The Concept of Unified Agency in Nietzsche, Plato, and Schiller

Paul Katsafanas

The Concept of Unified Agency in Nietzsche, Plato, and Schiller

PAUL KATSAFANAS

LATELY, THERE HAS BEEN A PROFUSION OF WORK discussing Nietzsche's views of unity of the self, freedom, and agency.[1] A widespread consensus on three points has emerged: (1) Nietzsche's notion of unity is meant to be an analysis of freedom; (2) unity refers to a relation between the agent's drives[2] or motivational states; and (3) unity obtains when one drive predominates and imposes order on the other drives.

In this essay, I argue that these three claims are philosophically and textually indefensible. The claims are philosophically indefensible because they fail to characterize correctly certain paradigmatic cases of agency and its absence. The claims are textually indefensible because there are passages indicating that Nietzsche rejects each of them. In contrast to the standard interpretations, I argue that (1′) Nietzschean unity is an account of the distinction between genuine actions and mere behaviors, rather than between free and unfree actions; (2′) unity refers to a relation between drives and conscious thought; and (3′) unity obtains when

---

[1] Many of the relevant essays are contained in Ken Gemes and Simon May, ed., *Nietzsche on Freedom and Autonomy* (New York: Oxford University Press, 2009). Important earlier works include Amy Mullin, "Nietzsche's Free Spirit," *Journal of the History of Philosophy* 38 (2000): 382–405; Brian Leiter, "The Paradox of Fatalism and Self-Creation in Nietzsche," in *Nietzsche*, ed. John Richardson and Brian Leiter (New York: Oxford University Press, 2001); Robert Guay, "Nietzsche on Freedom," *European Journal of Philosophy* 10 (2002): 302–27; Bernard Reginster, "What is a Free Spirit? Nietzsche on Fanaticism," *Archiv für Geschichte der Philosophie* 85 (2003): 51–85; John Richardson, "Nietzschean and Kantian Freedoms," *International Studies in Philosophy* 37 (2005): 149–62; and Mathias Risse, "Nietzsche and Korsgaard's Kant on the Unity of Agency," in *Nietzsche and Morality*, ed. Brian Leiter and Neil Sinhababu (New York: Oxford University Press, 2007).

[2] 'Drive' [*Trieb, Instinkt*] is a term of art for Nietzsche. A Nietzschean drive is a non-conscious disposition toward some characteristic type of behavior; this disposition manifests itself by generating conscious affects and desires. For example, the sex drive is a non-conscious disposition toward sexual activity; it manifests itself by generating emotions such as lust, attraction, sexual desire, desires to be with particular people, and so forth. For an extended analysis of Nietzsche's notion of drive, see Paul Katsafanas, "Nietzsche's Philosophical Psychology," in *The Oxford Handbook on Nietzsche*, ed. John Richardson and Ken Gemes (Oxford: Oxford University Press, forthcoming).

Paul Katsafanas is Assistant Professor of Philosophy at Boston University.

the agent's attitude toward her own action is stable under the revelation of further information about the action's etiology. I show that Nietzsche develops this account of unity by drawing on Plato's and Schiller's models of unified agency. Nietzsche's theory incorporates elements of both Plato's and Schiller's models, but goes beyond them in emphasizing the roles of non-conscious motivation and the ubiquitous forms of self-ignorance in agency. Accordingly, the account of unified agency that emerges from Nietzsche's works is considerably more sophisticated, and potentially more philosophically fruitful, than has yet been appreciated.

The essay proceeds in five sections. Section 1 introduces Nietzsche's criticisms of traditional models of agency, and argues that he aims to replace these accounts with a model of unified agency that overcomes their flaws. Section 2 examines Nietzsche's critique of Platonic unity, and argues against the standard interpretations of Nietzschean unity. Section 3 explicates Schiller's notion of unity, and suggests that Nietzsche draws on Schiller's ideas in developing his own conception of unity. Section 4 defends this reading by examining a paradigm case of disunity, the *Genealogy*'s ascetic priest, and drawing from it an account of what unity must be. Section 5 examines the broader significance of Nietzsche's notion of unity.

1. THE PHILOSOPHICAL ROLE OF THE CONCEPT OF UNITY

1.1 Nietzsche's critiques

Nietzsche's critiques of agency take two main forms. First, he argues that traditional accounts of agency overestimate the role of reflective, self-conscious phenomena in the production of action. In this vein, he argues that an agent who self-consciously deliberates about what to do is still "secretly guided and channeled" by his non-conscious drives and motives (*BGE* 3).[3] In addition, he claims that whenever an agent steps back from and reflects upon a motive, the agent's "intellect is only the blind instrument of another drive" (*D* 109). These passages play a debunking role: while certain accounts of agency rely on a distinction between acts produced by reflective, self-conscious episodes of deliberation, and acts brought about independently of deliberation, Nietzsche argues that no such distinction is available: every reflective activity contains an admixture of influence by the non-conscious.[4]

Second, Nietzsche argues that traditional accounts of deliberative or self-conscious agency may rely on a problematic conception of the agent. Consider a few descriptions of deliberative agency. Christine Korsgaard describes the Kantian

---

[3]I cite Nietzsche's texts using the standard English abbreviations of their titles: *A* is *The Antichrist*; *BGE* is B*eyond Good and Evil*; *CW* is *The Case of Wagner*; *D* is *Daybreak*; *EH* is *Ecce Homo*; *GM* is *On the Genealogy of Morality*; *GS* is *The Gay Science*; *HC* is "Homer's Contest"; *HH* is *Human, All Too Human*; *TI* is *Twilight of the Idols*; *UM* is *Untimely Meditations*; *WP* is *The Will to Power*; *Z* is *Thus Spoke Zarathustra*. The numbers following the abbreviations refer to section numbers (as well as part numbers, if applicable). I use the Kaufmann and Hollingdale translations, though I have sometimes made minor modifications to their translations.

