists can point out that B-ing is not an action. But, you might respond: “Call it ‘action’ or ‘schmaction’, I’m going to B”. This is a potential problem in two ways. First, constitutivists hold that the only behaviors for which there are normative reasons are *actions*, as specified by action’s constitutive aims. But if you can schmact, it is unclear why you could not have a reason to do so—or, if one prefers, a ‘schreason’, where a schreason is just like a reason except that it favors schmactions. Second, if on a given occasion you could have a (sch)reason to schmact and you don’t care about doing the thing constitutivists label ‘action’, it remains to be seen why you have any reason to perform the action. The possibility of schmactions thus casts doubt on the idea that reasons supposedly generated by action’s constitutive aims are *authoritative*.

Katsafanas responds to the Schmagency Objection by arguing that there can be no schmactions and hence no (sch)reasons to schmact. A further objection now looms, though: the Why Bother? Objection (*AFE* 63-7). We ordinarily think a person can have *more reason* to perform some actions than others. Constitutivists need to explain this. One obvious way to do so is by showing that the actions we have more reason to perform are those *more fully realizing* action’s constitutive aims. This rests on the idea that such aims are differentially realizable. Katsafanas criticizes extant constitutivisms (especially the broadly Kantian versions advanced by e.g. Velleman 2000; Korsgaard 2009) for failing to deliver a differentially realizable constitutive aim (*AFE* 68-108), arguing that they are thereby unable to explain why we have more reason to perform some actions than others. Nietzschean constitutivism yields a differentially realizable constitutive aim: power. Nonetheless, given that any action realizes that aim to some degree, we can still ask why one has more reason to realize this aim more fully—“why bother fulfilling [it] to the highest degree?” or “doing more than the bare minimum?” (*AFE* 207, 65). So the Why Bother? Objection raises a challenge to explain why, given a choice between realizing action’s constitutive aims more or less fully, you have more reason to do the former.

Whether Nietzschean constitutivism can deal with these two objections will be revealed by considering whether it can justify Privilege. We turn to this now.

B. The argument for Privilege

Imagine someone, call her Christine, who endorses a range of values Nietzsche opposes—“compassion, charity, equality, pleasure, the absence of suffering, altruism, peacefulness, and the like” (*AFE* 192). Suppose this set of values is internally coherent in the sense that, bracketing claims about will to power, coherence alone does not require Christine to reject them. Nonetheless, at least some of them conflict with will to power (*AFE* 192-6, 218-31). Privilege implies that Christine should reject such power-conflicting values. But what justifies Privilege?

Katsafanas’ answer is that we are *committed* qua agents to rejecting power-conflicting values. His argument is informed by two assumptions already encountered. First, power is an *inescapable* constitutive aim: “we are committed to valuing [aiming at] power” and “therefore have inescapable reasons for performing those actions that involve manifestations of power” (*AFE* 188). Second, we are not inescapably committed to valuing whatever other specific values we value; unlike power, these can be “reassessed and altered” (*AFE* 188). He then argues:

- (1) If an agent performs an action A-ing, she is committed to agential activity... to approving of her A-ing, and to having this approval be stable given further facts about A-ing’s etiology.
- (2) The etiology of every action includes will to power.
- (3) Thus, in order for an agent to be active, her approval must be stable given further facts about the ways in which will to power motivates her.
- (4) [So] the agent must approve [is inescapably committed to approving] of will to power as a motivating force. (*AFE* 207)

Therefore:

- (5) Agential activity commits any agent to disapproving of, and hence rejecting, any power-conflicting value.

In short, power has a privileged normative status because: as agents we are inescapably committed to agential activity and hence to approving of our will to power, and thereby committed to rejecting values and behaviors conflicting with our power; we thus have more reason to do things realizing greater degrees of power than things conflicting with will to power.

That is the argument for Privilege. Note, though, there are several ways one could interpret the crucial idea of *being committed* in (4) and thus (5). I'll first consider a non-normative reading followed by two normative interpretations, but argue that none vindicate Privilege.⁹

C. Non-Normative Interpretation

Interpreted non-normatively, the thought is that any agent necessarily values power and, therefore, *necessarily* would reject a power-conflicting value, at least insofar as she exercises her agency. Numerous *people* do *not* reject power-conflicting values, though. Indeed, the values Christine endorses—compassion, charity, equality, pleasure, absence of suffering, altruism, peacefulness, etc.—remain “predominant values within our culture” (AFE 192). The non-normative interpretation therefore has a strong implication: either Christine and the numerous people like her are not actually agents, or they fail quite systematically to exercise their agency fully. For simplicity, let's assume that Christine is an agent in the sense that she at least has the capacity to perform actions. Now it might *seem* that when Christine behaves in light of power-conflicting values she performs actions: given the low bar set for aiming at power, she appears to satisfy C2; she might also appear to successfully determine and control her behavior, wholeheartedly approving of it. Nevertheless, given that will to power is an inescapable part of who she is qua agent, and given that she endorses power-conflicting values, constitutivists could claim that she does *not* approve of her values in a *wholehearted* or *unified* way—hence her approval is *not stable*. She thereby fails to satisfy C1: the behavior she exhibits in light of her power-conflicting values does *not* count as (fully-fledged) action expressive of (fully-fledged) agency.

Although a strong implication, this is not incoherent. Even if we accept it, though, it does not vindicate Privilege since it does not explain why one *should* reject power-conflicting values. This is a *normative* matter. It is not explained by pointing out that people who behave in light of power-conflicting values are not performing (fully-fledged) actions. Given that people like Christine evidently fail to perform such actions, the normative question remains as to why they should perform actions expressive of fully-fledged agency and reject values inhibiting this.

It might be suggested that one should do so *in order to be* a fully-fledged or unified agent. But that invites a reiterated version of the question: why *should* one strive to be such an agent? (And merely

pointing out that one *is* an agent, or has the capacity to exercise fully-fledged agency, does not explain why one *should* do what one is not doing.) Note, further, that this non-normative interpretation is vulnerable to the Schmagency and Why Bother? Objections. For suppose that one can behave in light of power-conflicting values and hence intentionally behave in ways not aiming at action. In that case, constitutivism has not explained why on any given occasion one has reasons to perform only actions, rather than reasons (or schreasons) to engage in non-action behaviors expressing power-conflicting values (hence the Schmagency Objection). Or, if those behaviors count as actions, albeit not fully-fledged, so that one can have reasons to perform them, constitutivism needs to explain why one has more reason to perform fully-fledged actions (the Why Bother? Objection). Either way, the non-normative interpretation of *being committed* does not justify Privilege.

