Philosophy and
Phenomenological Research

Philosophy and Phenomenological Research
Vol. LXXXIX No. 1, July 2014
doi: 10.1111/phpr.623

Nietzsche and Kant on the Will: Two Models of Reflective Agency

PAUL KATSAFANAS
Boston University

According to the Kantian theory of action, desires do not determine the actions of self-conscious agents. Self-conscious agents have the capacity to check the motivational impulses associated with their desires, and to decide, freely and rationally, which desires to act upon. Thus, Kant tells us that the will "can indeed be *affected* but not *determined* by impulses . . . . *Freedom* of choice is this independence from being determined by sensible impulses" (*Metaphysics of Morals* 6:213–214).[1]

According to a standard reading of the Nietzschean theory of action, the opposite is the case. Our actions are the products of a chaotic mix of largely non-conscious desires and drives.[2] Our conscious thoughts are causally impotent, buffeted about by forces that we neither discern nor understand. Appearances of self-conscious decision are illusory.[3] Thus, Nietzsche claims that when an agent decides to do something, the agent is analogous to a boat "following the current," which "'wills' to go that way *because it—must"* (GS 360).[4]

It appears that Kant and Nietzsche are asserting diametrically opposed accounts of agency. Indeed, this is how the literature on Kant

[1] For discussions of these ideas, see Allison (1990, Chapters 3 and 5), Baron (1995, 189ff.), Korsgaard (1996, 93ff.), and Wood (1999, 51ff).

[2] "Drive" (*Trieb, Instinkt*) is a term of art for Nietzsche. It refers to a non-conscious disposition toward a characteristic type of activity; this disposition manifests itself, in part, by generating various conscious desires and affects that incline the agent to engage in the activity. For example, the reproductive drive is a non-conscious disposition toward sexual activity that manifests itself in part by generating desires, attractions, emotions, and so forth. For a discussion of Nietzsche's notion of drive, see Katsafanas (forthcoming).

[3] See, for example, Gemes (2009), Leiter (2001), and Risse (2007).

[4] I cite Nietzsche's works using the standard abbreviations of their English-language titles, followed by section number. For a key, see the References.

and Nietzsche typically proceeds. Nietzscheans charge Kantians with relying on an excessively intellectualized, empirically implausible conception of agency, which ignores the pervasive impact of non-conscious phenomena on our actions.[5] Kantians charge Nietzscheans with ignoring the very real role that choices and other conscious phenomena play in the production of action.[6] Accordingly, we seem to have a standoff.

In this essay, I am going to suggest that the debate between Kant and Nietzsche actually takes a far more interesting form. Nietzsche is not merely rejecting the Kantian picture of agency. Rather, Nietzsche is offering a subtle critique of the Kantian theory, denying certain aspects of it while preserving others. The resultant theory of agency is considerably more sophisticated than has yet been appreciated.

The paper falls into five sections. Section 1 considers a potential obstacle to any interpretation of Nietzsche on agency: Nietzsche seems to alternate between denying that there is any such thing as a will (conceived as a causally efficacious capacity for reflective choice) and relying on a conception of the will. I show that this apparent inconsistency is a result of a change in Nietzsche's view: in his early and middle works, Nietzsche accepts an incompatibilist account of willing, according to which the truth of causal determinism rules out the possibility of genuine willing. However, in the late works, Nietzsche modifies his view. He develops a compatibilist conception of willing, according to which our wills, though causally determined, are philosophically significant.

With this groundwork in place, Section 2 introduces the Kantian conception of the will. I distinguish three of Kant's central claims about willing: that reflection suspends the effects of motives, that motives do not determine choice, and that choice determines action.[7] Sections 3 and 4 show that Nietzsche endorses certain aspects of this model while rejecting others. Due to the fact that he develops a more complex account of motivation, Nietzsche concludes that reflection is not capable of suspending the influence of motives. Nonetheless, he

[5] See especially Knobe and Leiter (2007) and Risse (2007).

[6] For an example, see Gardner (2009).

[7] Kant and Nietzsche use different terminology in describing the states that incline us toward action: whereas Kant speaks of inclinations (*Triebfeder*) and desires (*Begehr, Begierde*), Nietzsche more often appeals to affects (*Affekt*), urges (*Drang*), feelings (*Empfindung*), and drives (*Trieb, Instinkt*). A primary topic in this essay is the relationship between the items on this list and reflective choice. Accordingly, it will be helpful to have a term that refers to all of the items on the list. I will use the term *motive* in this way: it should be understood as a genus that has inclinations, desires, urges, affects, feelings, and drives as its species. (This is not to deny that there are important differences between, e.g., affects and desires. However, the differences are not relevant for the arguments in this essay.)

agrees with Kant that motives do not determine choice: our motives could be the same, and yet we could choose differently. Moreover, he maintains that conscious choice plays a causal role in the production of action.

According to my interpretation, then, Nietzsche gives conscious thought a causal role in the production of action. However, some commentators read Nietzsche as denying that conscious thought plays any such role. Section 4 examines this issue, considering whether and in what sense Nietzsche might be an epiphenomenalist about conscious thought. I argue that the epiphenomenalist interpretation cannot be correct, for it fails to account for Nietzsche's claim that conscious thought can transform the motivational propensities of our affects.

Finally, Section 5 argues that in light of these results Nietzsche can be shown to have a philosophically significant conception of the will. I argue that this model preserves certain Kantian insights about the nature of self-conscious agency, while embedding these insights in a more plausible account of motivation.

1. The Development of Nietzsche's Account of Willing

Before we can compare Nietzsche and Kant on the will, we must show that Nietzsche has a consistent account. This task is not straightforward, for Nietzsche seems to alternate between denying and affirming the existence of the will. Below, I argue that this apparent inconsistency results from a change in Nietzsche's views. Section 1.1 argues that in his early and middle works, Nietzsche accepts both incompatibilism and eliminativism about willing. Section 1.2 shows that in his later works, Nietzsche shifts to a compatibilist account of willing, which gives him space to defend a robust conception of willing.

1.1. Nietzsche's Early Acceptance of Incompatibilism and Eliminativism

In works written prior to 1883, Nietzche is decidedly skeptical about the will.[8] *Daybreak* 124 is characteristic:

> 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'. . . . But, all laughter aside, are we ourselves ever acting any differently whenever we employ the expression: 'I *will*'? (D 124)

[8] The most notable works of this period are *Untimely Meditations, Human, All too Human, Daybreak,* and Parts I–IV of the *Gay Science*. Works published after 1882 include *Thus Spoke Zarathustra,* Part V of the *Gay Science, Beyond Good and Evil, The Genealogy of Morality, The Twilight of the Idols,* and *The Antichrist*.

In passages of this kind, Nietzsche suggests that the will has no causal connection to action; it plays as little role in producing our own actions as it does in producing the sun's movements.

What leads Nietzsche to this surprising claim? In the works of this period, Nietzsche devotes considerable attention to the relationship between the will and causal determinism. For example, he writes,

> Perhaps there exists neither will nor purposes, and we have only imagined them. The iron hands of necessity which shake the dice box of chance play their game for an infinite length of time; so that there *have* to be throws which exactly resemble purposiveness and rationality of every degree. *Perhaps* our actions of will and purpose are nothing but such throws . . . . (D 130)

Here, Nietzsche wonders whether our actions are causally determined. He suggests that if our actions are causally determined, subject to the "iron hands of necessity," then the will does not exist.

In this passage, we can see that Nietzsche is moving from *incompatibilism* to *eliminativism* about the will. Incompatibilism is the claim that the will is free only if it is causally undetermined. Eliminativism is the claim that the will does not exist. In early and middle-period works such as HH and D, Nietzsche assumes that incompatibilism is the correct account of willing. In other words, he understands the claim "X has a will" to mean that X has a capacity for reflective choice that is undetermined by prior events. However, as D 130 indicates, Nietzsche argues that our actions are causally determined by a host of factors other than reflective choice. Among these are culture, upbringing, the person's physiology, and facts about the drives and affects that a person harbors (HH I.39). Consequently, we do not have wills in the sense defined above; we do not have causally undetermined capacities for reflective choice. As Nietzsche puts it in D 130, if our acts are determined by "the iron hands of necessity," then "there exists neither will nor purposes." So he moves from the claim that we lack *free* will to the claim that we lack will.[9]

Of course, this argument is questionable in two ways. First, many philosophers are *compatibilists* about willing, claiming that our actions

[9] Although most of the relevant passages in pre-1883 works move directly from incompatibilism to eliminativism, Nietzsche sometimes seems to have a different argument in mind. Certain passages focus on the causal *efficacy* rather than the causal *antecedents* of willing. For example, D 124 (quoted above) suggests that the will is causally inefficacious; it has no impact on action. In light of these sorts of passages, we might interpret Nietzsche as offering the following argument: acts of will are not among the causal antecedents of our actions; therefore, we should be eliminativists about the will. I will return to this point below, in Section 4.

can be both causally determined and free. Second, even if we hold that freedom requires absence of causal determination, we need not be driven to eliminativism: our actions could be *both* causally determined *and* the products of our wills.