[4]Nietzsche interprets Kant's and Plato's accounts of agency as relying on this distinction. Whether they actually do rely on any such distinction is controversial, but will not be relevant for our purposes.

model of agency as follows: "When you deliberate, it is as if there were something over and above all of your desires, something which is you, and which chooses which desire to act on."[5] Other writers put the point differently, but agree with the underlying idea. For example, Michael Bratman writes,

> When a person acts because of what she desires, or intends, or the like, we sometimes do not want to say simply that the pro-attitude leads to the action. In some cases, we suppose, further, that the agent is the source of, determines, directs, governs the action and is not merely the locus of a series of happenings, of causal pushes and pulls.[6]

Bratman here claims that we need to distinguish the operations of the agent from the operations of the agent's attitudes. These philosophers invoke the idea of an agent or a self, who serves as something more than a mere container for the various desires and affects that cause acts. In cases of genuine action, the agent is somehow the "source of" the act. Yet Nietzsche denies that there is any self over and above the drives. The self, Nietzsche tells us, is just a "relation" or "social structure of drives and affects" (*BGE* 12). Accordingly, it is difficult to see how we could draw a distinction between acts caused by agents and acts caused by drives and affects.

Nietzsche thus seems deeply skeptical both of the idea that reflective choice is anything more than a precipitate of drives, and of the notion of selfhood typically employed in accounts of agency. However, it would be a mistake to conclude that Nietzsche rejects the distinction between genuine agency and mere undergoing.[7] On the contrary, Nietzsche is explicit about his reliance on such a distinction. Not only does Nietzsche tell us that activity [*Aktivität*] is one of his foundational concepts [*Grundbegriffe*] (*GM* II.12), he also repeatedly relies upon a distinction between genuine actions and their degenerate relatives. Thus, he praises the "sovereign" or "autonomous" individual, who is distinguished by the fact that he "has his own independent, protracted will" (*GM* II.2). While the acts of non-sovereign individuals are simply determined by whatever impulse happens to arise, the acts of sovereign individuals are controlled by the agent herself. For the sovereign individual is "strong enough to maintain [her commitments] even in the face of accidents, even 'in the face of fate.'" By contrast, the non-sovereign individual is "short-willed and unreliable"; he "breaks his word even at the moment he utters it." For the non-sovereign 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 non-sovereign individual will only fulfill his projects and goals if, through sheer luck, he encounters no temptations.

Elsewhere, Nietzsche develops these ideas, claiming that some agents have the power "not to react at once to a stimulus, but to gain control of all the inhibiting, excluding instincts. … [T]he essential feature is precisely not to 'will', to be able to suspend decision. All unspirituality, all vulgar commonness, depend on an in-

---

[5]Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), 100.

[6]Michael Bratman, *Structures of Agency* (New York: Oxford University Press, 2007), 91.

[7]For the moment, I want to remain vague on what exactly "genuine agency" is. I explicate this notion in the next section.

ability 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; she possesses no capacity to direct her own behavior. By contrast, the strong individual is able to check her impulses and resist stimuli.[8]

In these passages, Nietzsche claims that some individuals have the capacity to control their behavior. Moreover, these characterizations of agency seem to invoke the very images that Nietzsche elsewhere rejects: they suggest a self that stands apart from the drives and impulses, and exerts a controlling influence over them.

Nietzsche's characterizations of strong, sovereign agents controlling their actions suggest that there is something correct in our ordinary distinction between acts actively produced by the agent and acts in which the agent is a mere vessel for forces within. At the same time, Nietzsche critiques the way in which agency is normally understood. So he must have some alternative way of characterizing agency.

Is there a way of drawing a distinction between genuine agency and its lesser relatives, without relying on the idea of a self independent from the drives, and without denying that drives exert a pervasive influence on choice? I will argue that there is. Contemporary philosophers who attempt to distinguish genuine actions from their lesser relatives sometimes appeal to a distinction between being a genuine agent and being a mere locus of forces. Nietzsche does not accept the distinction between being an agent and being a locus of forces; this is part of what he means to deny in passages such as *BGE* 12 (quoted above). However, Nietzsche does distinguish between *different kinds* of loci of forces.