D. Normative Interpretation One

Constitutivists could nonetheless use aspects of the non-normative interpretation to address the normative question. Suppose again that Christine's accepting a power-conflicting value entails her acceptance of the (false) claim that there is a reason for her to A. What explains why there is *not* a reason for her to A is that A-ing is *not an action*. There is therefore more reason to perform actions than non-actions because there are no reasons to perform non-actions. That explains Privilege: you should reject power-conflicting values because they recommend behaviors that are not actions; and, qua agent, you are committed to, i.e. should be, performing *actions*. This would be a neat story. But it faces two difficulties.

First, it re-invites the Schmagency and Why Bother? Objections. Consider Christine. The values she endorses conflict with power. The behaviors she performs look very similar to actions; all they lack is the counterfactual feature specified by condition (ii) of C1. Thus, even if constitutivists reserve the label 'action' for behaviors satisfying C1 and C2, we can use 'schmaction' to denote behaviors like Christine's lacking feature (ii) in C1. We can then say there are reasons (or schreasons) for her to perform those power-conflicting schmactions—where (sch)reasons for schmaction work just like the constitutivist's reasons except they favor behaviors lacking the one counterfactual feature specified by (ii) in C1. Hence Normative Interpretation One fails to justify Privilege: agents could have (sch)reasons to engage in

activities (schmactions) informed by power-conflicting values; we then lack an explanation for why one is normatively committed to rejecting (i.e. should reject) those values.

There is a second problem. Given the low bar set for aiming at and manifesting power, behaving in ways recommended by power-conflicting values *does* aim at and manifest *some* modicum of power. One would thus be in a position to stably approve of one's behavior to some degree, thereby satisfying both C1 and C2 to some degree. Behaviors recommended by power-conflicting values could thus count as actions and be behaviors for which agents have reasons. Normative Interpretation One seeks to justify Privilege by saying such behaviors are *not actions*. So if they are actions, Normative Interpretation One fails to justify Privilege. The obvious option now would be to concede that behaviors motivated by power-conflicting values *are* actions, just not fully-fledged actions. That *might* avoid the Schmagency Objection (though see below). But since there could then be reasons to act in light of power-conflicting values, to justify Privilege constitutivism needs to show there is *more* reason not to; i.e. it still needs to answer the Why Bother? Objection. This takes us to a third interpretation of the argument for Privilege.

E. Normative Interpretation Two

According to this final approach, (a) behaviors motivated by power-conflicting values can be actions, but (b) agents should reject power-conflicting values because they are committed to aiming at greater realizations of power than acting in light of power-conflicting values allows. This subsection focuses mainly on point (b); but before doing so it is worth briefly considering (a).

One rationale for (a) is this: given that agents evidently do behave in light of power-conflicting values, allowing that such behaviors are actions gives a way to block the Schmagency Objection. This is part of Katsafanas' approach. More generally, he responds to the Schmagency Objection by setting the bar for meeting action's constitutive aims sufficiently low that it is impossible to schmact, whereby "action is *inescapable*" (AFE 54). He writes: "one cannot decide to perform a schmaction instead of an action... the very process of *deciding* or *trying* to produce a schmaction would itself be an action, and would therefore manifest action's constitutive aim"; therefore "[any] intentional activity... will count as an action... the idea that there could be a schmaction—an intentional activity that is not an action—is self-contradictory" (AFE 54). So, schmactions are impossible. In Christine's case, that she *decides* to behave in

ways recommended by her power-conflicting values renders what she does an action (so long as the decision manifests action's constitutive aim). Thus the Schmagency Objection "misses its mark" (*AFE* 54). This may be less conclusive than Katsafanas believes, though. First, it rests on the contentious view that mental events like decisions are themselves actions. Yet even if they are (e.g. because they manifest action's constitutive aims), it's not clear that action's constitutive aims as manifest in a decision *necessarily transmit* to the action they produce (I might stably approve of *my deciding* to A, but not stably approve of *my A-ing*). More significantly, it isn't obvious that someone who behaves in light of power-conflicting values must (decide to) do so in a reflective self-determining way. Christine might continue behaving in light of her power-conflicting value somewhat *unreflectively* (as one of the 'herd' about whom Nietzsche frequently writes). This looks like behavior which could fail C1, since she would not be aiming to determine or control her behavior for herself. (Or, even if she satisfies conditions (i) and (ii) in C1 to some degree, it's then not clear whether that really is sufficient for *self-determining* behavior.) If so, her behavior could fall short of action and count as schmaction. These worries do not deliver a knockdown case in support of the Schmagency Objection—the issues they raise require more detailed attention than there is space for here. Nonetheless, they cast doubt on how conclusive Katsafanas' response to the objection is. My tentative conclusion regarding the Schmagency Objection, then, is that Nietzschean constitutivism *may* remain susceptible to it. But *even if* it can avoid that objection, so that behaviors motivated by power-conflicting values are actions, it still needs to explain why we should reject, and not act in light of, power-conflicting values. Let's now turn to this, i.e. (b).

The claim is that agents should reject power-conflicting values because they are committed to aiming at greater realizations of power than acting in light of power-conflicting values allows; and that is because aiming at greater realizations of power *better fulfills* action's constitutive aims. This rests on two assumptions: that "there are different gradations of power", whereby "power is differentially realizable"; and that "we have more reason to realize the higher gradations than the lower gradations" (*AFE* 200). There is a caveat. The claim about 'more reason' applies to our *long-term projects* rather than each and every discrete action (*AFE* 202). This sensibly allows that "agents sometimes have [more] reason not to perform the action that best fulfills the constitutive aim" (*AFE* 198). As Katsafanas explains:

suppose that I can either loaf about and watch television (call this action A), or I can continue working on a difficult philosophical problem (call this action B). Action B generates far more resistance, and thus better satisfies will to power. However, having worked on the problem for several hours, I am tired and strongly motivated to take a break... In some cases of this form, the balance of reasons may favor A-ing. (*AFE* 198)

Nevertheless, “in the long run we typically have more reason to perform the action that better fulfills the constitutive aim” (*AFE* 199). So the basic thought—with analogues in contractarianism, indirect consequentialism, etc.—is that the specific actions we have most reason to do are those which best contribute (directly or indirectly, depending on context) to an agent’s maximally fulfilling action’s constitutive aim of power in the long-run. These ideas suggest the following three theses:

(Gradable): Action’s constitutive aim of power is differentially realizable: one’s actions (goals, projects, etc.) can aim at and manifest varying degrees of power.

