

# Nietzsche's Theory of the Will

*Brian Leiter*

*School of Law & Department of Philosophy  
The University of Texas at Austin*

© 2007 Brian Leiter

<[www.philosophersimprint.org/007007/](http://www.philosophersimprint.org/007007/)>

IT IS WELL KNOWN that Nietzsche has provocative views about the nature of the will and free will.<sup>1</sup> It is less often appreciated that his views on these topics have considerable merit. Nietzsche not only anticipates and lends argumentative support to the new wave of non-libertarian incompatibilism defended by philosophers like Derk Pereboom (2001) and Galen Strawson (1994) — the view that free will is incompatible with “determinism” and that there is no credible account of free will as outside the causal order in the offing — but his theory of the will also wins some support from recent work on the will in empirical psychology (see Wegner 2002). As a philosophical naturalist, Nietzsche thought of his theoretical endeavors as proceeding in tandem with empirical inquiry (Leiter 2002: 6–11). As befits his self-designation as “the first psychologist”, it turns out that Nietzsche anticipated results that psychologists only arrived at a century later.

In section 124 of *Daybreak*,<sup>2</sup> Nietzsche sets out the primary issues that shall occupy us here in trying to understand his theory of the will. Nietzsche writes:

We laugh at him who steps out of his room at the moment when the sun steps out of its room, and then says: “I will that the sun shall rise”; and at him who cannot stop a

1. In discussing Nietzsche’s “theory of the will”, I am concerned with the notion of “will” familiar from general philosophy of action, both contemporary and historical, namely, the idea of a human faculty, whatever its precise character, that stands in some kind of necessary relationship with action. Such a faculty may itself be causally determined, or it may be autonomous of the antecedent causal order; its status may implicate questions of moral responsibility; and such a faculty may not exist at all. A theory of the will is one that sheds some light on these issues. My subject in this essay, I hasten to add, is *not* every use of the word “will” in Nietzsche’s corpus, since those uses are quite various, and many have little to do with the familiar philosophical issues about the will just noted.
2. References to Nietzsche will be to the standard English-language acronyms: *Daybreak* (D), *The Gay Science* (GS), *Beyond Good and Evil* (BGE), *On the Genealogy of Morality* (GM), *Twilight of the Idols* (TI), and *The Will to Power* (WP). I have generally started with the translations by Clark & Swensen, Hollingdale, and Kaufmann, and then made changes based on the Colli & Montinari edition of the *Sämtliche Werke*.

wheel, and says: “*I will* that it shall roll”; and at him who is thrown down in wrestling, and says: “here I lie, but *I will* lie here!” But, all laughter aside, are we ourselves ever acting any differently whenever we employ the expression “*I will*”?

I take it to be uncontroversial that this last question is rhetorical, and that the intended answer to this question is: “No, it is no different at all.” Should anyone be confused by the interrogative form Nietzsche employs (and, alas, some scholars have been so misled on other occasions),<sup>3</sup> other textual evidence will be adduced soon enough that makes it clear that when we act and say “*I will*,” it is no different, and no more ridiculous, than when he “who steps out of his room at the moment when the sun steps out of its room ... says ‘*I will* that the sun shall rise.’”<sup>4</sup>

If it is really true that this analogy holds, then it follows that the experience of willing which precedes an action does not track an actual causal relationship: the experience of willing is epiphenomenal (in some sense to be specified) with respect to the action. As Nietzsche notes in *The Gay Science*, “the feeling of *will* suffices for” a person “to assume cause and effect” (GS 127), but it will be the burden of Nietzsche’s argument to show that this assumption is mistaken. Thus, an adequate account of Nietzsche’s theory of the will and action will require us to get clear about three matters: first, the phenomenology of “willing” an action, the experience we have which leads us (causally) to conceive

3. See Owen & Ridley (2003: 70–71), who apparently confuse Nietzsche’s rhetorical questions with genuine questions, perhaps because of a propensity to read Nietzsche too literally, conjoined with inattention to the other kinds of textual evidence discussed in Leiter (1998, 2002), as well as in this essay. I have, in fact, changed my view about the nature of Nietzsche’s epiphenomenalism, although not about his fatalism. I return to Owen & Ridley’s criticisms, below, in the long note 11.
4. Nietzsche himself makes the point clear just a few sections later in the same work, noting “that we really are not responsible for our dreams — but just as little for our waking life” (D 128).

of ourselves as exercising our will (to say “*I will*”); second, Nietzsche’s arguments that the experiences picked out by the phenomenology are not causally connected to the resulting action (or, alternatively, not causally connected in a way that would underwrite ascriptions of moral responsibility); and third, Nietzsche’s account of the actual causal genesis of action. On the latter score, we shall turn to some recent work in empirical psychology that, in fact, vindicates precisely Nietzsche’s skepticism that our “feeling” of will is a reliable guide to the causation of action.

### The Phenomenology of Willing

Nietzsche recognizes that we often *feel* as if we are exercising free will, but he is unusual among philosophers in scrutinizing that experience, breaking it into its component parts. The resulting account is admittedly a *revisionary* one — it is not ultimately designed to *vindicate* the epistemic reliability of the feelings involved — but it is, Nietzsche claims, the correct account of those feelings. The key discussion comes in section 19 of *Beyond Good and Evil*, whose account of the phenomenology of willing bears quoting at length:

[I]n every act of willing there is, to begin with, a plurality of feelings [*Gefühlen*], namely: the feeling of the state *away from which*, the feeling of the state *towards which*, and the feeling of this “away from” and “towards” themselves. But this is accompanied by a feeling of the muscles that comes into play through a sort of habit as soon as we “will,” even without our putting “arms and legs” into motion. Just as feeling — and indeed many feelings — must be recognized as ingredients of the will, thought must be as well. In every act of will there is a commandeering thought, — and we really should not believe this thought can be divorced from the “willing,” as if some will would then be left over! Third, the will is not just a complex of feeling and thinking; rather it is fundamentally an *affect*

[*ein Affekt*]; and specifically the affect of the command. What is called "freedom of the will" is essentially the affect of superiority with respect to something that must obey: "I am free, 'it' must obey" — this consciousness lies in every will. ... A person who *wills* —, commands something inside himself that obeys, or that he believes to obey.

Let us use an example to flesh out Nietzsche's account. Sitting at the computer, I wonder whether I should go downstairs to see what the children are doing. I "decide" to do so, and so begin to rise from my chair. I *feel* as if I have willed the movement: I feel the *moving away* from the desk and computer, the *moving towards* the door, and I *feel* the physical, or muscular, movement as well. Let us call all this complex of feelings, for ease of reference, "the bodily feelings".