In particular, Nietzsche tells us that some agents are disunified loci of forces, whereas other agents are unified. Thus, Nietzsche argues that agents are typically multiple and fragmented. He notes that "human beings have in their bodies the heritage of multiple origins, that is, opposite, and often not merely opposite, drives and value standards that fight each other and rarely permit each other any rest" (*BGE* 200). As a result, "our drives now run back everywhere; we ourselves are a kind of chaos" (*BGE* 224). Thus, "the belief which regards the soul as … a monad, as an *atomon:* this belief ought to be expelled from science! … [T]he way is open for new versions and refinements of the soul-hypothesis; and such conceptions as 'mortal soul,' and 'soul as subject-multiplicity' [*Subjekts-Vielheit*], and 'soul as social structure of the drives and affects,' want henceforth to have citizens' rights in sci-

---

[8] There are a number of similar passages. Consider a few passages from Nietzsche's notebooks. In *WP* 95, Nietzsche condemns nineteenth-century thinkers for being "deeply convinced of the rule of cravings. (Schopenhauer spoke of 'will'; but nothing is more characteristic of his philosophy than the absence of all genuine willing)." *WP* 928 speaks of great individuals controlling their affects: "Greatness of character does not consist in not possessing these affects—on the contrary, one possesses them to the highest degree—but in having them under control." *WP* 933 makes a similar point: "In summa: domination of the passions, not their weakening or extirpation!—The greater the dominating power of a will, the more freedom may the passions be allowed. The 'great man' is great owing to the free play and scope of his desires and to the yet greater power than knows how to press these magnificent monsters into service." *WP* 962 claims that a great individual "has the ability to extend his will across great stretches of his life." Cf. *GM* II.3 and *WP* 705.

ence" (*BGE* 12). These sections have a common theme: they deny that there is a *unitary* self, and assert that the self should be understood as a relation, complex, or social structure of drives.[9]

However, Nietzsche makes it clear that this is a contingent state. While most individuals are disunified, attaining a unified self is possible: Nietzsche claims that modern individuals can be made "whole" or unified (*TI* IX.41), and he presents Goethe as an example of someone who "disciplined himself to wholeness" (*TI* IX.49). Thus, although the self is typically disunified, it is possible to attain a unified self.

In sum, while Nietzsche rejects the distinction between being an agent and being a locus of forces, he introduces a distinction between disunified and unified loci of forces. I will argue that Nietzsche's account of unity is an account of genuine agency: what it is for an agent to play an active role in producing her action is for the agent to be unified in acting.

1.2 *Three distinctions among doings*

Before proceeding, though, it will be necessary to clarify what exactly Nietzsche's account of unified agents is meant to be an account *of*. Philosophers of action typically draw at least three distinctions between the movements issuing from an agent: mere behavior, action, and autonomous (or free) action. On most accounts of agency, autonomous action and action can come apart: when I give the mugger my wallet, or when I act unreflectively out of mere custom or habit, I am *acting,* but I may not be acting *freely.*[10] So we can distinguish, among the set of actions, between the free and unfree ones.

However, we can also distinguish, among the set of movements that issue from the agent, between *mere behaviors* and *genuine actions*. Consider the distinction between movements such as sneezing, coughing, falling asleep, and blinking, on the one hand, and reading, conducting conversations, getting married, and deciding to go to Bermuda, on the other. Each of these events counts as something that a person does, in a sense, but there seem to be important differences between, say, sneezing and getting married. The former is a reflex, something that happens to me, something that is not entirely under my control; the latter is a product of choice, something that I do, something that is to some extent under my control. We can mark this distinction by calling the sneeze and its ilk *mere behaviors*, and the marriage and its ilk *actions.*

Many philosophers believe that the class of mere behaviors includes not only reflex behaviors of the sort mentioned above, but also a variety of behaviors that seem to be brought about independently of the agent's reflective thought or deliberation. For example, David Velleman claims that any case in which an agent is *ignorant* of her action, or in which she discovers what she is doing *only by observing*

---

[9] For a helpful discussion of these points, see Ken Gemes, "Post-Modernism's Use and Abuse of Nietzsche," *Philosophy and Phenomenological Research* 52 (2001): 337–60.

[10] Here I am assuming that coercion and habitual action are examples of unfreedom. Of course, not all accounts of freedom will classify these as unfree actions. Readers who prefer an alternative account of freedom can substitute examples of their own.

herself doing it, counts as a mere behavior. Harry Frankfurt argues that any case in which the agent is not *identified with* his motives is a mere behavior.[11] The idea, here, is that full-fledged actions require the agent to play an active role in the production of her own activity; anything less counts as mere behavior.

Nietzsche seems to have something like this distinction in mind when he writes, "Nothing is rarer than a *personal* action. A class, a rank, a race, an environment, an accident—everything expresses itself sooner in a work or deed, than a 'person'" (*WP* 866). Here, Nietzsche claims that what appears to be a case of a person actively bringing about an action is better described as some force acting through the person. Or, to put the point in contemporary terminology, what looks like action is really mere behavior.[12]

In sum, we have three distinctions: mere behaviors, actions, and free actions. So we should ask which of these distinctions Nietzsche's concepts of unity and disunity are meant to mark. Does the disunified/unified distinction correspond to the mere behavior/action distinction, or to the unfree action/free action distinction?[13]

Commentators on Nietzsche have assumed that his remarks on unity are meant to distinguish free actions from unfree actions: an act is free if the agent is unified, otherwise the act is unfree.[14] However, I will argue that this is a mistake. Nietzsche's remarks on unity are meant to distinguish actions from mere behaviors: what makes something an action, as opposed to a mere behavior, is that the agent is unified. On this reading, there is a further question concerning whether all unified actions are free actions: given that freedom is more demanding than unity, an agent could be unified without being free.