Compatibilism is not an unusual position in the history of philosophy: Descartes, Hobbes, Spinoza, Locke, and Leibniz were all compatibilists; in a different fashion, Kant held that the will and causal determination could coexist. In his earlier works, Nietzsche seems to have made the uncharacteristically gross error of neglecting this position. Fortunately, by the time of GM and BGE, Nietzsche's views on willing are far more sophisticated. He no longer believes that claims about causal determination alone would be enough to settle the debate about the existence and freedom of the will.

We can see this by examining the important passage BGE 21, where Nietzsche focuses upon freedom of will. There, Nietzsche explicitly states that the quick inference from "our actions are causally determined" to "our actions are unfree" is illegitimate. The passage begins:

> The *causa sui* is the best self-contradiction that has been conceived so far . . . but the extravagant pride of man has managed to entangle itself profoundly and frightfully with just this nonsense. The desire for "freedom of the will" in the superlative metaphysical sense . . . the desire to bear the entire and ultimate responsibility for one's action oneself, and to absolve God, the world, ancestors, chance, and society involves nothing less than precisely to be this *causa sui* . . . . (BGE 21)

Freedom of will "in the superlative metaphysical sense" is incompatibilist freedom: having a causally isolated will. Nietzsche, following Spinoza, calls the individual with a causally isolated will a *causa sui* (cause of itself). Nietzsche claims that this idea is simply incoherent: "the *causa sui* is the best self-contradiction that has been conceived so far" (BGE 21). Accordingly, nothing in the world can answer to the incompatibilist conception of freedom.

Interestingly, though, Nietzsche no longer moves from the denial of incompatibilist freedom to eliminativism. He writes,

> Suppose someone were thus to see through the boorish simplicity of this celebrated concept of "free will" [as incompatibilist freedom] and put it out of his head altogether, I beg of him to carry his "enlightenment" a step further, and also put out of his head the contrary of this monstrous conception of "free will": I mean "unfree will," which amounts to a misuse of cause and effect . . . . The "unfree will" is mythology; in real life it is only a matter of *strong* and *weak* wills. (BGE 21)

The fact that we are not *causa sui*—that our wills are causally determined—indicates neither that we are free nor that we are unfree.[10] In other words, Nietzsche is attacking the assumption that willing requires absence of causal determination. He is rejecting incompatibilism.

Throughout his works, Nietzsche accepts some version of determinism. Some commentators have thought that Nietzsche's proclamations of determinism rule out the possibility of his having any account of willing. But now we can see that this is a mistake; while the truth of determinism would of course rule out incompatibilist conceptions of agency, Nietzsche does not subscribe to these models. On the contrary, he suggests that there is no coherent conception of willing that would be threatened by the truth of determinism.

Although all cases of willing are determined, this does not preclude there being ways of distinguishing strong and weak wills. Consider an example: even if we think that all actions are determined, we might still wish to distinguish between the alcoholic's drinking and the ordinary individual's drinking, or between the self-deceived individual's voting and the cognizant individual's voting, or between the victim of ideology's acceptance of a value and the clear-headed individual's acceptance of a value. To draw these distinctions, what matters isn't the mere fact that our actions are determined. What matters is *how* they are determined. To elucidate this point, let's look more closely at the relationship between willing and acting.

1.2. Nietzsche's Positive Conception of the Will

In one of his most extensive discussions of freedom, Nietzsche considers the "sovereign," "autonomous" individual (GM II.2). The sovereign individual's defining characteristic is that he possesses "his own independent, protracted will." That is, the sovereign individual is able to commit himself to a course of action and carry through with his commitment. He is "strong enough to maintain [his commitments] even in the face of accidents, even 'in the face of fate'." By contrast, an unfree individual is "short-willed and unreliable," he "breaks his word even at the moment he utters it." For the unfree individual is incapable of holding himself to a course of action in the face of accidents and temptations. Unable to regulate his own behavior, the unfree individual will only fulfill his projects and goals if, through sheer luck, he encounters no temptations.

Although explicit discussions of the sovereign individual are confined to GM II.2, the entirety of GM II and III appeal to the capacities

[10] Nietzsche makes a related point in A 15, listing as "imaginary causes" both "free will" and "unfree will."

exemplified by this individual. GM II discusses the emergence of the capacity to *promise*. While capacity might seem mundane and familiar, promising actually presupposes a certain conception of willing. As John Richardson puts it, the promisor "must include a strong *inhibitive* power, to refrain from acting immediately upon one's drives. The promisor is able to 'insert a pause' in which to consult its commitments. . ." (Richardson 2009, 139). Drawing attention to this point, Nietzsche writes that "between the original 'I will', 'I shall do this', and the actual discharge of the will, its *act,* a world of strange new things, circumstances, and even acts of will may be interposed, without causing this long chain of will to break" (GM II.1). Yet the promisor can maintain his commitments in the face of these temptations: he is, Nietzsche tells us, "strong enough for that" (GM II.2).[11] The promisor has a capacity to will.

Analogously, GM III considers ascetics—agents who counter their immediate desires, inclinations, and aversions, including their strong aversions to pain. These individuals hold themselves to courses of action that run counter to their natural instincts. They display the capacity to maintain their commitments in the face of competing urges.

Thus, throughout the *Genealogy* Nietzsche appeals to agents with the capacity to will, where willing involves consciously holding oneself to a particular course of action. These characterizations are echoed in other passages from Nietzsche's late works. In *Twilight,* Nietzsche identifies willing with the power "not to react at once to a stimulus, but to gain control of all the inhibiting, excluding instincts . . . the essential feature is precisely not to 'will', to be able to suspend decision. All unspirituality, all vulgar commonness, depend on an inability to resist a stimulus: one must react, one follows every impulse" (TI viii.6). In the same work, Nietzsche defines weakness as the "inability *not* to respond to a stimulus" (TI v.2). The weak individual's actions are determined by whatever impulse or stimulus happens to arise; he possesses no capacity to direct his own behavior. By contrast, the strong individual is able to check his impulses and resist environmental stimuli.

In these passages, as well as others,[12] Nietzsche seems to associate willing with the capacity to control one's behavior reflectively: the "strong" individual is able to decide how to act and ensure that her behavior conforms to her decision. It is important to be clear that Nietzsche's talk of resisting stimuli is most naturally construed as referring to a *reflective* capacity, for two reasons. First, Nietzsche's phrasings ("gaining control" over instincts, "suspending decision," having a "pro-

11 Ridley (2009) and Owen (2009) analyze this point at length.

12 See, in particular, D 560, GM II.3, KSA 11:34[96], and WP 928/KSA 13:11[353].

tracted, independent will'') seem designed to elicit images of reflective processes. Second, every animal *unreflectively* resists certain stimuli: a bird that sees a tasty morsel of food will have an immediate urge to eat it, but will ''resist'' that urge when it notices the cat lying in wait; a badger that sees an approaching predator will have an immediate inclination to flee, but will ''resist'' that urge in order to protect its young. These sorts of cases need not be described in terms of strong *wills*, but simply as one desire or emotion (self-preservation, protection of young) being stronger than another (hunger, fear). If having a strong will simply meant having some desires that are stronger than others, then every animal with desires would *eo ipso* have a strong will. This would make nonsense of Nietzsche's claim that only *some* individuals have strong wills. For these reasons, Nietzsche's talk of strong and weak *wills* must refer to a reflective capacity, rather than a mere conflict of desires.

A problem remains, though: while these passages indicate that Nietzsche endorses a model of willing, other passages seem to suggest just the opposite. In a number of passages from the late works, Nietzsche appears to claim that conscious thoughts, decisions, and acts of will play no role in the causation of our actions. For example, he writes

> The error of false causality . . . . We believe that we are the cause of our own will . . . . Nor did we doubt that all the antecedents of our willing, its causes, could be found within our own consciousness or in our personal 'motives' . . . . But today . . . we no longer believe any of this is true. The 'inner world' is full of phantoms and illusions: the will is one of them. The will no longer moves anything, hence does not explain anything—it merely accompanies events; it can even be absent. (TI vi.3)

So, in TI vi.3, Nietzsche claims that there is no such thing as a will; a few pages later, in TI viii.6 (quoted above), he speaks of strong wills controlling impulses and stimuli. There seems to be a glaring inconsistency.

However, there is a way of defusing the tension. We can take TI vi.3 to be rejecting *one conception* of the will, and TI viii.6 to be endorsing *an alternative* conception of the will. Notice that TI vi.3 speaks of a will that is, in the terminology of BGE 21, ''causa sui'': a will that is determined by nothing other than the agent herself, a will whose ''causes could be found within our own consciousness.'' TI viii.6, on the other hand, speaks of a ''strong'' but not causally isolated will. Accordingly, I suggest that we read these passages as referring to different conceptions of the will.