These bodily feelings are not, however, sufficient for the experience of will: they are merely *qualitative*, merely the "raw" feeling of "away", "towards", of muscles contracting, limbs moving. We still need the "commanding thought" — that's Nietzsche's label, which I will use in what follows — namely, the thought, "I will get up from the desk and go downstairs" or some suitable surrogate. But the bodily feelings and the commanding thought are still not enough, according to Nietzsche, for the experience of willing. This is perhaps Nietzsche's key claim. For the experience of willing is, according to Nietzsche, essentially the *meta-feeling* — the "affect" Nietzsche calls it — of commanding. "Affect" is not being used here in its Freudian sense of psychic energy or "charge" but as something closer to feeling again. By the "affect of the command", Nietzsche means the *feeling* that the *thought* (*i. e.*, the propositional content, such as "I will get up from the desk and go downstairs")<sup>5</sup> brings about these other bodily feelings, *i. e.*, of "away from", "towards", of, in a word, movement; and that this commanding is *who I am*. By *identifying* with the commanding thought — by taking that to be "who I am" (on this occasion) — we *feel* superior, we

5. It might be more apt to put the propositional content in the imperative form, so that it does not sound merely like a prediction!

experience this affect of superiority. So a person has the experience of willing when he identifies himself with a certain propositional content (the commandeering thought "I will get up from the desk and go downstairs") that he takes to be commanding the bodily feelings, *i. e.*, the feelings that attach to the "towards", "away from", and the muscular sensations; and this identification produces the meta-feeling of superiority which is the feeling of willing. In short, one experiences willing when one *feels* as if the bodily qualia are obeying the thought, and that the commanding thought is who I am.

As Nietzsche recognizes, there is something paradoxical here, since, as he observes in the same section of *Beyond Good and Evil*:

[W]e are, under the circumstances, both the one who commands *and* the one who obeys, and as the obedient one we are familiar with the feelings of compulsion, force, pressure, resistance, and motion that generally start right after the act of willing. On the other hand, however, we are in the habit of ignoring and deceiving ourselves about this duality by mean of the synthetic concept of the "I." (BGE 19)

In other words, talk of "I" — as in "I will go downstairs" — obscures an elemental fact: it's my body that is "obeying" my will, so "I" am also one who obeys, as well as one who commands. Of course, we don't experience it, or think of it, that way: we identify the "I" with the feeling of *commanding*, not the feeling of obeying. Thus, our experience of the "I", our identification of it with the commanding thought, itself requires an explanation: *why* do we not identify ourselves with the *commanded* feelings and movements? Why do we instead identify ourselves with the superiority of the *commanding thought*?

Here is what Nietzsche offers by way of explanation in the same passage:

[T]he one who wills believes with a reasonable degree of certainty that will and action are somehow one; he attributes the success, the performance of the willing to

the will itself, and consequently enjoys an increase in the feeling of power that accompanies all success. "Freedom of the will" — that is the word for the multi-faceted state of pleasure of one who commands and, at the same time, identifies himself with the accomplished act of willing.<sup>6</sup>

We identify, then, with the *feelings of command* rather than those of obedience because identifying with the former increases pleasure (it is pleasant to feel the body "obeying") — so the explanation might seem straightforwardly hedonistic. But this would be too quick, since the real explanation, as often happens in Nietzsche, is cast in terms of feelings of power which, *in turn*, produce pleasurable sensations. We need not resolve the issue here of the primacy of the desire for pleasure or desire for power as the fundamental explanatory mechanism;<sup>7</sup> for the phenomenology, all that matters is that it be true that there *is* a feeling of pleasure attendant upon the sensation of willing, *even if* that feeling derives from a feeling of power.

So now we have Nietzsche's account of the *phenomenology* of willing from *Beyond Good and Evil* (and I should add that I know of no more systematic account in the corpus that conflicts with it): we *feel* as

6. What then of *unsuccessful* actions, which presumably one can also experience as being freely willed, although not "accomplished"? Suppose, for example, I decide to get up from the computer to see what the children are doing, but fail to do so (perhaps one of my kids has glued me to the chair!). Although the action is unsuccessful, all the requisite components for the experience of free will are present (*e.g.*, I can feel the requisite bodily movements, even if they are not brought to completion). Trickier cases involve (as a referee pointed out) extreme fatigue or paralysis, though at least in the former case, it is not implausible that, as with the unsuccessful action, I can feel some distinctive muscular sensation, even if I do not move my limbs. Nietzsche's proposed phenomenology may not work for the paralyzed, but we would have to know more about what, if any, experience they really have of willing the motion of paralyzed limbs.
7. In *The Gay Science*, Nietzsche offers the apparently stronger claim on behalf of hedonism, that for all willing "a representation of pleasure and displeasure is needed" (GS 127). Unfortunately, there is no elaboration of the thought, and so it is not clear that the *representation* might not itself be prompted by an experience of *power*: in any case, nothing in the discussion there rules it out.

though we are exercising free will when we *identify* with the "commanding thought" which we feel is superior to, and being obeyed by, the myriad qualitative experiences involved in movement, that is, the bodily feelings; and we so identify because of the feelings of pleasure and power that arise from the "affect of superiority" that flows from that identification.<sup>8</sup>

### Why the Phenomenology Does Not Track a Causal Relationship

The crucial idea in Nietzsche's theory of the will is that the phenomenology of willing, no matter how vivid, does *not* in fact mirror or reflect or — as I shall say in what follows — *track* an actual causal relationship (or, more precisely, a causal relationship sufficient to underwrite ascriptions of moral responsibility, a crucial qualification to which we'll return). That is, the commanding thought, with which we identify because it gives us a feeling of superiority, is not in fact identical with anything that actually stands in a causal relationship with the resultant action.

As Nietzsche writes in the long section from *Beyond Good and Evil* that we have been scrutinizing:

[T]he one who wills [*i. e.*, who has the experience of willing] believes, in good conscience, that willing *suffices* for action. Since it is almost always the case that there is will only where the effect of command, and therefore obedience, and therefore action, may be *expected*, the *appearance* translates into the feeling, as if there were a *necessity of effect*. In short, the one who wills believes with a reasonable degree of certainty that will and action are somehow one; he attributes the success, the performance of the willing to the will itself, and consequently enjoys an increase in the feeling of power that accompanies all success ... (BGE 19)

8. In what follows, I am going to take for granted the accuracy of Nietzsche's account of the phenomenology.

To put it in quasi-Humean terms, Nietzsche claims that because the complex of bodily feelings, commandeering thought, and meta-feeling are fairly constantly conjoined with succeeding bodily actions, we naturally infer (since it increases "the feeling of power") that the will has caused the subsequent actions. How, then, has the phenomenology misled us, according to Nietzsche?

Remember that the experience of will for Nietzsche has three components: the bodily feelings; the commandeering thought; and the meta-feeling of superiority, *i. e.*, the feeling that the thought commands everything else. It is the meta-feeling (perhaps the pleasure attendant upon the meta-feeling) in turn that leads us to identify with the thought, rather than with the parts of our body that are commanded.