I think there is unambiguous textual evidence establishing that this is Nietzsche's view: unity is a necessary, but not a sufficient, condition for freedom.[15] To see this,

---

[11]See David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000), and Harry Frankfurt, *The Reasons of Love* (Princeton: Princeton University Press, 2004). Velleman actually uses the term "mere activity," rather than "mere behavior," to describe these cases.

[12]*D* 38 makes a similar distinction.

[13]Of course, a third possibility is that the remarks on unity and disunity do not correspond to any of these categories. The evidence in the following sections counts against this reading.

[14]The following works endorse this view: Alexander Nehamas, *Nietzsche: Life as Literature* (Cambridge: Harvard University Press, 1985); Reginster, "What is a Free Spirit?"; Risse, "Nietzsche and Korsgaard's Kant on the Unity of Agency"; and Leslie Thiele, *Friedrich Nietzsche and the Politics of the Soul* (Princeton, NJ: Princeton University Press, 1990). Ken Gemes also reads Nietzsche as identifying freedom with unity; see his "Nietzsche on Free Will, Autonomy, and the Sovereign Individual," in Gemes and May, *Nietzsche on Freedom and Autonomy.* However, Gemes distinguishes "agency free will" from "deserts free will." Whereas deserts free will focuses on the questions of desert, punishment, and responsibility, agency free will focuses on the question of "what constitutes an action as opposed to a mere doing." Thus, what Gemes is calling agency free will is close to what I am calling the distinction between action and mere behavior. Gemes claims that Nietzsche is interested in agency free will, not deserts free will. My interpretation is in agreement with most of these points. However, Gemes' approach collapses the threefold distinction between mere behavior, action, and free action into a twofold distinction between non-action and action. I think Nietzsche does have a threefold distinction, as I will argue below.

[15]John Richardson and Simon May accept versions of this view. Richardson argues that freedom requires both unity and genealogical insight, whereas May argues that freedom requires both unity and the overcoming of nihilism. See Richardson, "Nietzsche's Freedoms," and May, "Nihilism and the Free Self," both in Gemes and May, *Nietzsche on Freedom and Autonomy.*

notice that Nietzsche's exemplars of unity are, in general, the members of certain castes in ancient societies. Throughout his works, Nietzsche emphasizes that these individuals were free from the kinds of inner conflict that plague modern human beings: "They were more *whole* human beings" (*BGE* 257). Nietzsche claims that certain forms of social organization and certain systems of morality—the ones that were dominant in Homeric Greece, for example—are conducive to the production of unified individuals; other forms of social organization and moral systems, such as the ones dominant today, are conducive to the production of fragmented, conflicted persons. Accordingly, the Greeks of the Homeric era were, in general, unified individuals, whereas moderns are, in general, disunified individuals.[16]

However, when Nietzsche is discussing *freedom*, he often singles out traits that seem entirely absent in the Greek nobility. The discussions of freedom focus on "evaluating on one's own," "revaluating," creating new values, questioning traditional values, and actively inquiring into the history and the effects of values.[17] These are not the traits that spring to mind when we consider an Achilles or an Agamemnon. The Homeric nobles are paradigms of physical strength, health, self-assertion, and self-certainty. But they are clearly *not* examples of individuals struggling to gain independence from traditional values, or to win truth or self-understanding. Indeed, Nietzsche explicitly states that the critical stance involved in the pursuit of truth and the questioning of tradition is a distinctively modern achievement. As he puts it, "The ability to contradict, the attainment of a good conscience when one feels hostile to what is accustomed, traditional, and hallowed—that is still more excellent and constitutes what is really great, new, and amazing in our culture" (*GS* 297). Accordingly, while figures such as Achilles are unified, they do not seem to be *free*.

Thus, Nietzsche's exemplars of unity do not seem to possess the traits that are characteristic of free individuals. This suggests that unity and freedom are distinct. And in fact this point becomes clear when we consider one trait in particular: the free individual is said to be *liberated from* or *independent of* morality. As Nietzsche puts it in the *Genealogy,*

> The ripest fruit is the *sovereign individual*, like only to himself, liberated again from the morality of custom [*Sittlichkeit der Sitte*], autonomous and supramoral [*übersittliche*] (for 'autonomous' and 'moral' are mutually exclusive), in short, the man who has his own independent, protracted will. (*GM* II.2)

The sovereign individual is *liberated from the morality of custom*; he is autonomous, and therefore, Nietzsche claims, "supramoral." But Nietzsche believes that the ancient Greeks are precisely those who *embraced* the morality of custom. A few lines before the passage quoted above, Nietzsche refers the reader to his discussions of "the morality of custom" in *Daybreak*. Turning to the relevant passages in *Daybreak,* we learn that "all the communities of mankind … up to the present day" have lived under the "morality of custom" (*D* 14). Moreover, Nietzsche notes that Socrates

---

[16]For discussions of the way in which ancient societies tended to produce unified individuals, whereas modern societies tend to produce disunified individuals, see *BGE* 257, the whole of *GM*, and *HC*.