(Weight) The degree of reason you have to A in a given situation (as specified by the weight of the reason or set of reasons favoring your A-ing) depends on the degree to which your A-ing would contribute to your realizing action’s constitutive aim of power in the long-term.

(More Reason) Given the projects you have as circumscribed by what you stably value, you have more reason to engage in one project (or set of projects) P_1 than some other project (or set of projects) P_2 *if and because*: P_1 enables you to realize action’s constitutive aim of power more fully in the long-term than P_2 .

This gives a maximizing view: the projects there is *most* reason to pursue are those which, given the aims you have, would *most fully* (i.e. *maximally*) realize action’s constitutive aim of power. What ultimately settles what you have most reason to do is the degree to which pursuing and realizing your projects expresses long-term will to power. The thesis More Reason thus explains Privilege: you should reject power-conflicting values because there is more reason to realize higher rather than lower gradations of power. The critical issue for justifying Privilege thus concerns whether constitutivism can justify More Reason.

Let’s start by asking why agents are committed to realizing action’s constitutive aim of power *maximally*, i.e. why we have more reason to realize greater gradations of power. For, as Katsafanas writes:

All it takes for something to be an action is for it to fulfill G to some extent. So, if an agent is committed to performing actions, she is committed to fulfilling G to some extent. But why bother fulfilling G to the highest degree? (*AFE* 207)

This is the Why Bother? Objection. His response, via his defense of More Reason, comes in two parts.

The first draws again upon the initial argument for Privilege (1)-(5) introduced in 4.B. The combination of C1 and C2 implies that “[m]erely in virtue of acting, we are committed to approving of will to power... if we don’t approve of will to power, we won’t be able to approve of any particular action at all” (*AFE* 208). “Consequently”, Katsafanas suggests, “I cannot simultaneously approve of fulfilling my aim [power] in every instance of action and ask why I should strive to fulfill it. In approving of it, I commit myself to fulfilling it” (*AFE* 208). Since the constitutive aim of power is differentially realizable, an agent’s commitment to this aim brings with it commitment to approving of actions that more fully (maximally) realize that aim. (The differential realizability of will to power is built into, as part of, what an agent is committed to by approving of will to power.) And, because will to power “is an aim to which agents are committed in every instance of action” (*AFE* 208), we are committed to approving of actions that maximally realize the aim of power in the long-run and thus have reasons to perform those actions. Let’s grant this conclusion:

- (i) Agents are committed to approving of actions that maximally realize the aim of power, whereby agents have reasons to pursue power-maximizing projects.

Note, as Katsafanas emphasizes (*AFE* 197-200), the reasons here—to do what will satisfy the constitutive aim of power maximally—are *pro tanto*, not necessarily overriding. (Indeed “the Nietzschean theory is only committed to the claim that the constitutive aim generates *pro tanto* reasons. In this respect, the constitutive aim is on par with all other aims” (*AFE* 198).) Thus understood, however, (i) does not justify Privilege. To justify Privilege, constitutivism needs to show not just that we have *some* reason to reject power-conflicting values—for then these reasons could be overridden by reasons generated by power-conflicting aims—but that we *should* reject power-conflicting values because we have *more reason* to realize power-maximizing values. Constitutivism therefore requires a stronger conclusion than (i)—namely:

- (ii) Agents are committed to *approving more fully* of actions that more fully realize the aim of power, whereby agents have *more reason* to pursue power-maximizing (than, say, power-satisficing) projects.

(ii) does not follow from (i). So even if Katsafanas' argument secures (i), we don't yet have a solution to the Why Bother? Objection or a full justification for Privilege.

Nonetheless, the second part to the argument, if successful, would support (ii). Katsafanas summarizes it by saying: "due to the omnipresence of will to power, systematically neglecting its [maximal] fulfillment will involve systematically ignoring reasons for action" (*AFE* 208). This suggestion—that we should pursue power-maximizing projects because pursuing sub-maximal power ignores other reasons—assumes we should not ignore those other reasons, presumably because there is more reason to do what those other reasons recommend. He explains this as follows: the omnipresence of will to power gives us more reason to engage in power-maximizing projects because it generates *more reasons* to do so. Consider the choice between pursuing a loafing life or your philosophical projects:

[W]e have a series of loafing actions, A_1 to A_n , making up a larger action [or loafing project], A . And, as an alternative, we have a series of working actions, B_1 to B_n , making up a larger action [or philosophical project], B . The Nietzschean claim is that for each A_i , I have a will-to-power-derived pro tanto reason *not* to engage in it, whereas for B_i , I have a will-to-power-derived pro tanto reason to engage in it. I also have other motives, so for some—but usually not all—of the A_i 's, I may have a motive and thus a reason to engage in it. However, if each reason is weighted equally, I will typically have more reason to B than to A . As this example indicates, in the long run we typically have more reason to perform the action that better fulfills the constitutive aim. It is not the weight of the constitutive-aim-derived reasons that generates that normative conclusion; it is their *ubiquity* and *pervasiveness*. (*AFE* 198-9)

So: for each A_i I have a power-derived reason not to A_i ; for each B_i I have a power-derived reason to B_i ; for at least some A_i 's, I have a value-derived reason to A_i ; but, because power-derived reasons are more ubiquitous than other reasons, I have more reason to B (pursue my philosophical projects) than to A (loaf my life away) because there are *more reasons* to B and not- A than to A and not- B . This faces several difficulties, however.