In debunking the phenomenology as a reliable guide to causation, Nietzsche's target is the *commandeering thought*, rather than the bodily feelings or the meta-feeling. And his argument is brilliantly simple. He starts from another bit of phenomenology, namely that, "a thought comes when 'it' wants, and not when 'I' want" (BGE 17). Nietzsche's target in this *particular* passage is the famous Cartesian doctrine, "I think, therefore I am." But, Nietzsche points out, from the fact that there is thinking, it does not follow that *I, i. e.*, some subject or agent, is doing the thinking, and so it does not follow that I exist. As Nietzsche puts it:

It is ... a *falsification* of the facts to say that the subject "I" is the condition of the predicate "think." It thinks: but to say the "it" is just the famous old "I" — well that is just an assumption or opinion, to put it mildly, and by no means an "immediate certainty." In fact, there is already too much packed into the "it thinks": even the "it" contains an *interpretation* of the process, and does not belong to the process itself. (BGE 17)

Even if the explicit target in this particular passage is the Cartesian "I", the surrounding context makes clear the real target, namely, the

will. The preceding section of *Beyond Good and Evil*, for example, treats the "I think" and the "I will" as both being common, but mistaken, examples of "immediate certainties"; while the following section returns explicitly to an attack on "free will", only to be followed by the long passage, section 19 of *Beyond Good and Evil*, on which we have been concentrating.

Now what does Nietzsche's phenomenological claim here mean? What does it mean to say a thought comes when "it" wants, not when "I" want? Because we are talking about thoughts that "come", I take it Nietzsche must mean thoughts that come to consciousness. And his point is that our "thoughts" appear in consciousness without our having willed them: "*ein Gedanke kommt, wenn 'er' will, und nicht wenn 'ich' will.*" We need to be careful here since, after all, Nietzsche is engaged in an attack on the existence of will, so he can't believe that there is any sense in which I could genuinely will a thought into existence. But he can't, of course, presuppose that conclusion here without begging the question. Rather, we must take the talk of willing here to refer to the *experience of willing*, which Nietzsche, as we have seen all along, concedes is real enough.

So Nietzsche's phenomenological point then comes to this: a "thought" that appears in consciousness is *not* preceded by the *phenomenology of willing* that Nietzsche has described, that is, there is no "commandeering thought" preceding the conscious thought to which the meta-feeling (the affect of superiority) attaches. (Even if there were such a commandeering thought in some instance, this would just create a regress, since not every commandeering thought will be preceded by the experience of willing.) Since we do not *experience* our thoughts as willed the way we experience some actions as willed, it follows that no thought comes when "I will it" because the experience to which the "I will" attaches is absent.<sup>9</sup>

9. I agree with Katsafanas (2005: 11–12) that BGE 17 does not support the epiphenomenality of consciousness *per se*, as I had wrongly claimed in Leiter (1998), but it does, as I argue here, support the epiphenomenal character of those experiences related to willing.

Notice the rather clever structure of this argument, for its critique is entirely *internal* to the perspective of the agent who takes himself to possess a will. For what Nietzsche does is point out that the criterion of willing that agents themselves treat as reliable guides to a causal relationship — namely, the phenomenology described above — is, in fact, completely absent in the case of thoughts (or, at least, in the case of the thought that starts an inferential chain of thinking which involves the experience of willing). As an introspective matter, it seems to me Nietzsche is plainly correct about this point. But if we do not experience our thoughts as willed, then it follows that the actions that follow upon our experience of willing (which includes those thoughts) are not caused in a way sufficient to underwrite ascriptions of moral responsibility.

That conclusion came rather quickly, so let us pause for a moment and try to set out more clearly how we got there. Nietzsche's conclusion — that our experience of willing does not, in fact, make us morally responsible for our actions — requires two premises. The first premise is this: one component of the experience of will — namely, the commandeering thought — is, itself, causally determined, at some point, by something other than the will, since thoughts that come into consciousness are causally determined by something other than the will.<sup>10</sup> (What the "something else" is we shall return to.) The second premise — the one so far unstated and unargued for in our discussion of Nietzsche's theory of the will — is that being self-caused (*causa sui* as Nietzsche calls it) is a necessary condition for responsibility (see BGE 21).<sup>11</sup> Since we have shown that the "commandeering thought"

10. I am going to speak of "caused" and "causally determined" interchangeably, without meaning to prejudge issues about probabilistic causation — issues that, in any case, were unknown to Nietzsche.

11. Owen and Ridley object (2003: 73–74) that the rest of the passage (BGE 21) undermines the point noted in the text for two reasons. First, they write, "it must be a mistake to attribute to Nietzsche ... the view that the will is 'unfree' in the sense of being causally determined, since he explicitly rejects that position" (73). But in what sense precisely does Nietzsche "reject" that claim? Owen and Ridley omit the relevant text which, as I discussed in Leiter (2002: 22–23), makes clear that Nietzsche's *apparent* skepticism has nothing to do

with whether *the will is causally determined*, but with whether *anything* is causally determined: and on this score, Nietzsche is still, in BGE 21, in the grips of the neo-Kantian view (acquired from his reading of Friedrich Lange) that "cause and effect" are merely features of the phenomenal world, not of "things in themselves". If Nietzsche abandons this neo-Kantian view — as even Owen and Ridley acknowledge he does (2003: 74) — then Nietzsche's argument against being *causa sui* is unaffected: if nothing in the "phenomenal" world can be self-caused, and the modifier "phenomenal" is doing no work, then it follows that nothing can be "self-caused" *simpliciter*. Owen and Ridley, in response to my comments on an earlier draft of their essay, purport to acknowledge this point, which brings us to their second objection, which warrants quoting (2003: 74):

[I]t does not follow ... from Nietzsche's rejection of the *causa sui* that he instead either embraces or must embrace full-blown classical determinism, since, as Nietzsche makes clear, he regards the *causa sui* and classical determinism as symptoms of precisely the same "superlative metaphysical nonsense. ... [Even granting Nietzsche's rejection of Lange's neo-Kantian skepticism concerning the reality of causation], if, as Leiter accepts, the argument against the *causa sui* as a piece of "superlative metaphysical nonsense" still holds once the Neo-Kantian view has been abandoned, then so too does the argument against classical determinism. That Nietzsche later came to accept the reality of causation, in other words, has no effect on the logical point that we draw from the passage, namely, that the opposite of a piece of nonsense is itself a piece of nonsense.