[17]See *HH* Preface 3, *HH* 225; *D* 9; *GS* 347; *GM* II.2; *BGE* 44, 211, 227; *A* 54; *EH* IV.1.

was one of the first to attempt to step beyond the morality of custom (*D* 9). This implies that, on Nietzsche's view, the pre-Socratic Greek nobles lived under the morality of custom. While they were unified individuals, they could not have been free individuals.

Nietzsche must, then, distinguish between unity and freedom. One can be unified without being free: unity is compatible with uncritical, unreflective government by certain forms of morality, whereas freedom is not.[18] So unity is at best a necessary condition for freedom. For this reason, I take it that Nietzsche's distinction between unity and disunity is meant to capture the distinction between mere behavior and action, rather than the distinction between action and free action.

2. NIETZSCHE AND PLATO ON THE UNIFIED AGENT

At this point, we know what philosophical role the concept of unity is meant to play: it marks the distinction between genuine action and mere behavior. However, we still need to determine what unity *is*.

The claim that the self is initially multiple, and that unification is an achievement, has a distinguished philosophical pedigree: we find it already in Plato. It will be useful to begin by contrasting the Platonic model with the Nietzschean model, for two reasons. First, Nietzsche self-consciously opposes his model to the Platonic model. Second, I will argue that many commentators have failed to recognize crucial respects in which Nietzschean unity differs from Platonic unity.

2.1 *Nietzsche's departures from Plato*

To begin, we will ask two questions about the Platonic and Nietzschean models. First, what are the parts into which the self is divided? Second, what kind of relation among these parts is required in order for unity to be achieved?

In the *Republic,* Plato claims that the soul has three parts: Reason, Appetite, and Spirit (*Republic* 580d–581d). Certain relations among these parts render the agent disunified, whereas others render the agent unified. In particular, the agent is unified when Reason exerts a controlling influence over the agent's action, and disunified when Appetite or Spirit reigns. As Plato puts it, it is "appropriate for the rational part to rule, since it is really wise and exercises foresight on behalf of the whole soul" (*Republic* 441e; cf. 442d).[19]

Much more could be said about Plato's view, but for our purposes this spare characterization will suffice. We can view Plato's account as consisting of two claims:

(1) Platonic Parts: The self is divided into three parts: Reason, Appetite, and Spirit.
(2) Platonic Relation: The self is unified when one of these parts, Reason, dominates the other parts.

---

[18]I take it that Nietzsche would not claim that, for any evaluative system, one can embrace that evaluative system and be unified. Rather, he claims that there are certain evaluative systems the adoption of which is compatible with unity. For example, he suggests that the evaluative systems of certain classical societies were compatible with unity, whereas Judeo-Christian evaluative systems are not compatible with unity.

[19]John Cooper, ed., *Plato: Complete Works* (Indianapolis: Hackett, 1997), 1073.

Does Nietzsche agree with Plato? Well, one difference is immediately obvious. Whereas Plato thinks the soul has three parts, Nietzsche is much more profligate: throughout his corpus he names over one hundred distinct drives.[20] Some commentators believe that this is the primary difference between the Platonic soul and the Nietzschean self: the Nietzschean self has *more* parts. For example, Thiele writes, "Platonic opposition between reason and passion is fractured into the opposition between multiple passions, each with its own capacity for reason and will to dominate."[21] On this reading, Nietzsche basically agrees with Plato, differing only on the *number* of parts.

Thus, on Thiele's interpretation, Nietzsche modifies claim (1) in the following way:

(1′) Nietzschean Parts: The self is divided into more than three parts. These parts are drives.

Claim (1′) is a common interpretation of Nietzsche, shared by Mathias Risse, Ken Gemes, and others.[22] Sections 3 and 4 will argue that (1′) is false: the parts into which Nietzsche divides the self include more than just drives. While the drives jointly compose one part of the self, there is another part: the agent's self-conscious thought. For now, though, let us proceed to claim (2).

The texts make it clear that Nietzsche disagrees with (2). Throughout his works, Nietzsche inveighs against Plato's claim that Reason should dominate the other parts: Plato "turn[ed] *reason* into a tyrant" (*TI* II.10).[23] Although Nietzsche repeatedly claims that Plato is mistaken in endorsing Reason's dominance over the other parts, Nietzsche is not as explicit as one would like about *why* this is a mistake. However, commentators generally agree on two points.

First, Nietzsche argues that Reason cannot be disentangled from Appetite. As Section 1 mentioned, Nietzsche argues that the agent's rational faculties are pervasively influenced by drives and affects. Nietzsche therefore concludes that the dominance of Reason, as Plato understands it, is impossible (or, at best, vanishingly rare). This fact renders claims about Reason's dominance problematic: if we cannot disentangle Reason and Appetite, then the claim that Reason should dominate Appetite seems untenable.[24]

Second, Plato claims that there is only *one* way to achieve unity: Reason must predominate. Nietzsche suggests that there are *many* different ways to attain unity. There is no one drive that must dominate, in order for the agent to be unified. Rather, commentators often interpret Nietzsche as arguing that unity obtains when one drive—*any* drive—exerts a dominant influence. For example, Gemes claims that "it is when a strong will [i.e. drive] takes command, orders and organizes lesser drives" that a person manifests genuine agency.[25] Richardson suggests

---

[20]For a discussion of this point, as well as an extended analysis of Nietzsche's concept of drive, see Katsafanas, "Nietzsche's Philosophical Psychology."