To be sure, Nietzsche isn't explicit about the fact that TI vi.3 refers to one conception of the will, whereas TI viii.6 refers to a quite different conception. However, the possibilities are as follows:

(1) Nietzsche succumbs to a gross inconsistency: he fails to notice that, in the course of several pages, he both denies and affirms the existence of the will

(2) Nietzsche is sloppy with his terminology: he uses the same term to refer to different things in different contexts.

If we must choose between accusing Nietzsche of gross inconsistency or of terminological sloppiness, the latter seems far preferable. So I will proceed on the assumption that Nietzsche's denial of the will, in passages such as TI vi.3 and A 14, is a denial of the *causally isolated* will, rather than a denial of *all conceptions of* the will.[13]

If this is correct, then Nietzsche endorses *some* conception of willing. But what might it be? To answer that question, I will turn to a model of willing that loomed large in Nietzsche's mind: the model articulated by Kant.

2. Kantian Models of the Will

Kantian accounts of willing focus not on the general question of whether the will is causally determined, but on whether the will is causally determined *by the agent's motives*.[14,15] Kantian accounts of agency claim that when a self-conscious agent reflects on potential actions, she is committed to viewing her deliberation as capable of *suspending the effects of her motivational states*. Consider again the passage quoted in the first paragraph of this essay: the will "can indeed be *affected* but not *determined* by impulses . . . . *Freedom* of choice is this independence from being determined by sensible impulses" (*Metaphysics of Morals* 6:213–214). Elsewhere, Kant writes that the will is "a faculty of

[13] Other passages from the late works make it explicit that Nietzsche is rejecting only *some* conceptions of the will. For example, Nietzsche writes, "today we have taken [man's] will away altogether, in the sense that we no longer admit the will as a faculty [*Vermögen*]" (A 14). Notice that Nietzsche says he is rejecting the will *as a faculty*. While Nietzsche does not explain what he means by "faculty," it seems natural to assume that conceiving of the will as a faculty involves conceiving of it as causally isolated from the agent's drives and motives.

[14] Note that both of the following could be true: (1) the will is causally determined, and (2) the will is *not* causally determined *by the agent's motives*.

[15] I lack the space to defend an interpretation of Kant's complex account of agency. Accordingly, I distinguish between Kant's view and Kant*ian* views. The view that I discuss below is endorsed by many contemporary Kantians (see note 17), and seems to align with Nietzsche's interpretation of Kant on agency. Kant's actual view is, no doubt, more complex. Note also that I am using the term *motive* in a way that differs from Kant's use of *Bewegungsgrund* (often translated as "motive"); see note 7, above.

determining oneself from oneself, independently of necessitation by sensible impulses'' (*Critique of Pure Reason* A 534/B 562), and that ''an incentive [or desire] can determine the will to its action *only insofar as the individual has taken it up into his maxim''* (*Religion within the Boundaries of Mere Reason* 6:24).[16] Thus, reflective agents are capable of suspending the effects of their motivational states and choosing in a way that is not determined by these states.[17]

To see why Kant describes action in this way, consider a paradigmatic case of willing. We can distinguish three steps. First, the agent reflects on some set of data. Kant mentions reflection upon one's motives. We can also include reflection upon other factors, such as plans, projects, goals, commitments, and values, as well as facts about the world. Second, the agent makes a decision about how to act. Third, after reflecting and deciding how to act, the agent attempts to carry out her decision. If all goes well, she acts as she has decided to act. She then manifests a form of successful willing. In sum, exercising the will involves reflecting on data, deciding how to act, and acting in that way.

With this in mind, let's distinguish three components of this Kantian model of willing. First, there is a claim about the causal relationship between reflection and motivation:

---

16 Kant distinguishes two senses of the concept *will*. He uses *Wille* to refer to practical reason, which he treats as the source of the normative content governing our actions. He uses *Willkür* to refer to the capacity for choice, which acts under the governance of *Wille*. In this essay, I am concerned with *Willkür* rather than *Wille*. For a helpful discussion of these distinctions, see Allison (1990, 130–2).

17 A number of commentators discuss this aspect of Kant's theory. For example, Allen Wood writes, ''Kant holds that in the brutes, impulses operate mechanically to produce behavior predetermined by instinct . . . . This means that a brute cannot resist impulses, or decide whether to satisfy a desire, or even deliberate about how to satisfy it'' (1999, 51). On the other hand, ''Kant contrasts this with the human power of choice, which is 'sensitive' (affected by sensuous impulses) but also 'free'. . . . Only a free power of choice is a will . . . . Not only do rational beings have the capacity to resist impulses, but even when the rational faculty of desire acts on sensuous impulses, it is never determined by them mechanically . . . .'' (1999, 51). Henry Allison notes that ''incentives (*Triebfedern*) do not motivate by themselves by causing action but rather by being taken as reasons and incorporated into maxims'' (1990, 51). This ''requires us to regard empirical causes (motives) of the actions of sensuously affected and thoroughly temporal rational agents such as ourselves as 'not so determining' so as to exclude a causality of the will . . . .'' (1990, 52). For ''I cannot conceive of myself as [a rational agent] without assuming that I have a certain control over my inclinations, that I am capable of deciding which of them are to be acted upon (and how) and which resisted'' (1990, 41). Marcia Baron writes that ''Kant's theory of agency is very different [than the familiar causal models]. Our actions are not the result of a desire or some other incentive that impels us. An incentive can move us to act only if we let it'' (1995, 189). Korsgaard (1996, 94) and Reath (2006, 154) agree. On the other hand, Frierson (2005) and McCarty (2009, 67ff.) develop a very different reading of Kant's theory of agency, according to which there is a sense in which motives determine choice.

(Suspension) When an agent reflects on her motives for A-ing, she suspends the influence of the motives upon which she is reflecting.

Suspension is what occurs during the first stage of willing: the agent reflects on her motives (and other factors), but does not take her choice to be determined by these motives (and other factors).[18] Rather, she takes herself to have the capacity to still these motives and choose in independence of them.

Given Suspension, we can make a claim about the causal efficacy of motives. Suppose an agent engages in a bout of deliberation: she reflects on her motives and tries to decide what to do. The reflection will suspend the motivational effects of those motives; consequently, the motives will not necessitate any action. We can put the point this way:

(Inclination) In deliberative agency, motives incline without necessitating. The agent's motives could be the same, and yet she could choose differently.

This is what is at issue in the second stage of willing: the agent takes her decision about how to act to be independent of determination by her motives.

Finally, there is a related claim about the causal efficacy of choice:

(Choice) Typically, if I am faced with two actions that it is possible for me to perform, A-ing and B-ing, and I choose to A, then I will A.

This is manifest in the third stage of willing: the agent takes her choice to determine what she will do.

These three claims compose the Kantian model of willing. Let's see how Nietzsche reacts to this model of willing.

3. Can Reflection Suspend the Influence of Motives?

The passages on strong, sovereign individuals quoted in Section 1.2 suggest that Nietzsche agrees with Kant that motives need not act as brute forces compelling agents to act in particular ways. At least in some cases—cases in which the agent has a "strong" will—the agent is able to counteract the tendencies of her motives and determine her action via choice. Thus, Nietzsche seems to accept Inclination and Choice.

[18] I borrow the term "suspension" from Locke, who writes that the mind has "a power to *suspend* the execution and satisfaction of any of its desires." The mind can "consider the objects of [these desires]; examine them on all sides and weigh them with others. In this lies the liberty that man has" (Locke 1975, 263).

Interestingly, though, Nietzsche makes it clear that he rejects Suspension. Section 3.1 will review Nietzsche's grounds for rejecting Suspension. Given his rejection of Suspension, we are faced with a question: is it possible to develop a conception of agency that denies Suspension but maintains some version of Inclination and Choice? Section 3.2 discusses this possibility.

3.1. Nietzsche's Rejection of Suspension

Nietzsche claims that the agent's reflection is "secretly guided and channeled" by his drives and affects (BGE 3). In addition, he claims that whenever an agent steps back from and reflects upon a drive, the agent's "intellect is only the blind instrument of *another drive*" (D 109). Thus, "the will to overcome an affect is ultimately only the will of another, or several other, affects" (BGE 117). Our reflective thoughts, and indeed even our perceptions, are structured by drives and affects. For this reason, Nietzsche derides the quest for "immaculate perception," perception that is not influenced by any drives (Z II.15). He writes, "there is no doubt that all sense perceptions are entirely suffused with value-judgments" (KSA 12:2[95]/WLN 78).[19] If this is right, then every episode of reflective thought will involve the manifestation of some drive.

Elsewhere, I have argued that these claims should be interpreted as follows: motives manifest themselves by coloring our view of the world, by generating perceptual saliences, by influencing our emotions and other attitudes, and by fostering attractions and aversions.[20] Thus, Nietzsche's idea is that the way in which one experiences the world is, in general, determined by one's motives in a way that one typically does not grasp.