This response is problematic at a number of levels. First, I certainly did not claim that Nietzsche embraces "full-blown classical determinism" — indeed, I am explicit about distinguishing classical determinism from Nietzsche's fatalism (1998: 224–225; 2002: 82–83) — though a separate misunderstanding (to which I'll return in a moment) may explain why Owen and Ridley think the latter collapses into the former. Second, Nietzsche does not speak of "superlative metaphysical nonsense"; he says rather that the idea of the *causa sui* reflects a "longing for 'freedom of the will' in the superlative metaphysical sense", that is, in contemporary terms, some kind of libertarian free will. Even allowing that Nietzsche deems the idea of the *causa sui* nonsensical — he does say it "is the best self-contradiction that has even been conceived" even though he does not call it "superlative metaphysical nonsense" — it simply does not follow, logically or otherwise, that an opposed doctrine is also nonsensical, unless that opposed doctrine shares the premise that renders the first self-contradictory. But what makes the idea of the *causa sui* incredible is the idea that one can "pull ... yourself by the hair from the swamp of nothingness up into existence" (BGE 21), and *that* is precisely the idea that is repudiated by the view that the will is causally determined. Nietzsche's purported argument against this latter doctrine depends *entirely* on the neo-Kantian skepticism about causation that even Owen and Ridley concede he later abandons. This,

that is part of the experience of will is not *causa sui*, it follows that the will it helps constitute is not *causa sui*, and thus any actions following upon that experience of willing could not support ascriptions of moral responsibility.

Now, Nietzsche, it must be conceded, simply takes for granted that free will and moral responsibility are incompatible with causal determination of the will. Why he takes this incompatibilism for granted is not hard to see, though: it is deeply embedded in ordinary moral and religious thought. As Galen Strawson has noted, the incompatibilist idea of responsibility “has for a long time been central to the Western religious, moral and cultural traditions” (Strawson 1994: 8). Arguably, it is only certain philosophers who think the need to be a self-caused agent is superfluous, something that can be finessed via some adroit dialectical moves. (And even among philosophers, of course, discontent with compatibilism is widespread: see, *e.g.*, Nagel [1986: 113] for a striking example.)

The “argument from the phenomenology of thoughts” (as I shall call

---

in turn, explains precisely why there was no point in quoting that portion of the passage: the argument against the “unfreedom of the will” is a bad argument, depending on a neo-Kantian doctrine Nietzsche himself came to repudiate. By contrast, the argument against the *causa sui* is sound in its own right, independent of the neo-Kantian skepticism, and an argument consistent with other claims Nietzsche makes in this and subsequent works. (For similar reasons, there is no reason to take particularly seriously Nietzsche’s claim that “in real life it is only a matter of *strong* wills and *weak* wills”, since that claim is motivated *entirely* in this passage by the neo-Kantian skepticism regarding causation which Nietzsche abandons.)

As to why Owen and Ridley run together “classical determinism” and Nietzsche’s fatalism (or “causal essentialism” as I call it), the answer is suggested by their claim that “causal essentialism [is] ... the view that, at any given moment, *the* trajectory of a person’s life ... is determined by type-facts plus environment, a view which, given that type-facts plus environment just equals the totality of causally relevant facts, is indistinguishable from classical determinism” (2003: 74). Contrary to Owen and Ridley, causal essentialism (which they characterize correctly) is still distinguishable from classical determinism, since the latter requires that there be “laws of nature”, and it is precisely because Nietzsche is a radical empiricist skeptic about *laws* that he eschews the language of classical determinism. Nietzsche’s “official” view (strange as it may seem) is that ours is a world of *token* necessities, not lawful necessities, and that is true even if persons are of different *types*: the *types* interact with the environment in ways that are not law-governed.

it) — namely, the argument that they come when they want, not when I want — is not, however, the only consideration influencing Nietzsche’s conclusion that the will is causally determined. That is actually fortunate, since the argument from the phenomenology of thoughts doesn’t rule out the possibility that our experience of willing misleads us as to our *real will*, and that this *real will* does, in fact, stand in the appropriate causal relationship with actions. Of course, this “real will”, if it exists, had better be one that the agent can claim as his own if we are then to saddle the agent with responsibility for the actions it produces. But we will return to that issue below.

To see what the other argument for the epiphenomenality of the will might be, we need to remember — as I have argued elsewhere (Leiter 2002: 63–71) — that Nietzsche is very much influenced by the idea popular among German Materialists in the 1850s and later that human beings are fundamentally bodily organisms, creatures whose physiology explains most or all of their conscious life and behavior. Nietzsche adds to this Materialist doctrine the proto-Freudian idea that the unconscious psychic life of the person is also of paramount importance in the causal determination of conscious life and behavior.<sup>12</sup> Thus, Nietzsche, as I have argued, accepts what we may call a “Doctrine of Types” (Leiter 1998), according to which

Each person has a fixed psycho-physical constitution,  
which defines him as a particular *type* of person.

Call the relevant psycho-physical facts here “type-facts”. Type-facts, for Nietzsche, are either *physiological* facts about the person or facts about the person’s unconscious drives or affects. The claim, then, is that each person has certain largely immutable physiological and psychic traits that constitute the “type” of person he or she is. While this is not, of course, Nietzsche’s precise terminology, the ideas are familiar enough from his writings.

A typical Nietzschean form of argument, for example, runs as fol-

12. Nietzsche’s “official” view seems to be that physiology is primary, but he mostly concentrates on psychological claims, most obviously because he is no physiologist!

lows. A person's theoretical beliefs are best explained in terms of his moral beliefs; and his moral beliefs are best explained in terms of natural facts about the type of person he is (*i. e.*, in terms of type-facts). So Nietzsche says, "[E]very great philosophy so far has been ... the personal confession of its author and a kind of involuntary and unconscious memoir"; thus, to really grasp this philosophy, one must ask "at what morality does all this (does *he*) aim" (BGE 6)? But the "morality" that a philosopher embraces simply bears "decisive witness to *who he is*" — *i. e.*, who he *essentially* is — that is, to the "innermost drives of his nature" (BGE 6). Indeed, this explanation of a person's moral beliefs in terms of psycho-physical facts about the person is a recurring theme in Nietzsche. "[M]oralities are ... merely a sign language of the affects" (BGE 187), he says. "Answers to the questions about the *value* of existence ... may always be considered first of all as the symptoms of certain bodies" (GS P:2). "Moral judgments", he says, are "symptoms and sign languages which betray the process of physiological prosperity or failure" (WP 258). "[O]ur moral judgments and evaluations ... are only images and fantasies based on a physiological process unknown to us" (D 119), so that "it is always necessary to draw forth ... the *physiological* phenomenon behind the moral predispositions and prejudices" (D 542). A "morality of sympathy", he claims, is "just another expression of ... physiological overexcitability" (TI IX:37). He attributes *ressentiment* — and the morality that grows out of it — to an "actual physiological cause [*Ursache*]" (GM I:15). Nietzsche sums up the idea well in the preface to the *Genealogy*: "[O]ur thoughts, values, every 'yes,' 'no,' 'if' and 'but' grow from us with the same inevitability as fruits borne on the tree — all related and each with an affinity to each, and evidence of one will, one health, one earth, one sun" (GM P:2).