[21]Thiele, *Friedrich Nietzsche and the Politics of the Soul,* 56.

[22]See Risse, "Nietzsche and Korsgaard's Kant on the Unity of Agency," and Gemes, "Nietzsche on Free Will, Autonomy, and the Sovereign Individual."

[23]For further remarks to this effect, see *TI* II.1–12, *TI* V.1, *BGE* 191, and *WP* 848.

[24]For an extended discussion of this point, see Katsafanas, "Nietzsche's Philosophical Psychology."

[25]Gemes, "Nietzsche on Free Will, Autonomy, and the Sovereign Individual," 42.

that Nietzsche associates unity with the emergence of "a single dominant drive, or perhaps a ruling 'committee' of drives."[26] Thiele claims that unity is achieved "through the harnessing of violent and varied passions, and their placement under the rule of a predominant drive."[27]

Thus, a number of commentators assume that while Nietzsche disagrees with the details of Plato's account, he accepts Plato's basic claim that the self is unified when one part dominates and imposes order on the other parts. Suppose we break Plato's claim into two parts:

(2a) Platonic Relation: The self is unified when one part dominates, and

(2b) This part must be Reason.

On the standard reading of Nietzschean unity, Nietzsche accepts (2a) and rejects (2b).

However, I think this is a mistake. Nietzsche's departure from Plato is more radical, for he also rejects (2a). He denies the idea that unity is achieved via one part's dominance over the other parts. While Nietzsche does think that dominance of one part *frequently causes* unity, he denies that dominance is the *same thing as* unity. Below, I argue for this point. I will first consider textual evidence in favor of the standard reading of Nietzschean unity, which claims that unity obtains when any drive exerts a dominant influence on the other drives. I will then offer some philosophical and textual reasons for rejecting this interpretation.

2.2 *Textual evidence for the claim that unity is the predominance of one part*

Consider two passages from Nietzsche's *Nachlass:*

The multitude and disgregation of drives and the lack of any systematic organization among them results in a "weak will"; their coordination under a single predominant drive results in a "strong will"; in the first case it is the oscillation and lack of gravity; in the latter, the precision and clarity of direction. (*WP* 46)

The antagonism of the passions: two, three, a multiplicity of "souls in one breast": very unhealthy, inner ruin, disintegration, betraying and increasing an inner conflict and anarchism—unless one passion at last becomes master. Return to health—. (*WP* 778)

In the first passage, Nietzsche tells us that the will is strong when one drive predominates and coordinates the other drives. If we assume that Nietzsche's talk of "strong" wills is meant to refer to unified wills, then Nietzsche seems to be offering a straightforward analysis of unity: an agent is unified if one of his drives coordinates the other drives. The second passage makes a similar point: an agent is unhealthy and experiences inner conflict when there is no drive that has become master (i.e. no drive that is dominant).

This seems plausible. The predominance of one drive seems to provide opportunities for locating the agent in the production of action. In particular,

(a) We could identify the agent's acts with the acts caused by the dominant drive.

---

[26]Richardson, "Nietzsche's Freedoms," 134–35.

[27]Thiele, *Friedrich Nietzsche and the Politics of the Soul,* 63.

(b) Alternatively, if the agent's will were identified with the whole set of drives, then the agent's will would be unified.

By contrast, consider an individual whose drives exhibit anarchy: the drives are inconsistent, at odds with one another, and no one drive is predominant. Such an individual would lack both of the conditions described above:

(a′) There is no *one* drive whose operations could be regarded as expressive of the agent. For no drive enjoys predominance over the others.

(b′) If the agent's will were identified with the whole set of drives, then the agent's will would be disunified, a chaotic mix of warring fragments.

So we can see why it is tempting to think that predominance of one drive constitutes unity.

There is, however, a complication. Nietzsche frequently praises individuals who harbor diverse, inconsistent, conflicting drives:

> In contrast to the animals, man has cultivated an abundance of *contrary* drives and impulses within himself: thanks to this synthesis, he is master of the earth. … The highest man would have the greatest multiplicity of drives, in the relatively greatest strength that can be endured. Indeed, where the plant "man" shows himself strongest one finds instincts that conflict powerfully (e.g. in Shakespeare) but are controlled. (*WP* 966)

> A philosopher—if today there could be philosophers—would be compelled to find the greatness of man, the concept of "greatness," precisely in his range and multiplicity, in his wholeness in manifoldness. … Precisely this shall be called *greatness:* being capable of being as manifold as whole, as ample as full [*ebenso vielfach als ganz, ebenso weit als voll sein können*]. (*BGE* 212)

On the face of things, these passages count against the idea that unity consists in dominance of one drive. Nietzsche's exemplars of unified selfhood—above he mentions Shakespeare, and elsewhere Goethe, Napoleon, and Nietzsche himself are the paradigms—are praised precisely because they have diverse, powerful, and inconsistent drives, but are *in some other sense unified.*