It is easiest to illustrate this point with an example.[21] Suppose that an agent reflects on his jealousy. Part of what it is to be in the grip of jealousy is to see reasons for jealousy everywhere: in the fact that Sarah arrived home a bit later than usual; in the fact that she got off the phone rather quickly last night; in the fact that she is a bit withdrawn lately. (Or, to use a literary example: in the fact that Desdemona is missing a handkerchief.) Accordingly, jealousy and other attitudes can move an agent not simply by overpowering his capacity to resist their pull, but by influencing his judgment and perception. A jealous agent's

[19] For further remarks on these phenomena, see D119, 432, 539; GS 301; GM II.12; BGE 230; CW Epilogue; KSA 11:26[119]/WP 259, KSA 12:7[60]/WLN 139, KSA 13:14[184], KSA 12:14[186], KSA 12:2[148]/WLN 90, KSA 12:10[167]/WLN 201-3.

[20] See Katsafanas (forthcoming).

[21] I discuss this example in Katsafanas (2011).

attention will be drawn to certain features of his environment that another agent would scarcely notice. A jealous agent's trains of thought will return to details that another agent might regard as inconsequential. A jealous agent's deliberative process itself can be influenced by these attitudes; they can incline him to draw conclusions that are not supported by the evidence, to give excessive weight to certain features, and so on. All of this may occur without the jealous agent's recognizing that it is occurring.

Precisely because attitudes influence reflective thought, agents often fail to grasp the ways in which they are being moved by their attitudes. An agent who is moved by jealousy is rarely an agent who consents to be moved by his jealousy; indeed, an agent moved by jealousy need not even *recognize*, much less consent to, a fully formed attitude of jealousy. More often, the jealous agent will struggle to resist the jealousy, but succumb to it in subtler ways. The attitude influences the agent's reflective thought itself: the agent experiences herself as having a reflective distance from the attitude, as scrutinizing the attitude and asking herself whether there is a reason to act on it; but, all the while, the attitude influences the agent's reflective thought in ways that she does not grasp. The jealous agent sees the phone call as furtive, the lateness as suspicious, the handkerchief as damning; and these perceptions, were they accurate, would indeed justify the jealousy. Reflective assessment of the jealousy vindicates it precisely because the agent is being surreptitiously influenced by the very emotion on which she is reflecting.

This type of influence is easiest to detect when we consider an action retrospectively. A person can be dissatisfied with his past actions not because he submitted to or was overcome by a recalcitrant attitude, but because his attitude blinded him, leading him to have a restricted or distorted conception of the options that were open to him. Looking back on my jealous spat with Sarah, the problem was not that I deliberately yielded to jealousy: the problem was that, in the grip of jealousy, I took harmless factors to vindicate my jealous behavior. The problem was that I saw my rage as *warranted* by the fact that she arrived home a few minutes late. I now see that the rage was entirely unwarranted, that I was driven to rage in a way that I did not comprehend. In this way, an agent can act reflectively, yet still be moved by attitudes that operate in the background. (Again, a literary example may be helpful: Othello's problem is not a lack of reflection and deliberation on the grounds for his jealousy; his problem is the way in which this very reflection and deliberation is distorted by his jealousy.)

When in the grip of jealousy, reflective assessment of one's jealous motives will typically vindicate these motives, precisely because the jealousy will manifest itself by inclining the agent to see jealous responses

as warranted by the situation at hand. With this in mind, return to the Suspension claim:

> (Suspension) When an agent reflects on her motives for A-ing, she suspends the influence of these motives.

We can now see that this is false. When I reflect on my motives, it may *appear* that I am suspending their influence. After all, I am not simply impelled to perform the action that they suggest; the jealous agent is not simply impelled to act on his jealousy. Nonetheless, the motives continue to operate on the agent: the agent can scrutinize his motives, decide that there is a reason to act in a certain way, and yet, all the while, be under the thrall of some motive. The effects of the motive needn't be construed as brute compulsions that force an agent to act; rather, the motive moves the agent by influencing the agent's perception of reasons.

Thus, Nietzsche rejects one component of the Kantian account: the Suspension claim. Although reflection may appear to suspend the effects of motives, it typically fails to do so; the influence of the motives simply becomes more covert, operating *through* reflection itself.[22] This marks a crucial difference between Nietzsche and Kant.[23]

3.2 The Relationship between Suspension, Inclination, and Choice

Above, I distinguished three aspects of the Kantian model of willing: Suspension, Inclination, and Choice. I have shown that Nietzsche rejects Suspension. This seems to give us a way of specifying the model of willing that Nietzsche accepts. When Nietzsche rejects the will, he is rejecting either the incompatibilist will or the Suspension claim; when he endorses a conception of the will, he is accepting Inclination and Choice.

This reading would be tidy. Unfortunately, there are still two potential problems. First, Kant supports Inclination and Choice by appeal to Suspension. If Nietzsche rejects Suspension, what grounds might there be for maintaining that Inclination and Choice are true? In

---

[22] This is why Nietzsche writes that positing "the intention as the whole origin and prehistory of an action" is an error. For "everything about [the action] that is intentional, everything about it that can be seen, known, 'conscious', still belongs to its surface and skin—which, like every skin, betrays something but conceals even more. In short, we believe that the intention is merely a sign and symptom that still requires interpretation . . . ." (BGE 32) He goes on, in BGE 33, to claim that behind apparent motives lie deeper motives.

[23] Elsewhere, I have argued that empirical psychology indicates that Suspension is false. See Katsafanas (2011). Thus, the empirical psychology lends support to Nietzsche's model of willing.

particular, how could he claim both that reflection is pervasively influenced by—indeed, driven by—motives and that choice is not determined by motives?

Second, the Kantian model of the will gives a central role to conscious thought: both Inclination and Choice presuppose that conscious deliberation is causally efficacious. Choice says that willing to A plays a causal role in determining whether I A; Inclination says that motives alone do not determine the course of deliberative agency. But this model of willing might seem altogether too *reflective* for Nietzsche. After all, Nietzsche displays considerable skepticism about the importance of conscious thought in our actions. He complains of the "ridiculous overestimation and misunderstanding of consciousness" (GS 11), and writes that "by far the greatest part of our spirit's activity remains unconscious and unfelt" (GS 333).

The next section addresses these questions, articulating a way in which we can delimit the causal role of conscious thought while still treating it as playing a signal role in the production of action. In particular, I will show that Nietzsche treats conscious thought's primary role as the *redirection* of affects: consciousness' effects are thus gradual and incremental. Yet, I argue, Inclination and Choice are still true.

4. Nietzsche on the Role of Conscious Thought

Some commentators have argued that Nietzsche treats conscious thought in general, and conscious willing in particular, as causally inert.[24] If this is correct, Nietzsche could not accept any version of Inclination or Choice. Thus, in this section I consider Nietzsche's claims about the causal efficacy of conscious thought. Section 4.1 reconstructs Brian Leiter's argument that Nietzsche is an epiphenomenalist. The following sections attempt to rebut this view by showing that conscious thought has a causal role. In particular, Nietzsche makes it clear that conscious reflection alters our affects. In Section 4.2, I will produce textual evidence for the claim that we can alter the motivational tendencies of our affects. Section 4.3 examines how *conscious interpretations* of affects, in particular, bring about these shifts. Section 4.4 considers what these claims imply about the causal role of conscious thought.

4.1. Leiter's Reading of Nietzsche as an Epiphenomenalist

What evidence is there for the claim that Nietzsche views conscious thought as causally inert? The most sophisticated defenses of this

[24] See especially Leiter (2001), Leiter (2007), and Gemes (2009).

interpretation are due to Brian Leiter, who has authored a series of articles exploring Nietzsche's claims about conscious thought. Leiter argues that it is a mistake "to conceive of ourselves as exercising our will" (2007, 2). Our experience of willing does not "track an actual causal relationship," but instead "systematically misleads us as to the causation of our actions" (2007, 2). Let's review Leiter's argument for this conclusion.

In his first article on the topic, Leiter argued that Nietzsche views all conscious mental states as epiphenomenal (2001, 294). According to this article, "conscious states are only causally efficacious in virtue of type-facts about the person," where 'type-facts' are "either physiological facts about the person, or facts about the person's unconscious drives and affects" (2001, 294). Put simply, whenever an action seems to be caused by a conscious state, it was actually caused by some non-conscious state (such as a drive or a physiological state).