We can see Nietzsche's Doctrine of Types clearly at work in his theory of the will when we move from *Daybreak* and *Beyond Good and Evil* to another significant discussion of the will in the late Nietzschean corpus: "The Four Great Errors" section of *Twilight of the Idols*. Three of the errors Nietzsche discusses here pertain to causation: he calls them "the error of confusing cause and effect" (1–2), "the error of false

causality" (3), and "the error of imaginary causes" (4–6). The fourth "great" error, by contrast, is "the error of free will" (7–8), though there is, in fact, no argument given in these concluding sections of the chapter for why free will is an error. (Instead Nietzsche offers one debunking explanation of why people might be motivated to believe in free will apart from its reality.) The inference the reader is plainly supposed to draw is that the "error of free will" follows from the errors about causation discussed in the preceding sections. If we can reconstruct the argument that emerges from those sections, then, we shall have identified Nietzsche's other main reason for viewing the will as epiphenomenal and thus for rejecting free will.<sup>13</sup>

The first error, that "of confusing cause and effect", can be summarized as follows. Given two regularly correlated effects E<sub>1</sub> and E<sub>2</sub> and their mutual "deep cause", we confuse cause and effect when we construe E<sub>1</sub> as the cause of E<sub>2</sub>, missing altogether the existence of the deep cause. Let us call this error "Cornarism" after the example Nietzsche uses:

Everybody knows the book of the famous Cornaro in which he recommends his slender diet as a recipe for a long and happy life. ... I do not doubt that scarcely any book (except the Bible, as is meet) has done as much harm. ... The reason: the mistaking of the effect for the cause. The worthy Italian thought his diet was the *cause* of his long life, whereas the precondition for a long life, the extraordinary slowness of his metabolism, the consumption of so little, was the cause of his slender diet. He was not free to eat little *or* much; his frugality was not a matter of "free will": he became sick when he ate more. (TI VI: 1)

In other words, what explains Cornaro's slender diet *and* his long life is the same underlying fact about his metabolism. Cornaro's mistake was

13. We shall return in the final section of the paper to the ambiguity regarding the strength of the epiphenomenalist doctrine as articulated so far.

to prescribe his diet for all without regard for how individuals differed metabolically, metabolism being the relevant type-fact in this context.

Even if we grant Nietzsche all the facts as he presents them, this would not yet show that there is no free will, unless the error involved in Cornarism extended beyond cases such as diet and longevity. But that is exactly Nietzsche's contention, since in the very next section he saddles morality and religion quite generally with Cornarism. According to Nietzsche, the basic "formula on which every religion and morality is founded is: 'Do this and that, refrain from that and that — then you will be happy! Otherwise ...'" Cornaro recommended a slender diet for a long life; morality and religion prescribe and proscribe certain conduct for a happy life. But, says Nietzsche,

[A] well-turned out human being ... *must* perform certain actions and shrinks instinctively from other actions; he carries the order, which he represents physiologically, into his relations with other human beings and things.

So morality and religion are guilty of Cornarism: the conduct they prescribe and proscribe in order to *cause* a "happy life" are, in fact, *effects* of something else, namely the physiological order represented by a particular agent, one who (as Nietzsche says) "*must* perform certain actions", just as Cornaro *must* eat a slender diet (he is "not free to eat little or much"). That one performs certain actions *and* that one has a happy life are themselves both effects of the physiological order. If we grant Nietzsche the Doctrine of Types, then there is indeed reason to think that Cornarism is a feature of morality too, since morality fails to recognize the crucial role of type-facts in determining what one does, even what morality one accepts.

That brings us to the next "error", that of "false causality", the mistake of thinking we know what causation is because of our introspective confidence in what we take to be the causal powers of our own mental life. Nietzsche explains:

We believed ourselves to be causal in the act of willing;  
we thought that here at least we caught causality in the

act. Nor did one doubt that all the antecedents of an act, its causes, were to be sought in consciousness and would be found there once sought — as "motives": else one would not have been free and responsible for it. Finally, who would have denied that a thought is caused? That the "I" causes the thought? (TI VI:3)

We already know, of course, from the phenomenology-of-thought argument, who it is that denies that thoughts are caused by the "ego", by some internal agency, such that one would be "free and responsible" for them. And Nietzsche soon makes clear in this section of *Twilight of the Idols* that his view remains unchanged:

The "inner world" is full of phantoms ... : the will is one of them. The will no longer moves anything, hence does not explain anything either — it merely accompanies events; it can also be absent. The so-called *motive*: another error. Merely a surface phenomenon of consciousness — something alongside the deed that is more likely to cover up the antecedents of the deeds than to represent them. ... What follows from this? There are no mental [*geistigen*] causes at all. (TI VI:3)

In the last line, Nietzsche must mean only that there are no *conscious* mental causes (or, at least, no conscious mental causes of those actions with which morality is concerned).<sup>14</sup> Indeed, in other passages,

14. Katsafanas (2005) is correct, I think, to object — *contra* Leiter (2002) as well as Gilles Deleuze — that Nietzsche cannot hold that consciousness *per se* is epiphenomenal, though I am ultimately skeptical about his own intriguing, often ingenious presentation of Nietzsche's account of consciousness, since it floats free of the texts at crucial points. (It also makes the Freudian idea of unconscious wishes unintelligible, since their content must be conceptually articulated, though Freud was quite correct, in my view, to have found the germ of this idea in Nietzsche.) But there is plainly a need to individuate *those conscious* states whose causal efficacy is impugned by Nietzsche. Katsafanas's own interesting distinction between the epiphenomenality of the Ego (the "substantive faculty" of "Consciousness") versus causally efficacious conscious mental states (those mental states marked by their conceptually articulated content) (2005: 13) might be one way of marking the relevant distinction for

he is explicit that the target of this critique is the picture of conscious motives as *adequate* to account for action.<sup>15</sup> As he writes in *Daybreak*, “[W]e are accustomed to exclude all [the] unconscious [*unbewusst*] processes from the accounting and to reflect on the preparation for an act only to the extent that it is conscious” (D 129), a view which Nietzsche plainly regards as mistaken, both here and in the passage quoted above. Indeed, the theme of the “ridiculous overestimation and misunderstanding of consciousness” (GS 11) is a recurring one in Nietzsche. “[B]y far the greatest part of our spirit’s activity”, says Nietzsche, “remains unconscious and unfelt” (GS 333; cf. GS 354). To be sure, there is a somewhat suspect overreaching in these passages: are we to believe that no conscious belief is part of the causal explanation of *any* action? I do not see that Nietzsche needs to defend this radical thesis, for what he is interested in debunking is the causal nexus between the conscious experience of will and actions of moral significance, that is, the actions for which moral praise and blame might be ascribed.<sup>16</sup>

---

Nietzschean purposes. Freud’s account — in which those conscious mental states which conjoined an Idea and an Affect were the ones that were causally determined — does not, obviously at least, map on to Nietzsche’s usage.