First, the account relies on the idea that, for each A_i , you have a power-derived reason not to A_i . This is crucial: there need to be enough power-derived reasons to make it the case that there are more reasons to B than A . Yet it is unclear why we should think there are power-derived reasons not to A . It

cannot be explained by the idea that A-ing *does not aim at or realize* power: for *every* action, loafing included, aims at encountering and overcoming *some* resistance. Probably, such reasons are supposed to arise from the fact that the A-ing project does not *maximize* power. But it is likewise unclear how that explains things. Suppose that the fact that B-ing *maximizes power* gives me a reason to B. However, it does not follow from this that there is *also* a reason not to A; and it is hard to see why one would think there is. In general, the fact which gives me a reason to do one thing *x* does not necessarily, or even typically, give me an *additional* reason not to do another thing *y*. Suppose that the fact that *x*-ing will maximally help Ann gives me a reason to *x*; that does not entail a reason not to help Bob. Not helping Bob may be a side-effect of helping Ann; but that doesn't show there is a *reason* not to help Bob. Yet, even if one insists that the reason to help Ann implies that there is a reason not to help Bob (because helping Ann implies not helping Bob, say), this does not seem to be an *additional* reason over and above the reason to help Ann. It is just a different description of the same reason, where acting upon that reason has two effects (helping Ann, not helping Bob). (To illustrate with an analogy: for utilitarians the fact that *x*-ing maximizes happiness, which gives me a reason to *x*, does not give me an *extra* reason *not* to do the second most happiness-conducive thing.) Nonetheless, for sake of argument let's grant that for each A_i I do have a power-derived reason not to do it.

Second, though, we're then left in the following position: since each loafing action would realize one of my values and manifest some power, for each A_i I have both a value-derived reason and a power-derived reason to A_i ; and since philosophizing would realize one of my values and manifest power, for each B_i I have a value-derived reason and a power-derived reason to B_i . So there are *just as many* power-derived (and value-derived) reasons to A as there are to B. Now add into the equation the power-derived reasons *not* to A. (These are the reasons the obtaining of which Katsafanas has not explained, but that we are granting anyway.) Then, though, I would also presumably have value-derived reasons *not to B*: for if my project is to loaf, and if philosophizing gets in the way of loafing, my values give me value-derived reasons not to philosophize; and so, since not philosophizing realizes power, for each B_i I have a power-derived reason not to B_i . Thus, with respect to each A_i and each B_i , there are just as many reasons for and against (whereby, with respect to A and B, there are just as many reasons for as against). Putting these calculations together, there are *just as many* value-derived reasons to A and not to B as there are power-

derived reasons to B and not to A. Therefore, the power-derived reasons to B and not to A are *not more ubiquitous*. So if there is more reason for me to B than to A, this is not explained by the greater ubiquity of power-derived reasons. Note, moreover, this result threatens to generalize in a potentially devastating way: the account fails to explain why we ever have more reason to perform some actions than others.

Third, the ubiquity account relies on the assumption that *each and every* possible reason to act has an *equal weight*. But that assumption is highly contentious (and under-motivated). Here is a standard way to explain why a set of reasons favoring an option x is more weighty than a set of reasons favoring another option y (whereby you have more reason to x than y): (i) individual reasons can vary in weight; (ii) the weight of a set of reasons favoring an option depends on the weight of the reasons comprising it; and (iii) what you have most reason to do is determined by the weights of the various sets of reasons for and against each option. It initially looked as though Katsafanas' use of Gradation would support a constitutivist reading of (i): the weight of a reason depends on the degree to which the action it favors realizes power, so that agents have *stronger* reasons to pursue projects more fully realizing power. However, the ubiquity account denies (i). It is thereby committed to denying the in-principle-possibility that reasons ever differ in weight (e.g. that a reason to help Ann right now because she is in desperate need could be stronger than a reason to help Bob out of a mild pickle or a reason to continue playing chess, etc.). That all reasons are equally weighty is a highly controversial claim standing in need of justification.

Normative Interpretation Two promised a way to vindicate Privilege by giving an account of weighting for reasons that explains why agents have more reason to pursue power-conducive than power-conflicting values. Absent a convincing justification for that, Normative Interpretation Two faces the Why Bother? Objection. Katsafanas' response rests on the plausible idea that power is a differentially realizable constitutive aim. The problem is that this differentially realizable aim does not license *differential weights of reasons*. This results in an independently questionable model of reasons' weights and, moreover, fails to explain why one has more reason to pursue power-maximizing goals. It is thereby unable to help constitutivism answer the Why Bother? Objection or explain why we should reject power-conflicting values as Privilege enjoins. Normative Interpretation Two therefore looks unable to justify Privilege.

F. Summary

These troubles run deep into Nietzschean constitutivism. Let's summarize how. Action has two constitutive aims: agential activity and will to power. Via Success and Aim-Reason, these generate normative reason claims from two sources: stable values and will to power. This opens up potential for conflicting reason claims. These are to be resolved via Privilege: power has a privileged normative status. Our central question was why we should accept Privilege. Katsafanas' argument for it relies on the idea that agents are *committed* to valuing power and rejecting power-conflicting values. I considered three ways to understand this. The Non-Normative Interpretation fails to explain the normative claim Privilege makes. Normative Interpretation One would explain this but, like the Non-Normative Interpretation, is susceptible to the Schmagency and Why Bother? Objections. I suggested that Normative Interpretation Two may remain subject to the Schmagency Objection. Yet even if it avoids that worry, it is unable to answer the Why Bother? Objection and thereby fails to vindicate Privilege. Thus the theory cannot resolve power-value conflicts in favor of power and cannot explain why we are required to critically assess our values in light of power. This leaves us with reasons to do anything that would count as acting, but no requirement to choose some actions over others, and hence an unconstrained license to act in light of whatever we happen to stably value. Hence the troubles run deep. This section concludes with a substantive worry falling out from these difficulties.

G. Substantive implications

Katsafanas suggests his account of the weight of reasons is best directed at discerning which *long-term* activities one has more or less reason to pursue. It is also better suited to evaluating activities and reasons *retrospectively* than prospectively: "to diagnose and correct courses of action and sets of values that we already embrace", by telling us "how to go on, not how to begin" (*AFE* 202, 203). So consider a long-term activity in progress. Fred is a ruthless serial killer. He doesn't care about much; but he values killing and the subsidiary activities this involves. Being a successful serial killer involves encountering and overcoming significant degrees of resistance over a long period. So, given that he stably approves of his murderous values and projects, there seem pervasive (power- and value-derived) reasons to engage in it.