Katsafanas (2005) argued that this could not be the correct characterization of Nietzsche. There, I surveyed a number of passages in which Nietzsche clearly attributes a causal role to conscious thought. In response, Leiter conceded a version of my point:

> 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. (Leiter 2007, 5)

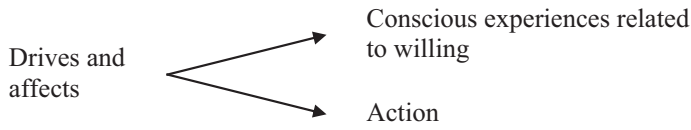
Thus, Leiter accepts my claim that Nietzsche cannot view *all* conscious thought as epiphenomenal. However, Leiter maintains that Nietzsche holds a restricted version of the epiphenomenality thesis: while some conscious states are causally efficacious, the conscious states (or, as Leiter puts it above, experiences) *related to willing* are epiphenomenal. He writes, "*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" (Leiter 2007, 10–11). Conscious states whose propositional content makes them appear causally connected to the action would, presumably, be intentions, deliberation, choices, and so forth. Thus, I will interpret Leiter as claiming that these kinds of conscious states and events are epiphenomenal.[25]

[25] Gemes endorses an analogous claim, writing that "Nietzsche is on the whole fairly dismissive of the import of consciousness," and he goes on to claim that "Nietzsche often claims that conscious willing is largely epiphenomenal, and sometimes seriously flirts with the conclusion that all conscious phenomena are totally epiphenomenal" (Gemes 2009, 48 and 48n19).

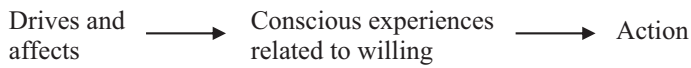
In addition, Leiter argues that Nietzsche's texts are actually ambiguous between two different models of willing. In Leiter's terminology, we can read Nietzsche as claiming either that the will is epiphenomenal or that the will is a "secondary cause" (2007, 13). According to the Will as Epiphenomenal model, the conscious mental states and events associated with willing play no part in the causal chain leading to action. According to the Will as Secondary Cause model, the conscious mental states and events associated with willing are part of the causal chain leading to action, but they are not the primary cause. In other words, they are efficacious only in virtue of other causes.

It helps to illustrate these views with a diagram. If arrows represent directions of causal determination, we have:

Epiphenomenal model:



Secondary cause model:



According to the epiphenomenal model, drives and affects cause both action and conscious experiences of willing. The conscious experiences of willing, however, are not themselves causally connected to the action. According to the secondary cause model, the drives and affects cause conscious experiences of willing, which in turn cause actions.

Thus, Leiter argues that Nietzsche embraces either the epiphenomenal or the secondary cause model of willing. In the following section, I argue that this is a mistake: Nietzsche actually rejects *both* of these views. His position is considerably more complex than either of these views allows. To show this, I will focus on Nietzsche's analysis of the relationship between conscious thoughts and affects. Nietzsche argues that conscious thought can transform the motivational tendencies of affects. This gives conscious willing a rather different role in the production of action than is suggested by either of Leiter's models.

4.2. Nietzsche on the Causal Efficacy of Conscious Interpretations

Nietzsche argues that the primary way in which conscious thought influences action is by influencing our *motives*. To see this, we need to

investigate one of Nietzsche's most counterintuitive claims: that pleasure and pain do not have a determinate motivational impact on human actions.

Nietzsche writes, "what really arouses indignation against suffering is not suffering as such but the meaninglessness of suffering" (GM II.7). What moves us, Nietzsche argues, is not sensation as such, but sensation *coupled with a thought about its meaning*. Agents do not object to the sensation of suffering as such, but rather to suffering that is perceived as *meaningless*. The sensation alone isn't aversive; the sensation coupled with an interpretation is aversive.

Pedestrian examples can illustrate Nietzsche's point. A number of agents seek out the suffering induced by vigorous exercise, competitive sports, and the like, precisely because they regard this suffering as justified (for its health benefits, or for the enjoyment of participating in sports). The selfsame sensations, if induced by illness, a drug, and so forth, would be aversive. Some of these justifications take an instrumental form: we seek pain in order to achieve future pleasure. But others don't. Some agents enjoy the pain induced by running, sport, and so forth for its own sake; that is, they interpret the sensation itself as attractive.

Although Nietzsche takes suffering as paradigmatic, his argument applies to sensations quite generally: the particular way in which a sensation moves us is dependent upon the interpretation that accompanies the sensation. Nietzsche's core argument for this point is present in a crucial passage at the end of the *Genealogy*. He writes,

> Precisely *this* is what the ascetic ideal means: that something *was lacking,* that an enormous *void* surrounded man—he did not know how to justify, to explain, to affirm himself; he *suffered* from the problem of his meaning. He suffered otherwise as well, he was for the most part a *diseased* animal: but suffering itself was *not* his problem, rather that the answer was missing to the scream of his question: "*to what end* suffering?" Man, the bravest animal and the one most accustomed to suffering, does *not* negate suffering in itself: he *wants* it, he even seeks it out, provided one shows him a *meaning* for it, a *to-this-end* of suffering. The meaninglessness of suffering, not the suffering itself, was the curse that thus far lay stretched out over humanity—*and the ascetic ideal offered it a meaning!* Thus far it has been the only meaning; any meaning is better than no meaning at all [. . .]. The interpretation—there is no doubt—brought new suffering with it, deeper, more inward, more poisonous, gnawing more at life: it brought all suffering under the perspective of *guilt*. . . . But in spite of all this—man was *rescued* by it, he had a *meaning* [. . .]. now he could *will something*—no matter for the moment in what direction, to what end, with what he willed: *the will itself was saved.* (GM III.28)

Nietzsche makes a number of claims in this important passage: (i) that we have a desire to regard events in our lives as meaningful or justified; (ii) that particular sensations and emotions, such as occasions of suffering, do not move us except insofar as they relate to the aforementioned desire; (iii) that we will seize upon interpretations that increase our suffering so long as they provide us with a perception of justification

Claim (ii) is crucial for our purposes. It attacks the idea that uninterpreted states have determinate motivational impacts. According to this claim, sensation and emotion acquire motivational directions only in light of interpretations.[26] If I interpret suffering in one way, it will be aversive; if I interpret it in another, it will be attractive. As Nietzsche puts it elsewhere, "that a violent stimulus is experienced as pleasure and pain is a matter of the *interpreting* intellect, which, to be sure, generally works without our being conscious of it; and one and the same stimulus *can* be interpreted as pleasure or pain" (GS 127).

This is true not just of suffering, but of affect quite generally. Consider an important passage from *Daybreak:*

> *Drives transformed by moral judgments*.—The same drive evolves into the painful feeling of *cowardice* under the impress of the reproach custom has imposed upon this drive: or into the pleasant feeling of *humility* if it happens that a custom such as the Christian has taken it to its heart and labeled it *good.* That is to say, it is attended by either a good or a bad conscience! In itself it has, *like every drive*, neither this moral character nor any moral character at all, not even a determinate accompanying sensation of pleasure or displeasure: it acquires all this as a second nature only when it enters into relations with drives already baptized good or evil, or is noted as a property of beings that have already been morally ascertained and assessed by the people.—Thus the older Greeks felt differently about *envy* from the way we do; Hesiod counted it among the effects of the *good,* beneficent Eris, and there was nothing offensive in attributing the gods something of envy: which is comprehensible under a condition of things the soul of which was contest; contest, however, was evaluated and determined as good. (D 38)

Nietzsche claims that envy and the desire to avoid distinguishing oneself (which the Greeks called "cowardice" and we call "humility") acquire different motivational propensities depending upon the way in

[26] Of course, Nietzsche's claim is not that sensations actually tend to occur independently of interpretations. When a sensation or emotion is instantiated in a person, it tends to occur together with an interpretation. But these interpretations can be altered.

which we interpret them. (The passage goes on to give additional examples.[27])

Nietzsche draws attention to a generalized version of this point in *The Gay Science*, writing that what has "caused me the greatest trouble" is

> to realize that *what things are called* is unspeakably more important than what they are. The reputation, name, and appearance, the worth, the usual weight and measure of a thing—originally almost always something mistaken and arbitrary, thrown over things like a dress . . . has, through the belief in it and its growth from generation to generation, slowly grown onto and into the thing and has become its very body: what started as appearance in the end nearly always becomes essence and functions [*wirkt*] as essence! [. . .] Let us not forget that in the long run it is enough to create new names and valuations and presumptions in order to create new 'things'. (GS 58; cf. GS 44)

Interpretations gradually transform the thing that is interpreted.

Notice that we can interpret Nietzsche's claim in two ways:

- Motives as causally inert: motives have no causal tendencies until coupled with interpretations.

- Motives as causally indeterminate: motives have causal tendencies, but the particular behaviors that they characteristically cause are dependent on the associated interpretation.

The first thesis seems implausible: if an animal stumbles into a fire, the pain sensations are going to cause it to withdraw and flee independently of any associated interpretations. However, the second thesis is far more plausible: while uninterpreted sensations of pain are aversive, appropriate interpretations can render these sensations attractive.[28] Although Nietzsche's remarks seem neutral between these two formulations, I think it is best to interpret him as endorsing the latter claim.

4.3. Interpreting and Redirecting our Affects

So far, I have argued that Nietzsche treats motives as causally indeterminate: the particular behavior that a given motive characteristically causes is dependent on the associated interpretation.