15. Notice that the fact that conscious mental causes are not *adequate* to account for morally significant actions does *not* mean they are irrelevant to the best account. What is at stake in the claim about their lack of adequacy is only the thought that they are *type*-epiphenomenal, that is, their causal relevance is dependent upon the relationship in which they stand to physiological or unconscious causes. (Thanks to an anonymous referee for pressing this point.)
16. Gemes (forthcoming, esp. at note 8) claims that Nietzsche was not mainly interested in the connection between free will and moral responsibility, but bases this on the quite surprising (and undocumented) assertion that Kant was not centrally interested in this issue. Gemes introduces a doctrine he calls “agency free will” (which he says certain later German Idealists — with whom, of course, Nietzsche had little familiarity or interest — were primarily interested in) which, he says, is more interested in “autonomy” than “moral responsibility”. “Autonomy” on Gemes’s view answers the question “what constitutes an action as opposed to a mere doing”. To be sure, any view about free will and moral responsibility will have to distinguish between “doings” and “actions”, but what Gemes owes us is both (1) an account of “actions” which somehow prescind from claims about responsibility, and (2) evidence that this was Nietzsche’s primary concern. As to (2), the best Gemes can do is

Notice, of course, that there is, as yet, no *argument* given for these claims, just conclusory statements about the causal inertness of putative conscious mental causes. But once again, if we accept the Doctrine of Types underlying the first error, then the claims here also make some sense. That is, if morally significant actions and the conscious mental states that precede them are themselves the product of type-facts, then it follows that *the conscious mental states* that precede the action and whose propositional contents would make them appear to be causally connected to the action are, in fact, epiphenomenal, either

---

call attention to the beginning of the Second Essay of the *Genealogy*, where Nietzsche writes about “the sovereign individual ... with his own independent long will. ... This being who has become free, who is really *permitted* to promise, this lord of the *free* will, this sovereign ...” (GM II: 2). Even putting to one side the idiosyncratic nature of this passage — it is hard to see why this one passage should be thought more significant than the extensive textual evidence considered in the body of this paper — its meaning is far less evident than Gemes thinks. For Nietzsche quickly puts ‘free’ in quotes, to signal, of course, that this “sovereign individual” is not free in the sense requisite for moral responsibility (which is nowhere mentioned in the passage). And as I wrote about this section previously (Leiter 2002: 228), the “sovereign individual”:

is distinguished by one overriding trait: he has “*the right to make a promise*” (GM II: 2) because he can actually pull it off, *i. e.*, his behavior is sufficiently regular and predictable so that he can be “answerable for his own future” (GM II: 1), and he is able to remember what he has promised and honor that memory. But we know already from the opening lines of the Second Essay that this just means he is a certain kind of “animal” that has been bred the right way and nothing more. Indeed, should we be misled by the rhetoric of GM II:2, Nietzsche soon reminds us of the real point when he asks in the very next section: “How do you give a memory to the *animal man*?” (GM II: 3, emphasis added).

The answer to that latter question about *breeding* involves a whole series of mnemonic techniques (many involving torture of one kind or another) that would, indeed, explain how an animal could be trained that is able to “keep a promise”. Yet far more is at stake in the question of free will and moral responsibility than breeding animals who can make and discharge their promises, and nothing in the *Genealogy* suggests otherwise. (I should add that I am largely sympathetic to Gemes’s thesis that Nietzsche views the *self* as a hierarchy of drives, but am unclear why he thinks Nietzsche conflates this point with the theory of the will I reconstruct in this essay.)

as *tokens* or as *types*: that is, they are either causally inert with respect to the action or causally effective *only in virtue of* other type-facts about the person.

That brings us to the final error about causation, what Nietzsche calls the error of “imaginary causes” (TI VI:4). This error occurs when we invent, *post hoc*, causes to explain certain phenomena in our experience, phenomena that are, in reality, the cause of our invention. Nietzsche uses the striking example of dreams, though I’ll modify the content of his example. Suppose that while I am dreaming in the early morning hours, a police car, with siren wailing, passes by my window, but does not wake me. More often than not, into my dream will emerge a narrative which *explains* the sound: perhaps in my dream I suddenly find myself being pursued by police, in their cars, with *their* sirens wailing. As Nietzsche puts it, “The representations which were *produced* by a certain state have been misunderstood as its causes” (TI VI:4). That is, the dream police car, and the dream siren — themselves *actually* the product of the real, external sound of a siren — are now, in the dream, treated as the *causes* of that sound.

Once again, what is significant for Nietzsche is that, as he puts it, “the whole realm of morality and religion belongs under this concept of *imaginary causes*” (TI VI:6). Let us take just one of Nietzsche’s examples. Christians, he says, might “explain” “agreeable general feelings” as being produced by “faith, charity, and hope ... the Christian virtues” (TI VI:6). One feels well, at peace, content, because one practices these Christian virtues — or so the religious explanation goes. But, objects Nietzsche,

[A]ll these supposed explanations are resultant states and, as it were, translations of pleasurable or unpleasurable feelings into a false dialect: one is in a state of hope *because* the basic physiological feeling is once again strong and rich ... (TI VI:6)

So the Christian says, “That you have practiced the Christian virtues explains why you feel well and are at peace with yourself.” In fact, says

Nietzsche, there is a physiological explanation for why an agent who feels at peace with himself feels that way, and it is also an explanation for why he practices hope, faith, and charity. The structure of the criticism suggests, in fact, that the “error of imaginary causes” is just an instance of the first error, that “of confusing cause and effect”, since, once again, one mistakes an “effect” (*e.g.*, the feeling of hope) for the cause of something else (*e.g.*, being at peace with oneself), when both are effects of an unrecognized “deep cause”, *i.e.*, “the basic physiological feeling” as Nietzsche has it in this example.<sup>17</sup> And as with the earlier error, this one seems to depend entirely on accepting Nietzsche’s Doctrine of Types, his doctrine that the psycho-physical facts about a person explain their conscious experience and behavior.

How does it follow from the three errors about causation — really, *two* errors, since the last is just an instance of the first — that “free will” is also an error? The error of confusing cause and effect is a *general* error that afflicts morality because morality is based on a mistaken picture of agency: we think that certain moral prescriptions will bring about certain consequences for those who follow them, yet the ability and disposition to act on the prescriptions, and the enjoyment of the consequences, are possible only for certain *types* of persons. An exercise of free will plays no role. Notice that this argument makes no claim about the phenomenology of willing.

The error of false causality is an error because we wrongly infer that we know what causation is from our *experience* of the will being causal; but the will is not, in fact, causal, which follows from the Doctrine of Types. But, on any account of free will and moral responsibility, the will must be causal (even if not *causa sui*), in order for agents to have free will and be morally responsible for their actions. Therefore, if the error of false causality is a genuine error, then it follows that there is no free will. Only this second error implicates the phenomenology of

17. So, too, in the case of dreams: one treats the dream police car as cause of the dream siren, when in fact both are caused by the *real* siren. What makes the case of “imaginary causes” a special instance of the error of confusing cause and effect is that in this case, E2 is itself a reflection of the deep cause. That difference, as far as I can see, does not matter for our purposes in the text.

willing, since it claims that we are in error in thinking we know what causation is based on our *experience* of the will. And the argument says we are in error here because our experience of the will misleads us as to the causal powers of the will: “there are no mental causes at all”, Nietzsche tells us.