Indeed, there so far looks little to rule out the inference that Fred could have had, and continues to have, *more reason* to pursue serial killing than to do anything else.¹⁰

Is this an objection? On one hand, Katsafanas thinks we should not expect a Nietzschean theory to deliver all the substantive results associated with orthodox moral outlooks—after all, Nietzsche is a critic of traditional morality (*AFE* 212, Ch.8). He also says that, although “there is a universally valid normative standard” generated by will to power, “the particular results generated by this standard vary across different types of individual” (*AFE* 218), depending on what they value. That Fred has more reason than not to pursue his murderous projects might just be an implication of, rather than objection to, constitutivism—an implication we are committed to if there is a persuasive case for constitutivism. On the other hand, though, Katsafanas emphasizes that Nietzsche does not condone vicious action *per se* (but advances an ethic focused on realizing higher forms of flourishing and creativity; *AFE* 212-4); and he is understandably reluctant to condone such conduct himself. We might then expect Nietzschean constitutivism, qua *ethical* theory, to say that Fred has more reason not to pursue his murderous projects than to do so.¹¹ So can it?

Although Katsafanas considers someone who might on a given occasion have most reason to perform “a particular action of murder” (*AFE* 201; 201-3), he does not directly address longer-term activities like serial killing. Nonetheless, one might appeal to the indirect model to question whether a project like Fred’s really involves overcoming maximal long-term resistance—perhaps by suggesting that Fred will likely get caught and imprisoned, and that incarceration will reduce his long-term opportunities to encounter and overcome resistances, whereby the balance of reasons overall favors not killing. I doubt this will succeed, though. For one thing, it implies that whether Fred has more reason to continue or stop his murderous activities is contingent on the likelihood of being caught. If, as matters have so far proven, the likelihood is sufficiently low, this does not attenuate the concern. Furthermore, suppose that Fred is caught and imprisoned after several years of successful killing. It’s unclear why we should consider the long-term effects on his ability to overcome resistance when assessing whether, in the successful years he had, he did something he then had (in)sufficient reason to do. Moreover, even if we do consider it, there are surely many opportunities to overcome considerable resistances in prison—by killing others, say. So, it again looks like Fred could have had more reason to pursue his project than not.¹² Such an indirect

strategy therefore looks precarious. Although an extreme example, the worry for constitutivism is that it generalizes rather too widely: to any range of activities we deem horrific (and that we might expect an ethical theory to forbid) but that some vicious people valorize.

The constitutive aims of stable approval and will to power therefore generate serious substantive qualms. Absent a persuasive account of the weights of reasons to assuage such concerns, Nietzschean constitutivism is on shaky ground. It would nonetheless have delivered a significant breakthrough if, despite these worries, it can show how we get normativity from constitutive aims. The next section considers this meta-normative matter.

5. FOUNDATIONAL ISSUES

A common objection is that constitutivist theories move illegitimately from purely descriptive (non-normative) claims about constitutive aims to normative claims. Part of the worry is that this appears to “violate that old Humean chestnut about deriving an ‘ought’ from an ‘is’” (*AFE* 60). More generally, if constitutivism cannot explain how constitutive aims generate normativity, it has not explained how we have any normative reasons. I’ll call this ‘the Unexplained Normativity Objection’, or UNO.

To see what is at issue, consider the following two ways to understand the relation between activities and reasons (*AFE* 48-9, 53):

- (1) If you participate in an activity α , the constitutive aim of α is reason-providing.
- (2) The constitutive aim of α is reason-providing iff you *have reason* to participate in α .

Katsafanas observes that constitutivism’s critics (e.g. Enoch 2011) usually assume that it endorses claims of the form given by (2) (*AFE* 49). However, Katsafanas advances a version endorsing claims of the form represented by (1), thereby avoiding problems attending (2). Although there are also concerns with (1), he believes he can avoid these. It will be useful to briefly outline worries with each.

According to a standard objection to claims like (1), the fact that you are participating in an activity with a constitutive aim (be it chess, serial killing, or whatever) does not show there is any normative reason for you to realize that aim. For if there is no reason for you to participate in the activity in the first place, the mere fact that you are participating cannot bootstrap into existence reasons to do what successful participation involves (that would generate the peculiar claim that you have reason to do

things conducive to realizing an aim there is no reason to realize). This reflects the worry that, by bootstrapping reasons into existence, we fail to explain where normativity comes from, i.e. UNO.

That worry is averted by (2). Yet (2) is also problematic. One difficulty is that constitutivism seeks to derive all reasons from action's constitutive aims. (2) implies that a constitutive aim generates a reason only if there is already a reason to engage in the activities regulated by that aim. This *more basic reason* is not a reason favoring some specific action, but a reason *to participate in the very activity of acting*—a reason to act “as such” (AFE 51-3). Constitutivism would then need to explain the obtaining of this more basic reason (AFE 49). But just as the mere fact that chess has a constitutive aim does not generate a reason to play chess, it is hard to see how the fact that action has a constitutive aim generates a *reason to engage in activities regulated by that aim*—a reason to act *as such*. If constitutivism cannot explain this, then since by (2) the obtaining of all other reasons (for specific actions) presupposes this more basic reason, it fails to explain how *any reasons* obtain. Katsafanas raises a further problem. This idea of a *reason to act as such* is nonsensical:

Asking whether there is reason to A presupposes that there is some alternative to A-ing... An agent who asks whether he has a reason to play chess is asking whether, rather than playing chess, he should perform some alternative action... But when an agent asks whether there is reason to perform any actions at all, the question must have a different sense. It is not as if the agent can do something *other* than performing actions; action is inescapable... [Thus] the question whether there is reason for action as such does not make sense... there is no such thing as a reason to perform action as such... the very idea of a reason to act as such is incoherent (AFE 52-3)

So, the idea of a *reason to act as such* is nonsensical because for agents action is *inescapable*; it makes no sense to ask whether agents have reason to participate in the very activity, action, to which they are inescapably committed.¹³ Katsafanas therefore rejects (2) and favors (1). Our question is whether there is a constitutivist way to read (1) that explains how we get normativity from constitutive aims. Katsafanas believes there is.