---

27 This passage is from a pre-1883 work, in which Nietzsche still seems to endorse incompatibilism and eliminativism. However, we can see that even in these early works, Nietzsche gives interpretation and judgment a role in altering drives and affects.

28 Relevant here are Nietzsche's repeated claims about sublimating or altering the objects of drives. See, for example, BGE 189, BGE 229, and TI vi.3.

This claim gives an important causal role to conscious thought. The remarks on suffering indicate that the motivational tendencies of psychological states are dependent upon our *interpretations* of the states. Interpretations enjoy a causal role in determining the motivational tendencies of even our most basic sensations, such as pleasure and pain. Presumably, these interpretations of our affects can be conscious phenomena.[29] Nietzsche makes this explicit in the passages from the *Genealogy:* the ascetic priests offer religious views that constitute interpretations of suffering. But, if these interpretations are *altering* the motivational propensities of the affects, then it straightforwardly follows that conscious thought is causally efficacious: the interpreting of our affects plays a causal role in the production of action.

To clarify the point, consider a simplistic example: an agent experiences suffering, and is inclined to alleviate it. However, the agent then reflects on the alleged fact that suffering is a punishment from God. This interpretation leads the agent to experience the suffering as partially attractive. Hence, he seeks to perpetuate the suffering. In this fashion, the agent's conscious reflections on his own sensations have a causal impact on his actions.

There are more familiar, everyday examples of this phenomenon. A religious individual interprets sexual activity as sinful, and hence experiences sexual urges in a complex way: the urges will be accompanied by a sense of shame and guilt, and consequently will be partially aversive. At some point, he abandons his religion, coming to see it as an illusion. Accordingly, he no longer takes sexual activity to be sinful. Over time, he comes to experiences sexual urges as alluring rather than (partially) aversive. The motivational propensity of the affect depends on the associated conscious interpretation.[30]

4.4. The Causal Role of Conscious Thought

The above remarks should make it clear that Nietzsche cannot accept Leiter's epiphenomenal model of choice, according to which conscious

---

[29] See, for example, GS 127, quoted above. There, Nietzsche says that these interpretations "generally" [*zumeist*] occur without our being conscious of them, which implies that they *sometimes* occur consciously.

[30] Of course, Nietzsche cannot mean that affects are *completely* malleable. The ascetic or masochist will interpret pain in such a way that he finds it partially alluring, rather than fully aversive; nevertheless, the pain will continue to be partially aversive. After all, part of the point of asceticism and masochism is that one overcomes one's own resistance to aversive sensations. Consequently, the reinterpretation of pain cannot eliminate the aversive qualities, which are the very source of the resistance. The reinterpretation must, instead, couple the aversive qualities with attractive ones.

thought plays no role whatsoever in the production of action. Conscious thoughts have a causal impact on our motives, and hence on our actions.

What about Leiter's secondary cause model? According to this model, conscious thoughts do have a causal impact on action, but the causal impact is unidirectional: motives determine conscious thoughts, and these conscious thoughts then determine actions. According to the evidence adduced above, Nietzsche must reject this view as well. Our conscious thoughts and deliberations are capable of altering our motives, so the series of causes leading from motive to action is more complex than Leiter's model allows. To illustrate this, imagine two agents with identical motives. Suppose these agents experience pity upon witnessing another agent in distress. One agent might reflect, deliberate, develop a certain interpretation of his motives, experience the pity as attractive, and help the agent in distress. The other agent might reflect, deliberate, develop a different interpretation of his motives, experience the pity as aversive, and ignore the agent in distress.

We can picture Nietzsche's view as follows. Motives causally impact the conscious experiences related to willing, which in turn causally influence the motives; out of this process, we get a potentially reconfigured set of motives, with new motivational propensities. This new set of motives might again causally influence the conscious experiences related to willing; and so on. Action results from all of this. Rather than a unidirectional causal path from motives to willing to action, then, we have a play of interacting forces that modify one another and eventually result in action.

If this is right, though, what are we to make of Nietzsche's invectives against our "ridiculous overestimation" of consciousness' role in the production of action (GS 11)? The next section addresses this question.

5. Nietzsche's Model of Willing

There is no denying that Nietzsche critiques our ordinary understanding of reflection's role in the production of action. But doesn't my interpretation have him accepting much of this ordinary understanding? In this section, I argue that Nietzsche is best interpreted as making two points about the role of reflection in action. First, Nietzsche argues that whereas we ordinarily conceive reflective thought as operating in an instantaneous fashion, its effects are actually gradual and incremental. Second, Nietzsche claims that whereas we ordinarily take reflective thought to be decisive in the production of action, it is merely one causal factor amongst many others. So reflective thought's role is far more modest than we have believed. Sections 5.1 and 5.2 examine these points in turn.

5.1. The Incremental Nature of Consciousness's Effects

Part of the explanation for Nietzsche's invectives against conscious thought is that we misunderstand how conscious thought operates. We imagine atomic, momentary acts of choice altering our actions. On Nietzsche's view, though, consciousness' effects are gradual and aggregative.

In the examples given above, conscious reflection on a motive leads to a new interpretation of the motive, and hence to a new motivational propensity. But not all shifts of motives occur in this straightforward, immediate fashion. Indeed, Nietzsche emphasizes that most shifts in motives are gradual and incremental. *Daybreak* 38, quoted above, is exemplary. There, Nietzsche is not primarily interested in individual reactions to particular affects. Rather, he looks at the gradual, aggregative way in which cultures have reflectively reinterpreted the selfsame drives and affects. This didn't happen overnight: as Nietzsche elsewhere puts it, a new interpretation must be "constantly internalized, drilled, translated into flesh and reality" (GS 301); "from generation to generation, slowly grown onto and into the thing," until it "has become its very body" (GS 58).[31]

Just as cultures reinterpret affects in incremental ways, so too with individuals. To see this, consider Nietzsche's frequent remarks on *pity*. Nietzsche claims that "pity in your sense" is "pity with social 'distress', with 'society' and its sick and unfortunate members" (BGE 225). That is, pity is a negative feeling associated with the perception of sickness, misfortune, and, more generally, suffering. Accordingly, Nietzsche claims that pity involves a desire to alleviate another's suffering. However, Nietzsche argues that alleviating suffering would diminish human flourishing:

> You want, if possible—and there is no more insane "if possible"—*to abolish suffering*. And we? It really seems that we would rather have it higher and worse than ever. Well-being as you understand it—that is no goal, that seems to us an *end*, a state that soon makes man ridiculous and contemptible—that makes his destruction *desirable*. The discipline of suffering, of *great* suffering—do you not know that only *this* discipline has created all enhancements of man so far? (BGE 225)

Nietzsche claims that suffering has produced "all enhancements of man so far"; suffering has acted as a spur to greatness. For this reason, Nietzsche claims that he "beholds your very pity with indescribable anxiety" (BGE 225). Pity, in aiming to eliminate suffering, runs the risk of diminishing our achievements.

---

[31] A similar discussion: the *Genealogy's* claim that the bad conscience is reinterpreted as guilt.

So Nietzsche has effected a shift in the way that pity moves him. Whereas most contemporary individuals experience pity as motivating them to alleviate another's distress, Nietzsche interprets pity in a way that stills this motivational tendency. Pity, for Nietzsche, motivates nothing but the desire to rid himself of a misleading and dangerous emotion. We should deal with pity in the same way that we deal with a headache: get rid of it.

Of course, we might suspect that it won't be easy for Nietzsche to witness the suffering of another; we might suspect that pity will retain some of its traditional motivational propensities. And indeed, there are passages indicating that Nietzsche is still affected by the suffering of others. Lamenting the fact that his philosophical commitments require him to attack traditional values such as the positive valuation of pity, Nietzsche writes that "one is not always bold, and when one grows tired then one of us, too, is apt to moan 'It is so hard to hurt people—oh, why is it necessary!'" (GS 311). In his translation of the *Gay Science,* Walter Kaufmann appends to this passage a relevant extract from Nietzsche's August 20, 1880 letter to Peter Gast: "To this day, my whole philosophy totters after an hour's sympathetic conversation with total strangers: it seems so foolish to me to wish to be right at the price of love, and not to be *able to communicate* what one considers most valuable lest one destroy the sympathy." Although Nietzsche has reinterpreted his sensation of pity, on occasion it nonetheless manifests its original motivational tendency.

In passages of this form, we can see Nietzsche struggling—and sometimes failing—to shift his motives. Thus, Nietzsche writes that "we have to *learn to think differently*—in order at last, perhaps very late on, to attain even more: *to feel differently*" (D 103). A shift in thinking does not immediately result in a shift in motives.