However, proponents of the predominance model claim that we can account for this point by distinguishing different forms of predominance by one drive. Chief among these are tyranny and mastery.[28] Tyranny consists of one drive's achieving a predominant status by suppressing or extirpating other drives (*HH* I.228, *GS* 347). In other words, drive *A* tyrannizes drives *B* and *C* when *A* becomes stronger than *B* and *C* by weakening or eliminating *B* and *C*. Nietzsche offers asceticism as a paradigmatic form of tyranny. Mastery, by contrast, consists of one drive's being predominant, but still allowing other drives expression. In other words, drive *A* masters drives *B* and *C* when *A* becomes stronger than *B* and *C*, and modulates the expression of *B* and *C*, yet does not weaken or eliminate *B* and *C*. An example might be a dominant drive toward intellectual activity modulating the expression of, say, the hunger drive and the sex drive; the intellectual drive might master these drives in the sense that the agent allows the latter drives expression only when doing so does not interfere with the expression of the intellectual drive. The

---

[28]Again, the connection to Plato should be clear. Plato distinguished five different types of unity and disunity: aristocracy, timocracy, oligarchy, democracy, and tyranny.

idea, then, would be that Shakespeare, Goethe, and Nietzsche's other exemplars harbor a drive that predominates through mastery.[29]

Thus, we seem to have an account of unity. The self is unified when one drive predominates, and exerts a coordinating influence on the other drives. Conflict among drives does not have to be *eliminated*; it just has to be *managed*.

2.3 *Textual and philosophical reasons for rejecting the reading of unity as predominance*

The predominance model seems to provide a tidy answer to the question of what Nietzsche's account of unity is. However, I think that upon examination this account becomes far less plausible, for two reasons.

First, there is an obvious philosophical problem with the view that unity is dominance by one part of the self: it assumes that the dominant part of the self has some special claim to being expressive of the self. But this assumption is unwarranted, for we often distinguish actions produced by the agent from acts caused by the agent's dominant motive. For example, imagine an alcoholic who ardently craves a drink, but judges that he should resist. Suppose the craving eventually overpowers the agent's resistance. Here, the craving for alcohol is the strongest motivational force, but it would be perverse to say that when the agent acts on that craving, he manifests agential control. On the contrary, the voice of the agent seems to reside in the weak, overpowered element of resistance. It is for good reason, then, that we distinguish acts that are expressive of the agent from acts that are expressive of the strongest motivational force.

A proponent of the unity-as-predominance model might respond to this objection by claiming that the alcoholic urges should be regarded as operating through tyranny rather than mastery. That is, if unity requires predominance in the form of mastery, and if alcoholics exhibit predominance only in the form of tyranny, then alcoholics would not serve as counterexamples to the unity-as-predominance model.[30]

However, this response on behalf of the predominance model does not seem promising. In order for the response to succeed, one would have to show that there are *no* cases in which alcoholism operates via mastery. This is implausible. Recall that mastery simply requires that one drive modulates, but does not weaken or extirpate, other drives. A number of alcoholics seem to fit this description. For example, "high-functioning" alcoholics are defined as those who maintain stable and successful lives, often for many decades, despite an addiction to alcohol. Many of these agents have rich arrays of passions and drives that are subordinated to, but not weakened or extirpated by, their craving for alcohol.[31] Thus, the alco-

---

[29]Reginster's "What is a Free Spirit?" discusses the notions of anarchy, tyranny, and mastery in detail. See also Ken Gemes, "Freud and Nietzsche on Sublimation," *Journal of Nietzsche Studies* 38 (2009): 38–59. Gemes uses the terms 'sublimation' and 'repression' to pick out conditions similar to those that Reginster labels 'mastery' and 'tyranny.' Employing these notions, Gemes provides a very helpful discussion of the possible configurations of drives.

[30]Thanks to an anonymous referee for asking me to address this point.

[31]Ernest Hemingway is sometimes cited as an example of a high-functioning alcoholic. Hemingway reportedly drank a quart of liquor per day for most of his adult life. Sometimes he drank even more:

holic urges should count as mastering the other drives. If the proponent of the predominance model insists that these high-functioning alcoholics, despite their manifold accomplishments and varied activities, are tyrannized by their drive for alcohol, then one starts to lose any grip on what the distinction between tyranny and mastery is supposed to be.

(In further support of this point, notice that high-functioning alcoholism usually does not last for the individual's whole life. In typical cases, the alcoholic's condition lasts for several years and sometimes even several decades, but eventually deteriorates. The alcoholic becomes incapable of maintaining his usual routines: he loses his job, his other passions, his friends, and so on. After this collapse in functioning, the craving for alcohol operates in a very different way: the craving extirpates or severely weakens competing drives, and becomes all-consuming. Given the definitions of mastery and tyranny, the correct characterization seems to be this: prior to the collapse in functioning, the alcoholic craving operates via mastery; after the collapse, it operates via tyranny.)

I conclude that certain alcoholics do indeed serve as counterexamples to the unity-as-predominance model. Alcoholics who are mastered by their alcoholic urges would count as unified according to the predominance model, and yet it seems perverse to claim that when such an alcoholic succumbs to his addiction he is manifesting agential control.