He makes three salient points. First, it is not the inescapability of *action* that generates normativity but the fact that action has constitutive *aims*. Thus, “[t]he move is not from ‘you inescapably ϕ ’ to ‘you should ϕ ’” (AFE 56). That would violate Hume’s dictum. Rather, “the aim-based version of

constitutivism moves from ‘you inescapably *aim* at ϕ -ing’ to ‘you should ϕ ’” (AFE 56). Second, though, “It isn’t the [aim’s] *inescapability* that is reason-providing. The aim itself—any aim—is reason-providing. The inescapability is just a point about how ubiquitous the aim is, not about why it is reason-providing” (AFE 57). The question therefore remains how non-normative aims, inescapable or otherwise, generate normativity (without violating Hume’s law). The third and key point is that *aims*, but not actions, have *standards of success*—and these are supposed to license the move from constitutive aims to normativity: “All that we need, in order for the constitutivist project to work, is the claim that aims in general generate standards of success” (AFE 57). More specifically:

All but the most ardent [non-]Humeans¹⁴ accept the claim that if you aim at A, you have reason to take steps toward realizing A. This principle—which I labeled Success...—is all that the constitutivist needs in order to generate normative results. So long as we accept Success, the mere fact that we have an aim does entail that we have reasons to fulfill it. (AFE 60-61)

This is peculiar, though. According to Success: if X aims at G, G is a standard of success for X. On a widespread view, a normative sentence either is or conceptually entails a normative reason claim (see AFE 7). Success makes no express mention of reasons. The thought must therefore be that Success entails claims about reasons. This is why I earlier suggested that constitutivism needs not just Success but also Aim-Reason. So recall Aim-Reason:

(Aim-Reason) For any (stable) aim ϕ you have and any action A: if your A-ing would fulfill ϕ , you have a reason to A.

Let’s grant that constitutivism generates normative results *if* we grant Aim-Reason. However, UNO is asking why we should accept a thesis like Aim-Reason. It seems to move from a non-normative antecedent to a normative consequent.¹⁵ The idea must be that there is something about standards of success that generate normativity and license Aim-Reason. We therefore need to consider whether constitutivism can get from the non-normative thesis Constitutive Aim, via Success, to the overtly normative thesis Aim-Reason. I’ll argue that it cannot. To do so, I’ll draw a threefold distinction between purely *descriptive* claims, first-order *evaluative* claims (couched via narrowly valoric concepts like *good* and *bad*), and first-order *normative reason* claims. I’ll develop the argument in three stages. First, I’ll show that the concept *success* can be understood descriptively or evaluatively, and that constitutivism has not entitled

itself to the evaluative conception it needs. Next, I'll reveal an ambiguity in the concept *success*, using this to explain why constitutivists might *think* Success generates evaluative claims. Finally, I'll show that even the relevant evaluative reading of Success does not support Aim-Reason.

Stage One. If G is a standard of success for X, X is successful only if (and to the degree that) X meets G. For action we then get:

(Action-Success) If action A aims at G, then: A is successful only if A satisfies G.

Because constitutivism claims that a standard of success is an evaluative standard, it needs to interpret Action-Success as making an evaluative claim. For instance:

(Action-Success^E) If A aims at G, then: A is good¹⁶ only if A satisfies G.

So, are constitutivists entitled to move from Action-Success to Action-Success^E?

There are two ways to understand the concept *success* in Success and Action-Success: *descriptively* or *evaluatively*. On a purely descriptive interpretation: a standard of success states a necessary (perhaps also sufficient) condition for achieving an aim; and being successful marks the fact that one has met that standard. Thus, if a constitutive aim of action is overcoming resistance, overcoming resistance is a standard of success for action—in the sense that *overcoming resistance is a necessary condition for realizing the aim of action*, i.e. for acting—such that you realize the aim of acting (i.e. you act) only if you overcome resistance. This does not make or entail any evaluative claims. It illustrates how the concept *success* deployed in schemata like Success and Action-Success can be used to make *purely descriptive* claims. Hence, a mere standard of success need not generate evaluative content. Constitutivism therefore needs to show that we *should* interpret Action-Success evaluatively (via Action-Success^E), by showing that standards of success generate evaluative content. Otherwise it fails to explain how we get evaluative content.

Stage Two. Constitutivism faces a problem getting to Action-Success^E, however. Consider the following thesis:

(Chess) If a constitutive aim of playing chess is checkmating your opponent: your chess-playing is successful only if you checkmate your opponent.

This is ambiguous. The embedded conditional could mean either:

(Chess₁) You successfully play chess only if you checkmate your opponent.

(Chess₂) You play chess successfully only if you checkmate your opponent.

Chess₁ is intended as a purely descriptive claim: you count as playing chess only if you checkmate your opponent (which is false). If Chess₂ makes an evaluative claim, it must be read as saying (something like): you played chess *well enough* only if you checkmated your opponent (which might be true). A similar ambiguity afflicts the concept *success* in the case of *action*. You could *successfully act*, in the descriptive sense that you realize the constitutive aims of action. But that does not entail that you *acted successfully*, in the evaluative sense constitutivism needs according to which your action was good. *Stage One* already argued that constitutivism has not yet shown that it is entitled to the evaluative reading. However, we can now see why constitutivists might think they are entitled to it: they surreptitiously move from descriptive claims about *successfully acting* (i.e. doing something that counts as action) to evaluative claims about *acting successfully* (doing something that is good). The question therefore remains whether constitutivism is entitled to these evaluative claims, as Action-Success^E requires. If it just assumes that it is entitled to schemata like Action-Success^E (without explaining how), it presupposes the very thing it needs to explain: how standards of success generate evaluative content.¹⁷

Stage Three. Suppose, though, that constitutivists could explain that. Even so, Action-Success^E does not mention reasons. So a further step is needed to reach Aim-Reason. One issue here is that it remains contentious whether *any* evaluative claims by themselves entail claims about normative reasons to act (many Humeans deny this, for instance (e.g. Williams 2001)). Significant as that is, suppose that at least some claims of the form ‘A is good’ imply claims about what particular agents have reasons to do. There nonetheless remains a further obstacle in getting from Action-Success^E to Aim-Reason.