So we should not underestimate the difficulty of shifting our interpretations of motives. I cannot simply decide, in a moment of choice, that I will henceforth experience suffering as alluring or pity as aversive. Individuals, Nietzsche thinks, will need to do a great deal of work to shift these accreted interpretations:

> Man has for all too long had an 'evil eye' for his natural inclinations, so that they have finally become inseparable from his bad 'conscience'. An attempt at the reverse would in itself be possible—but who is strong enough for it?—that is, to wed the bad conscience to all the unnatural inclinations, all those aspirations to the beyond, to that which runs counter to sense, instinct, nature, animal, in short all ideals hitherto, which are one and all hostile to life . . . . (GM II.24)

Not only will this be difficult—some aspects of our affects may be completely immutable:

> Learning changes us . . . but at the bottom of us, really 'deep down', there is of course something unteachable, some granite of *spiritual fatum* . . . . (BGE 231)

Of course, Nietzsche's claim that reflectively shifting affects is piecemeal, difficult, and sometimes unsuccessful does not imply that doing so is impossible. It simply implies that Nietzsche is realistic about the vicissitudes of human psychology: our conscious thoughts, though causally efficacious, are not guaranteed to have a *decisive* causal impact.[32]

5.2. *What Model of Conscious Willing Remains?*

So far, we have the following picture of reflective agency:

> Reflection as gradual and aggregative: reflection modifies the passions in a gradual, incremental fashion.
>
> Reflection as influenced by the passions: reflection does not enjoy any independence from the passions; on the contrary, it is everywhere influenced by them. (Denial of Suspension)
>
> The passions as influenced by reflection: the passions do not enjoy any independence from reflection; on the contrary, they are everywhere influenced by reflection.

On this model, the (Humean) division between inert reason and efficacious passion looks spurious. So, too, does the Kantian division between active reason and passive sensation. Passion and reason are both efficacious. Just as Nietzsche inveighs against treating the will as causally isolated from motives, I have suggested that he would reject the idea that *motives* are causally isolated from the will.

In this section, I will ask whether this minimal role for conscious thought leaves room for anything that deserves to be called *willing*. What kind of role does conscious thought—in particular, conscious choice—play in the production of action?

In Section 2, we saw that the Kantian theory of action maintains Suspension, Inclination, and Choice: simply put, reflection suspends the effects of motives, motives don't determine choice, and choice

[32] Presumably, this is part of why Nietzsche refers to "strong" and "weak" wills: shifting the motivational propensities of our affects is not something that happens automatically and effortlessly. On the contrary, it requires protracted engagement with those affects.

determines action. Given Nietzsche's rejection of Suspension and his model of conscious thought as operating in an incremental manner, he will need to reinterpret Inclination and Choice. However, I will show that he needn't reject them.

There are two ways of picturing Inclination and Choice: we might picture them according to the *triggering* model or the *vector* model. On the triggering model, agents have various motivational states, such as desires and affects. These motivational states *incline* or *tempt* the agent to pursue various courses of action. However, when the agent deliberates, the motivational states are *incapable* of causing the agent to act. They must await the consent of the will. Thus, the will has a triggering role; it can endorse a desire, which thereby becomes causally efficacious. This triggering model, which is endorsed by the Kantian theory of action,[33] incorporates strong versions of Inclination and Choice. Motives are capable merely of inclining us to act, and choice alone is causally efficacious. I think this has become our commonsense conception of action.

But we might also model Inclination and Choice in a more modest way. On the *vector model*, the will is simply one source of motivation among many others.[34] It can reinforce other motives, by placing its motivational weight behind them. For example, the agent's decision to go to the store produces one more motive that inclines him to go to the store. But this motive is not uniquely efficacious; the individual's action is determined by the vector of motives, including the will.

This distinction between the triggering model and the vector model can be illuminated by imagining two cases of deliberative action. In the first case, an agent is tempted to eat some ice cream. Reflecting on the desire, he decides that eating the ice cream isn't worth the calories. So he doesn't eat. This seems to fit the triggering model: a desire inclines the agent to pursue a course of action, but the desire cannot move him without the consent of the will. In the second case, an alcoholic craves another drink. The vodka is before him, but he reflects on the craving, and decides that he should resist. He does resist, for a while, but in the end he drinks after all. This seems to fit the vector model: the addiction and the motive produced by the agent's deciding not to drink compete, and in the end the addiction wins.

Nietzsche certainly rejects the triggering model of the will, for reasons that we examined above. First, the will is continuously acted upon by the agent's drives and affects, and therefore does not operate

[33] See especially the passages from Allison, Baron, Korsgaard, and Wood cited above in note 17.

[34] Compare Richardson: "Agency [i.e. the capacity to choose] is indeed a kind of drive itself" (2009, 137). It is a disposition that competes with other dispositions.

independently of them: drives and affects are continuously leading us to act and react in various ways, influencing our perceptions of the world, our reflective thoughts, and the course of our deliberation. Second, the will does not enjoy a unique capacity to determine the agent's actions; rather, the agent's actions are determined by a set of motivational forces that includes the will, drives, and affects. The conscious states are forces, too, but they are only one part—perhaps a very small part—of the total set of forces.

While Nietzsche rejects the triggering model of the will, he does not reject *every* account of the will. As we noted above, he contrasts the sovereign, strong agents, who are capable of controlling their behavior through acts of will, with the non-sovereign, weak agents, who are simply buffeted about by their drives and affects. With this point in mind, notice that the vector model of the will is far more modest than the triggering model. By accepting the vector model of the will, we can find a place for the will in the production of the action, without committing ourselves to the faculty psychology model of the will, or to the idea that the will enjoys independence from the agent's motives. When we speak of the agent's will, we simply refer to the agent's capacity to choose. These decisions are influenced and perhaps even determined by antecedent events; they are not uniquely efficacious, being one causal factor amongst others; and these episodes of choice are pervasively influenced by drives and affects. Nevertheless, the motives produced by the act of choice are, sometimes, sufficiently strong to enable the agent to act as he has chosen to act.

Suppose, in other words, that when I consciously decide to A I acquire a new motivation—possibly a very slight one—to A. My decision to A is doubtless influenced by background motives, and does not enjoy independence from my motivational states. Nonetheless, the decision yields a new motive, which may alter the antecedent balance of forces. If my motives were more or less evenly balanced beforehand, the additional motive could tip the scales. This, I suggest, is Nietzsche's model of willing.

This brings us to another point. Although every self-conscious agent has the power to make decisions, the strength of this capacity could vary across individuals. Earlier, we saw that Nietzsche wants to replace the idea of free and unfree wills with the idea of strong and weak wills (BGE 21). Again, the vector model of willing gives us a natural way of reading that passage. The capacity deliberatively to form intentions, and to remain resolute in their realization, is something that might well vary across individuals. Indeed, we already know that in certain cases it does vary: certain individuals seem to manifest more self-control than others.[35]

[35] For discussion of the empirical evidence, see for example Baumeister, Mele, and Vohs (2010) and Holton (2009).

In light of this, I suggest that Nietzsche accepts the vector model of the will. For Nietzsche allows that human beings are capable of self-conscious reflection upon their own drives and affects. Moreover, he thinks that reflective thought can have a causal impact on our motives and on our actions. Accordingly, an agent who reflects and decides to act in a certain way will, sometimes, thereby bring it about that she acts in that way.

This is why, in the later works, Nietzsche never denies that there is such a thing as *willing*. Rather, he argues that we can account for willing without committing ourselves to problematic accounts of the will, which reify the will as a faculty or treat the will as enjoying the capacity to trigger motives (cf. KSA 10:24[15]/WP 667).[36] If we confine ourselves to the modest account sketched above, there is nothing wrong with speaking of the will.

Thus, the problem with the Kantian model of willing is *not* that it includes Inclination and Choice. Inclination and Choice, interpreted in a psychologically realistic way, are true. The problem with the Kantian model is that it couples Inclination and Choice with Suspension, and is thereby led to a triggering model of the will. According to this psychologically unrealistic model of the will, the will operates as a faculty independent of the motives, enjoys causal independence from the motives, and is uniquely capable of causing action. Nietzsche roundly rejects this triggering model of the will. But he accepts the vector model, which denies Suspension and incorporates psychologically realistic versions of Inclination and Choice. According to this model, conscious thought, episodes of decision, and motives are all treated as causal forces interacting with one another. None enjoys a privileged position in the production of action.[37]

---

[36] Nietzsche never explicitly states what it is to treat the will as a faculty. However, he seems to associate treating the will as a faculty with treating it as a capacity that is independent of any influence by motives (see note 13, above). Thus, when Nietzsche claims that the will is not a faculty, or that the will just is a relation of drives, we can read these passages as emphasizing the pervasiveness of the drives' influence upon reflective thought and choice. This is just what the vector model entails: reflective thought and choice do not enjoy any independence or position of causal isolation from the drives.