Let us return, now, to the question with which we started: Why, according to Nietzsche, is the phenomenology of willing not a reliable guide to the causation of action or, more precisely, the causation of action such that we are morally responsible for the actions so caused? Nietzsche now has given us two arguments for the conclusion: the argument from the phenomenology of thoughts; and the argument from false causality, which depends on his Doctrine of Types.

### The Real Genesis of Action

If the *experience* of willing does not, according to Nietzsche, illuminate how actions are brought about, what, then, *really* explains our actions? The “Four Great Errors” section of *Twilight of the Idols*, and the Doctrine of Types on which it depends, suggest an answer. *Type-facts* — facts about the unconscious psychology and the physiology of agents — explain our actions. That, alas, is to paint with a broom, not a brush, and does not yet shed much light on the actual causal structure of agency or its connection to the phenomenology of willing painted in such detail by Nietzsche in the section from *Beyond Good and Evil* we discussed at the start.

Here is where it may help to turn to some recent work in empirical psychology, powerfully synthesized by Daniel Wegner in his recent book *The Illusion of Conscious Will* (Wegner 2002). Wegner, like Nietzsche, starts from the *experience* of willing, and, like Nietzsche, wants to undermine our confidence that the experience accurately tracks the causal reality. To do so, Wegner calls our attention to cases where the *phenomenology* and the *causation* admittedly come apart. One set of cases involve “illusions of control”, that is, “instances in which people have the feeling they are doing something when they actually are not doing anything” (2002: 9). (Think of a video game in which

you feel your manipulation of the joy stick explains the action on the screen, when in fact, the machine is just running a pre-set program.) Another set of well-documented cases involve the “automatisms”, that is, cases where there is action but no “experience of will” (2002: 8–9). (Examples would include ouija board manipulation and behaviors under hypnotism.) Wegner remarks:

[T]he automatisms and illusions of control ... remind us that action and the feeling of doing are not locked together inevitably. They come apart often enough to make one wonder whether they may be produced by separate systems in the mind. The processes of mind that produce the experience of will may be quite distinct from the processes of mind that produce the action itself. (2002: 11)

If the cases in question do, indeed, show that the phenomenology of willing is not always an accurate guide to causation, they certainly do not show that this is *generally* true. But Wegner wants to establish Nietzsche’s claim, namely, that the phenomenology of willing systematically misleads us as to the causation of our actions. And in the place of the “illusion of free will” as Wegner calls it, he proposes a different model according to which “both conscious willing and action are the effects of a common unconscious cause” (Holton 2004: 219), but the chain of causation does not run between the experience of willing and the action; rather, in Nietzschean terms, some type-fact about persons explains both the experience *and* the action (see the helpful diagram in Wegner & Wheatley 1999: 483; Wegner 2002: 68). As Wegner sums up his alternative picture of the causal genesis of action:

[U]nconscious and inscrutable mechanisms create both conscious thought about action and the action, and also produce the sense of will we experience by perceiving the thought as cause of the action. So, while our thoughts may have deep, important, and unconscious causal connections to our actions, the experience of conscious will

arises from a process that interprets these connections, not from the connections themselves. (Wegner 2002: 98)

Before we turn to one part of Wegner's empirical evidence for this strong claim, it is important to recall an ambiguity that emerged in our earlier discussion of Nietzsche's epiphenomenalism, an ambiguity well illustrated in the causal story suggested by Nietzsche's example of Cornaro. On one reading — call this *the Will as Secondary Cause* — slow metabolism (the relevant type-fact about Cornaro) explains why Cornaro ate a slender diet, and the fact that he ate a slender diet explains his longevity. If we take this version as an analogue of willing, then the will is, indeed, causal, but it is not the *ultimate* cause of an action: something *causes* the experience of willing and then the will causes the action.

On another reading — *the Will as Epiphenomenal* — the slow metabolism explains *both* the slender diet and the longevity, but there is no causal link between the latter two. The Cornaro example itself most plausibly suggests the Will as Secondary Cause (surely the slender diet makes a *causal* contribution to the long life), yet the passage with which we began, from *Daybreak*, suggests the Will as Epiphenomenal instead: if the "I will" is really analogous to the person "who steps out of his room at the moment when the sun steps out of its room, and then says: 'I will that the sun shall rise'" (D 124), then there is no causal link between the experience of willing and the resulting action, just as there is no causal link between the person who wills the sun to rise and the rising of the sun.

Nietzsche's texts on this subject are, I believe, generally ambiguous as to which view of the will he decisively embraces. Wegner's empirical evidence, by contrast, is offered in support of the Will as Epiphenomenal. If Wegner is right, that is good reason, as a matter of interpretive charity, to read Nietzsche as committed to the latter view,<sup>18</sup>

18. In the terms used in Leiter (2002: 91–92), this would be to read Nietzsche as embracing Token Epiphenomenalism, contrary to the interpretation emphasized in my earlier work.

that is, to read him as holding the view that is (a) supported by his texts, and (b) most likely to be correct as a matter of empirical science.<sup>19</sup>

Wegner adduces a variety of kinds of evidence to support the idea of the Will as Epiphenomenal, and it will not be possible to review most of them here. I want to concentrate on one illustrative bit, already well known to students of the free will literature but probably less familiar to those interested in Nietzsche. These are the studies by Benjamin Libet and colleagues (discussed in Wegner 2002: 50–55) examining the brain electrical activity (the "readiness potential" or "RP") that precedes an action (such as moving a finger) and the experience of willing. What the researchers found is that "the conscious willing of finger movement occurred at a significant interval *after* the onset of the RP (and also at a significant interval before the awareness of the movement)" (Wegner 2002: 53). According to Wegner, "These findings suggest that the brain starts doing something first ... [t]hen the person becomes conscious of wanting to do the action" that the brain has already initiated (Wegner 2002: 53).<sup>20</sup> Wegner (2002: 54) quotes Libet summing up the import of his findings as follows:

[T]he initiation of the voluntary act appears to be an unconscious cerebral process. Clearly, free will or free choice of whether *to act now* could not be the initiating agent, contrary to one widely held view. This is of course also contrary to each individual's own introspective feeling that he/she consciously initiates such voluntary acts; this provides an important empirical example of the possibility that the subjective experience of a mental causal-

19. To the extent that the Will as Epiphenomenal is not vindicated by empirical research — the verdict is plainly out — then the argument from interpretive charity may ultimately cut the other way. (Thanks to Eddy Nahmias for pressing this point.)