Action-Success^E states that ‘A is *good* only if A meets G’. This was a shorthand way to capture evaluative content. But it is potentially misleading. The term ‘good’ in ‘A is good’ ordinarily functions as an evaluative predicate to denote a form of non-attributive value. However, constitutivism seems committed to treating ‘good’ as a predicate modifier to denote a form of attributive value. For recall that Constitutive Aim make a claim about what it is for a thing to be a token of a certain type: A *counts as an instance of action* only if A meets G. Action-Success^E would then license evaluative claims relativized to whether A meets G: claims like ‘A is a *good instance of action* insofar as A meets G’, i.e. A is a good G-instantiator. Here, ‘good’ functions as a predicate modifier identifying a form of attributive value. However, claims about attributive value do not entail claims about value in some further non-attributive

sense (and, it is commonly agreed, claims about attributive value can be analyzed in purely descriptive terms). Suppose we say “Fred is a good serial killer”: i.e. Fred is a good instance of the type *serial killer*—as far as serial killing goes, Fred is pretty effective at it. This does not imply that Fred or serial killing is good. Similarly with ‘good’ and *action*: sentence (i) “A is a good instance of the type action” does not entail (ii) “A is good” non-attributively. This is problematic for constitutivism. As noted, it is controversial whether sentences like (ii) imply claims about normative reasons to act. But suppose they do. Katsafanas has not shown how constitutivism is entitled to (ii). At best, he has shown that constitutivism is entitled to sentences like (i). But (i) does not make a normative (reason-implying) claim (it does not follow from the claim that “x is a good y” that there is a reason to realize x or to aim at y); nor does it entail (ii). Thus: (i) does not itself generate reason claims; even if (ii) does, (i) does not support (ii). So even if constitutivism were entitled to Action-Success^E, that does not vindicate Aim-Reason.

Katsafanas has not justified the move from non-normative constitutive aims to the normative thesis Aim-Reason. Maybe we can derive normative claims from constitutive aims, *so long as one grants* Aim-Reason. But constitutivism’s entitlement to Aim-Reason is precisely what UNO demands an explanation of. I’ve argued that the prospects for this explanation look bleak.

6. CONCLUSION

The last two sections argued that Nietzschean constitutivism faces significant difficulties. By way of conclusion I’ll draw out three wider implications.

First, Nietzschean constitutivism falters in many of the same places as other constitutivisms. This is revealing. It fuels the thought that such problems arise not just from the content of specific constitutivist proposals but constitutivism’s very *structure*: the possibility of non-action behaviors leaves constitutivism vulnerable to Shmagency Objections; even *if* Nietzschean constitutivism avoids that objection, like other constitutivisms it is unable to show that we have more reason to do things which more fully realize action’s constitutive aims—i.e. the Why Bother? Objection; and there remain serious doubts about whether the constitutivist apparatus really generates or explains normativity.

As a result, secondly, constitutivism fails to answer Katsafanas’ two foundational questions. E1 asked how normative (ethical) claims are justified. If constitutivism fails to explain how we get

normativity, it cannot answer this. It thereby fails to answer E2—how *universal* normative (ethical) claims are justified. Even if Nietzschean constitutivism had answered E1, though, the answer this implies for E2 is somewhat attenuated. It might have shown that every agent has reasons to do things which aim at power and express stable values. More specific normative claims would apply universally only if their content meshes with specific values stably approved of by every agent. Yet, in Nietzschean vein, Katsafanas denies there are any such values. Thus, specific normative ethical claims are not universal.¹⁸

Finally, although the preceding arguments haven't attended to any fine-grained Nietzsche-exegesis, if the philosophical case against Nietzschean constitutivism is sound, those of who think Nietzsche is important for contemporary ethics would do better to reject this constitutivist interpretation of him. Indeed, we should recall how Nietzsche himself (1886: §186) likens the search for such foundations in ethics to that inevitably doomed quest for the fantastical philosophers' stone.¹⁹

REFERENCES

- Enoch, David (2011). 'Schmagency Revisited', in *New Waves in Metaethics*, ed. M. Brady. New York: Palgrave Macmillan: 208-33.
- Huddleston, Andrew (Forthcoming). 'Normativity and the Will to Power: Challenges for a Nietzschean Constitutivism', *Journal of Nietzsche Studies* 47/3.
- Hurka, Thomas (2007). 'Nietzsche: Perfectionist', in *Nietzsche and Morality*, ed. B. Leiter and N. Sinhababu. Oxford: Oxford University Press: 9-31.
- Katsafanas, Paul (2013). *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.
- Korsgaard, Christine (2009). *Self-Constitution*. Oxford University Press.
- Nietzsche, Friedrich (1886). *Beyond Good and Evil*. London: Penguin (1990).
- Velleman, J. David (2000). *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Williams, Bernard (2001). 'Postscript. Some Further Notes on Internal and External Reasons', in *Varieties of Practical Reasoning*, ed. E. Millgram. Cambridge: MIT Press: 91-7.

¹ Katsafanas emphasizes that it is this *aim* which partly constitutes something as an action (*AFE* 62).

Although that might be necessary, it seems too weak: behavior you *aim* to control could then count as action even if you *lack* adequate control—something Katsafanas sometimes denies (*AFE* 134-6). One way to reconcile these strands (implied at *AFE* 62, 111) is to say that something counts as an action only if it satisfies the aim to some degree.

² Equilibrium is something you aim at whenever you act (*AFE* 111). Every action aims to satisfy the counterfactual (ii) because we aim to be unified agents, which involves approving of action's second constitutive aim of power (see esp. §4).

³ The account of action may nonetheless be problematic. Regarding C2, the “empirical” thesis (*AFE* 165) that all actions are drive-motivated (aim at power) risks being objectionably unfalsifiable in ways empirical theses shouldn't. Regarding C1, stable approval might itself be passive—i.e. something that happens to you, is not controlled by you—and hence not sufficient for agential activity. For other concerns, see Huddleston forthcoming.

⁴ Or any action *appropriate* to the kind of activity it is related to, e.g. by being a suitably challenging but effective way to pursue one's goals (*AFE* 177); this proviso can be assumed throughout.

⁵ He suggests that C1 by itself “doesn't seem to generate much normative content” (*AFE* 146), elsewhere making a similar point about C2 (*AFE* 185). His real worry, though, is that they each generate *so much* normative content which by themselves they are unable to critically adjudicate or constrain (*AFE* 145-6).