[37] A potential objection: in certain passages, Nietzsche seems to attribute willing to individual drives or affects, rather than whole persons. For example, in BGE 117 Nietzsche writes that "the will to overcome an affect is ultimately only the will of another, or several other, affects." The vector model, on the other hand, suggests that willing is a product of the person. How should we make sense of this? I take it that when Nietzsche speaks of individual drives and affects "willing" things, he simply means that these drives and affects strongly dispose the person to pursue certain ends. In addition, Nietzsche uses this phrasing to draw attention to the way in which these drives and affects pervasively influence the person's conscious deliberations. Thus, willing—in the relevant sense—is indeed an attribute of the whole person. I discuss these points in detail in Katsafanas (forthcoming).

6. Conclusion

I have argued that in his early works, Nietzsche accepts the conjunction of incompatibilism and eliminativism about willing. By 1883, however, Nietzsche develops a more sophisticated conception of willing, which draws in certain respects on the Kantian theory of agency. To clarify the relationship between Nietzsche and Kant, I distinguished three of Kant's central claims about reflective agency: (i) that reflection suspends the effects of motives, (ii) that motives do not determine choice, and (iii) that choice determines action. I argued that Nietzsche endorses certain aspects of this model while rejecting others. In particular, Nietzsche endorses a complex account of motivation, which entails that reflection is not capable of suspending the influence of motives; thus, he rejects (i). Nonetheless, he maintains (ii): our motives could be the same, and yet we could choose differently. Moreover, he accepts a version of (iii), claiming that conscious choice plays a causal role in the production of action.

This interpretation of Nietzsche runs counter to a standard reading, according to which Nietzsche denies that conscious thought plays any role in the production of action. I argued against this epiphenomenalist reading of Nietzsche by showing that Nietzsche is committed to the claim that conscious thought can transform the motivational propensities of our affects.

In light of these results, I argued that we should distinguish two ways of picturing reflective agency. On the triggering model, motives are incapable of causing us to act until they are triggered by the will. On the vector model, the will is merely one motive among others; it can throw its weight behind certain motives, but it does not occupy a privileged position in the determination of action. I argued that while many Kantians are led to the acceptance of the triggering model—in part because of their acceptance of claim (i), above—Nietzsche endorses the vector model.

In sum, I have argued that Nietzsche develops a substantive and philosophically sophisticated conception of willing. *Pace* the standard readings, Nietzsche does not merely reject the Kantian conception of willing in its entirety. Rather, he critically engages with that model, shedding the components of it that seem psychologically unrealistic or predicated on problematic conceptions of motivation. So Nietzsche's model preserves certain Kantian insights about the nature of self-conscious agency, while embedding these insights in a more complex account of motivation.

In closing, notice that it remains to be seen whether Kantians could accept this more complex account of agency while preserving Kant's normative commitments, particularly insofar as these commitments are

based on Kant's account of autonomy. Nietzsche himself suggests not—he claims that Kant's account of autonomy rests on a "psychological misunderstanding," which "has invented an *antithesis* to the motivating forces, and believes one has described another kind of force; one has imagined a *primum mobile* that does not exist at all" (KSA 12:10[57]/WP 786). Thus, Nietzsche suggests, "the world to which alone [Kant's] moral standards can be applied does not exist at all" (KSA 12:10[57]/WP 786). In other words, Kant's account of autonomy and attendant moral theory rests on a robust conception of willing that Nietzsche denies. Whether Nietzsche is correct—whether the rejection of claim (i) and the triggering model of will vitiates Kant's moral theory—is a large and difficult topic, which must await another occasion.[38]

References

Reference edition of Nietzsche's works: *Friedrich Nietzsche: Sämtliche Werke, Kritische Studienausgabe in 15 Bänden*, herausgegeben von G. Colli und M. Montinari (Berlin: Walter de Gruyter, 1967–1977).

List of Abbreviations of Nietzsche's Works:

A *The Antichrist*, trans. W. Kaufmann (Viking, 1954)
BGE *Beyond Good and Evil*, trans. Kaufmann (Modern Library, 1968)
D *Daybreak*, trans. R.J. Hollingdale (Cambridge University Press, 1982)
EH *Ecce Homo*, trans. Kaufmann (Modern Library, 1968)
GM *On the Genealogy of Morality*, trans. Kaufmann (Modern Library, 1968)
GS *The Gay Science*, trans. Kaufmann (Vintage, 1974)
HH *Human, All too Human*, trans. Hollingdale (Cambridge, 1996)
KSA *Kritische Studienausgabe*
TI *Twilight of the Idols*, trans. Kaufmann (Viking, 1954)
WLN *Writings from the Late Notebooks*, trans. Kate Sturge (Cambridge, 2003)
WP *The Will to Power*, trans. Kaufmann and Hollingdale (Vintage, 1967)
Z *Thus Spoke Zarathustra*, trans. Kaufmann (Viking, 1954)

[38] For helpful comments on drafts of this essay, I would like to thank Tom Bailey, Marco Brussotti, Maudemarie Clark, João Constancio, David Dudrick, Charles Griswold, Walter Hopp, David Liebesman, Luca Lupo, Alexander Nehamas, Frederick Neuhouser, Bernard Reginster, Simon Robertson, Sally Sedgwick, Herman Siemens, Susanne Sreedhar, Daniel Star, Owen Ware, and audiences at Boston University, Leiden University, and Temple University. Thanks also to an anonymous reviewer for insightful critiques.

Allison, Henry. (1990), *Kant's Theory of Freedom*. New York: Cambridge University Press.
Baron, Marcia. (1995), *Kantian Ethics Almost Without Apology*. Ithaca: Cornell University Press.
Baumeister, Roy, Alfred Mele and Kathleen Vohs. (2010), *Free Will and Consciousness: How Might They Work?* Oxford: Oxford University Press.
Frierson, Patrick. (2005), "Kant's Empirical Account of Human Action," *Philosophers' Imprint* 5(7): 1–34.
Gardner, Sebastian. (2009), "Nietzsche, the Self, and the Disunity of Philosophical Reason," in Gemes and May (2009).
Gemes, Ken. (2009), "Nietzsche on Free Will, Autonomy, and the Sovereign Individual," in Gemes and May (2009).
—— and Simon May. (2009), *Nietzsche on Freedom and Autonomy*. Oxford: Oxford University Press.
Holton, Richard. (2009), *Willing, Wanting, Waiting*. Oxford: Oxford University Press.
Kant, Immanuel. (1996), *The Metaphysics of Morals*, Mary Gregor (ed.). New York: Cambridge University Press.
——. (1998), *Groundwork of the Metaphysics of Morals*, Mary Gregor (ed.). New York: Cambridge University Press.
——. (1999a), *Critique of Practical Reason*, Paul Guyer and Allen Wood (eds.). New York: Cambridge University Press.
——. (1999b), *Religion within the Boundaries of Mere Reason*, Allen Wood and George di Giovanni (eds.). New York: Cambridge University Press.
Katsafanas, Paul. (2005), "Nietzsche's Theory of Mind: Consciousness and Conceptualization," *European Journal of Philosophy* 13: 1–31.
——. (2011), "Activity and Passivity in Reflective Agency," in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics* 6: 219–254.
——. (forthcoming), "Nietzsche's Philosophical Psychology," in John Richardson and Ken Gemes (eds.), *The Oxford Handbook on Nietzsche*. Oxford: Oxford University Press.
Knobe, Joshua and Brian Leiter. (2007), "The Case for Nietzschean Moral Psychology," in Leiter and Sinhababu (2007).
Korsgaard, Christine. (1996), *The Sources of Normativity*. Cambridge: Cambridge University Press.
Leibniz, G.W. and S. Clark. (2000), *G.W. Leibniz and Samuel Clark: Correspondence*, Roger Ariew (ed.). Indianapolis: Hackett Publishing.
Leiter, Brian. (2001), "The Paradox of Fatalism and Self-Creation in Nietzsche," in John Richardson and Brian Leiter (eds.), *Nietzsche*. New York: Oxford University Press.

——. (2007), ''Nietzsche's Theory of the Will,'' *Philosophers Imprint*, 7.7: 1–15.
—— and Neil Sinhababu. (2007), *Nietzsche and Morality*. New York: Oxford University Press, 2007.
Locke, John. (1975), *An Essay Concerning Human Understanding*. Oxford: Oxford University Press.
McCarty, Richard. (2009), *Kant's Theory of Action*. Oxford: Oxford University Press.
Owen, David. (2009), ''Autonomy, Self-Respect, and Self-Love: Nietzsche on Ethical Agency,'' in Gemes and May (2009).
Reath, Andrews. (2006), *Agency and Autonomy in Kant's Moral Theory*. Oxford: Oxford University Press.
Richardson, John. (2009), ''Nietzsche's Freedoms,'' in Gemes and May (2009).
Ridley, Aaron. (2009), ''Nietzsche's Intentions: What the Sovereign Individual Promises,'' in Gemes and May (2009).
Risse, Mathias. (2007), ''Nietzschean 'Animal Psychology' Versus Kantian Ethics,'' in Leiter and Sinhababu (2007).
Wood, Allen. (1999), *Kant's Ethical Thought*. Cambridge: Cambridge University Press.