20. This experimental result has a rather striking resonance with a phenomenological claim made by Nietzsche in *Nachlass* material, namely, that "in cases of sudden pain the reflex comes noticeably earlier than the [conscious] sensation of pain" (WP 699). (Thanks to an anonymous referee for calling this passage to my attention.)

ity need not necessarily reflect the actual causative relationship between mental and brain events.<sup>21</sup>

In other words, about a century after Nietzsche, empirical psychologists have adduced evidence supporting his theory that the phenomenology of willing misleads as to the actual causal genesis of action.

This is not the only place that empirical psychology has “caught up”, as it were, with Nietzsche — or, more charitably, provided confirmation for Nietzsche’s theory of the will. Wegner adduces support, for example, for what Nietzsche calls the “error of false causality” (Wegner 2002: 64 ff.), as well as the “error of confusing cause and effect” (Wegner 2002: 66 ff., 96 ff.). What is worth emphasizing here, however, is the extent to which the alternative picture of the genesis of action suggested by Libet and Wegner — namely, that unconscious processes explain both the action and the experience of willing — resonates with Nietzsche’s own, more detailed remarks on the genesis of action that come in the fascinating passage in Section 109 of *Daybreak* on “self-mastery”.

Nietzsche begins this discussion by canvassing six different ways of “combating the vehemence of a drive [*eine Triebe*]”. What follows is Nietzsche at his most psychologically astute, as he documents six different ways of mastering a powerful urge: for example, by avoiding opportunities for gratification of the drive, thus weakening it over time;

21. Alfred Mele (2006: 30–48) has recently mounted a sustained attack on Libet’s interpretation of his results. Mele says that the onset of the conscious experience of willing after the RP “leaves it open that ... rather than acquiring an intention or making a decision of which he is not conscious [at the time of the RP], the agent instead acquires an *urge* or *desire* of which he is not conscious”, and which only becomes effective in virtue of the conscious intention (2006: 33). Although Mele’s point gains some support from an interpretation of other experiments by Libet (discussed by Mele at 34–40), he establishes only that there is an *alternative* interpretation, not that his alternative is correct. But the resolution of the Libet-Mele dispute does not matter for Nietzsche’s purposes, since Nietzsche is an incompatibilist, and the Libet results, even on Mele’s rendering, show that *the causal trajectory* (whether that is an urge or an intention) leading to the action begins prior to the conscious intention to perform the action (whenever that occurs), and that is sufficient to defeat the *causa sui* conception of freedom of the will.

or by learning to associate painful thoughts with the drive, so that its satisfaction no longer has a positive valence.

Interesting as these observations are, the real significance of this passage for our purposes lies elsewhere. For Nietzsche is also concerned to answer the question as to the “ultimate motive” for “self-mastery”. He explains it as follows:

[T]hat one *wants* to combat the vehemence of a drive at all, however, does not stand within our own power; nor does the choice of any particular method; nor does the success or failure of this method. What is clearly the case is that in this entire procedure our intellect is only the blind instrument of *another drive*, which is a *rival* of the drive whose vehemence is tormenting us. ... While “we” believe we are complaining about the vehemence of a drive, at bottom it is one drive *which is complaining about the other*; that is to say: for us to become aware that we are suffering from the *vehemence* of a drive presupposes the existence of another equally vehement or even more vehement drive, and that a *struggle* is in prospect in which our intellect is going to have to take sides.

Even if the intellect must “take sides” (*Partei nehmen*), this plainly does not mean that the intellect determines which side prevails: to the contrary, the intellect is a mere spectator upon the struggle. Thus, the fact that one masters oneself is *not* a product of “free will” but rather an effect of the underlying type-facts characteristic of that person: namely, which of his various drives happens to be strongest. There is, as it were, no “self” in “self-mastery”: that is, no conscious “self” who contributes anything to the process. “Self-mastery” is merely an effect of the interplay of certain unconscious drives, drives over which the conscious self exercises no control. A “person” is an arena in which the struggle of drives (type-facts) is played out; how they play out determines what he believes, what he values, what he becomes. But *qua* conscious self or “agent”, the person takes no active part in the process. As Nietzsche

puts the same point, later, in *Beyond Good and Evil*: "The will to overcome an affect is, in the end, itself only the will of another, or several other, affects" (117; cf. also GM III:17). The will, in other words, or the experience of willing (in self-mastery), is itself the product of various unconscious drives or affects. Which is, in slightly different terms, exactly the theory of the will that some empirical psychologists have arrived at one hundred years after Nietzsche.<sup>22</sup>

### References

- Holton, Richard. 2004. Review of Wegner (2002), *Mind* 113: 218–221.
- Gemes, Ken. Forthcoming. "Nietzsche on Free Will, Autonomy and the Sovereign Individual", in *Nietzsche on Freedom and Autonomy*, ed. K. Gemes & S. May (Oxford: Oxford University Press).
- Katsafanas, Paul. 2005. "Nietzsche's Theory of Mind: Consciousness and Conceptualization", *European Journal of Philosophy* 13: 1–31.
- Leiter, Brian. 1998. "The Paradox of Fatalism and Self-Creation in Nietzsche", in *Willing and Nothingness: Schopenhauer as Nietzsche's Educator*, ed. C. Janaway (Oxford: Oxford University Press).
- . 2002. *Nietzsche on Morality* (London: Routledge).
- Mele, Alfred. 2006. *Free Will and Luck* (Oxford: Oxford University press).
- Nagel, Thomas. 1986. *The View from Nowhere* (New York: Oxford University Press).
- Owen, David, and Aaron Ridley. 2003. "On Fate", *International Studies in Philosophy* 35/3: 63–78.
- Pereboom, Derk. 2001. *Living Without Free Will* (Cambridge: Cambridge University Press).
- Strawson, Galen. 1994. "The Impossibility of Moral Responsibility", *Philosophical Studies* 75: 5–24.
- Wegner, Daniel M. 2002. *The Illusion of Conscious Will* (Cambridge, Mass.: Harvard University Press).
- Wegner, Daniel M., & Thalia Wheatley. 1999. "Apparent Mental Causation: Sources of the Experience of Will", *American Psychologist* 54: 480–492.

22. This paper has been in circulation so long that there is already a secondary literature replying to it. I have refrained from making too many changes responding to published (or circulating) critiques, but hope to do so at a later date. I am grateful to the graduate students in my fall 2003 seminar on "Nietzsche and Ethics" at the University of Texas at Austin for their help in thinking about these issues; to the audience at my keynote address on this topic at the meeting of the Friedrich Nietzsche Society at the University of Sussex in September 2004; to the discussants at the conference on "Nietzsche and Naturalism" organized by Mathias Risse at the Radcliffe Institute at Harvard University in November 2004; to audiences at Cambridge University, the University of Manchester, Washington University in St. Louis, Georgia State University, and the University of Minnesota, Minneapolis-St. Paul; and to Reid Blackman, Daniel Came, Richard Holton, Paul Katsafanas, Derk Pereboom, and two anonymous referees for detailed comments.