⁶ Katsafanas identifies this thesis with Success (*AFE* 57, 184). Aim-Reason must be a particular way of unpacking Success, though: they are not identical, since Success does not mention reasons—see §5.

⁷ The account focuses on reasons to maximally express *one's own* power (rather than, say, power in the world more generally). This raises questions about the overall shape of the theory—e.g. whether there are agent-neutral reasons. I suspect Katsafanas (and Nietzsche) would deny that (though see Hurka 2007); but I leave these matters aside.

⁸ Some brief clarifications: (1) We'll see in §4 that Katsafanas rules out *genuine* conflicts between (a) and (b): these are merely *apparent*, since a value-derived reason claim is true only if the value is something the agent *stably* approves of—where stable approval requires that the value mesh with power (thus, value-derived reasons recommend the same actions as power-derived reasons). (2) He variously suggests that

power-value conflicts are between what you *take yourself* to have reasons to do (AFE 193), the reasons you *take yourself* to have and the reasons there *actually* are (AFE 192), and the *content* of reason *claims* (AFE 189, n.17; also 194-5, 198-9); and, although he elucidates conflicts as *contradictions* (AFE 189, n.17; 194-5, 198-9), much of his discussion uses a conception weaker than (logical) contradiction (AFE 193-9). Power-Value Conflict is intended to accommodate these various ideas. (3) He also distinguishes cases of (i) “*internal* conflict” in which “valuing V entails taking there to be reason to ϕ , whereas willing power entails taking there to be reason *not* to ϕ ” (AFE 193), and (ii) “*external* conflict... the value [claim] entails that we have reason to perform some action A... [but] A-ing ultimately reduces our capacity to will power” (AFE 194-5).

⁹ Katsafanas has confirmed that he intends the third interpretation (which below I label Normative Interpretation Two). It is worth considering all three, though, to ward off alternative defenses and see what might push one towards the third interpretation.

¹⁰ A common Humean response to such cases (e.g. Schroeder 2007) is to (a) be permissive with respect to the range of reasons agents have, whereby we can have reasons to do horrific things, (b) argue that we should not trust our intuitions about when people do *not* have such reasons, and (c) supply an account of reasons’ weights ruling out worries about people having *most reason* to do those things. §4.E raised doubts about whether Nietzschean constitutivism could discharge (c).

¹¹ Katsafanas seems to want to resist substantively horrific implications when suggesting that constitutivism can condemn people like Fred on grounds that their dreadful activities conflict with what *we* value (AFE 234-5). (This doesn’t show that *Fred* has more reason to refrain from his projects, of course; it suggests *we* have reasons to deplore him, protect ourselves against him, etc.)

¹² It might depend on specifics of the penal system, including whether Fred will be isolated from others. But since what he has most reason to do is contingent on these things, it remains possible that he has most reason to pursue his murderous projects. Such thoughts generalize to anyone with deep-rooted tastes best fulfilled by overcoming resistances one easily finds in prison. Even if the penalty Fred faces were capital, it is unclear that this gives him more reason to refrain from his murderous endeavors: that could depend on whether there are less vicious alternatives Fred stably approves of, how strongly he values the risky life of serial killing he finds meaningful over a safer life he would find mundane, etc.

¹³ I'm likewise skeptical about *a reason to act as such*, though on different grounds to Katsafanas: the concept of *a reason to act* implicates a relation holding between some agent and some (more or less) specific action; since the substantive claim that an agent has a reason to act *as such* does not specify a specific action amongst its relata, it is hard to make sense of. Katsafanas sometimes articulates his own skepticism as follows: the question whether I have a reason to act as such does not make sense because it is asking "*whether I should perform any actions at all*" (*AFE* 52). This is ambiguous, though. It could be asking (i) whether I should perform *actions rather than schmactions*, or (ii) whether there are *any* actions I *should* perform. However, even if asking (i) were not to make sense, asking (ii) does: for we haven't yet established that there are *any normative reasons at all*. Katsafanas might be conflating these, assuming that since (i) is nonsensical so is (ii). At any rate, constitutivism cannot respond to UNO merely by moving from the claim that there are no reasons to act *as such* to the conclusion that there are normative reasons.

¹⁴ The text says "All but the most ardent Humeans...", but this must be a typo. Elsewhere, he suggests that Success is "relatively uncontroversial; even the most minimal accounts of practical reason, including most variants of the Humean account, accept it" (*AFE* 184). I doubt it's as widely accepted as Katsafanas urges (numerous non-Humeans and some Humeans deny it). Either way, the present issue is whether constitutivism is entitled to it.

¹⁵ Recall that it is the aim itself, not its inescapability or ubiquity, which is reason-providing. But even if there were something about inescapability that generates normativity, constitutivism still needs to explain how we get from inescapable non-normative aims to normative reasons; or if inescapability is itself normative, it needs to explain *that*.

¹⁶ Other valoric terms may be more appropriate in various contexts; I'll use 'good' as a placeholder to cover relevant alternatives, returning to the locution 'A is good' in *Stage Three*.

¹⁷ Katsafanas has suggested the following riposte: "I don't see this as an assumption but rather as a foundational move in the argument. The constitutivist is claiming that what it is for something to count as acting successfully is for it to realize the constitutive aim." The objection, however, is that this foundational move—from something being an effective instance of action (which can be specified non-normatively) to the claim that it is good in the normative, reason-generating sense—is what needs

justifying. I suspect our disagreement, ultimately, is whether constitutivism does need to justify this. *Stage Three* provides further grounds for thinking it does.

¹⁸ E2 also asked how the apparent *authority* of ethical claims is justified. Katsafanas' account implies that, although the truth of a specific reason claim does not depend *solely* on the agent's actual goals (etc.), nor are such claims true *regardless* of her (actual, stable) goals. Authoritative reasons are often understood as 'categorical reasons'—reasons one has irrespective of one's specific desires, goals, etc. So Nietzschean constitutivism would not have vindicated that form of normative authority.

¹⁹ Many thanks to *JPR*'s two (initially anonymous) referees, Paul Katsafanas and Andrew Huddleston; responding to their suggestions has helped improve the paper immensely. Thanks also to an excellent MA class at Cardiff University for our many discussions about Nietzschean constitutivism: Cameron Crook, John Davies, Matt Jenkins, Seb Jones, Osian Lewis.