So the first problem with the predominance model is that we often distinguish acts caused by the agent's strongest motive from acts produced by the agent herself. But there is also a second problem with the predominance model: there is textual evidence that Nietzsche dissociates unity and dominance. After all, he derides those of us who, like the alcoholic, become "as a whole the victim of some part of us [*als Ganzes das Opfer irgend einer Einzelheit an uns werden*]" (*BGE* 41). More decisively, one of Nietzsche's paradigms of dominance by one drive is *also* a paradigm of disunity. Nietzsche claims that Richard Wagner's personality is controlled by one drive:

> The dramatic element in Wagner's development is quite unmistakable from the moment when his ruling passion became aware of itself and took his nature in its charge: from that time on there was an end to fumbling, straying, to the proliferation of secondary shoots, and within the most convoluted courses and often daring

---

over one sixth-month period in the early sixties, he seems to have consumed eighteen bottles of liquor and 660 bottles of wine. Aside from the sheer amount that he drank, the signs of alcoholism were clear: he was unable to stop drinking when his doctors advised him to do so, and he seemed highly dependent on alcohol—witness his claim that "You wake up in the night and things are unbearable and you take a drink and make them bearable" (quoted in Jeffrey Meyers, *Hemingway: A Biography* [New York: De Capo, 1999], 539). This intense drinking seems to have begun in the 1920s. The physiological effects started to appear in the 1950s, and progressively worsened until Hemingway's suicide in 1961—a reporter who interviewed Hemingway in 1958 noted that Hemingway's "liver was bad. You could see the bulge of it stand out from his body like a long, fat leech" (Meyers, *Hemingway*, 539). Nevertheless, for many decades Hemingway maintained a rich, varied, and productive life. By any plausible standard, we must conclude that his alcoholism allowed his other drives expression—after all, he maintained an active social life, he flourished as a writer, he had diverse intellectual interests, he traveled extensively, and so on. Hemingway thus seems to be a paradigmatic case of an individual whose alcoholic cravings master, rather than tyrannize, the other drives. Nonetheless, Hemingway seems passive in the face of his alcoholism.

> trajectories assumed by his artistic plans there rules a single inner law, a will by which they can be explained. (*UM* III.2)

Nietzsche here claims that Wagner's rich and diverse personality, with manifold interests and passions, was dominated by one ruling passion or drive.[32] Yet Nietzsche treats Wagner as a paradigm of *disunity*: he is "the *most instructive* case" for a philosopher interested in a "*diagnosis of the modern soul*" (*CW*, Epilogue; italics in original). As we saw above, Nietzsche defines the 'modern soul' as the disunified soul. If Wagner is supposed to be an exemplar both of one drive's dominance *and* of disunity, then unity cannot be identical with one drive's dominance.[33]

Thus, there are compelling philosophical and textual reasons for distinguishing unity and the predominance of one drive. So we need a new account of unity.[34]

3. SCHILLER ON UNITY AS HARMONY

On Plato's account, the alcoholic individual would be regarded as dominated by Appetite, rather than by Reason. Thus, Plato would have good reason to judge the alcoholic defective as an agent. This seems to be the right result. The suggested reading of Nietzsche, above, was supposed to be neutral on which drive dominated; unity was supposed to consist of domination by *any* drive. So the suggested reading cannot judge the alcoholic agent to be defective.

What Plato's model is capable of capturing, and the predominance model is not, is the fact that when the agent's reflective judgments conflict with the agent's predominant motive, we regard the agent as overpowered by a part of himself. Nietzsche himself would agree with Plato on this much. For, as we saw in Section 1, Nietzsche associates genuine agency with the ability to control one's behavior via choice. We are still attempting to explain exactly how Nietzsche understands these notions, but the unity as predominance model simply ignores them. So we need a different model.

At this point, it will be helpful to introduce another model of unity, which would have loomed large in Nietzsche's mind. Friedrich Schiller proposed a model of unified agency, which he conceived as an alternative to the Kantian model of

---

[32]The claim that Wagner is dominated by one drive is present not only in early works such as *UM*, but also in some of Nietzsche's very last works. For example, in *CW* Nietzsche repeatedly states or implies that Wagner is dominated by one drive. Nietzsche writes that "one cannot figure out Wagner until one figures out *his dominant drive*" (*CW* 8; cf. *CW* 11). He explains, "One does not understand a thing about Wagner as long as one finds in him merely an arbitrary play of nature, a whim, an accident. He was no 'fragmentary', 'hapless', or 'contradictory' genius, as people have said. … If anything in Wagner is interesting it is the logic with which a physiological defect makes move upon move and takes step upon step in practice and procedure, as innovation in principles, as crisis in taste" (*CW* 7). Nietzsche's suggestion, then, is that some drive or "physiological defect" dominated Wagner and imposed an overall order on his life.

[33]See also *GM* III.4, where Nietzsche claims that Wagner experienced "a deep, thorough, and even frightful identification with and descent into medieval soul-conflicts."

[34]I do not deny that Nietzsche is interested in the psychic conditions of predominance. My claim is simply that Nietzsche does not identify the condition of predominance with the unity that is necessary in order for the agent to play an active role in producing the action. The account of mastery is nothing more than what Nietzsche explicitly says: it is an account of "strong" or "healthy" wills (*WP* 46, *WP* 778). But a will can be strong, in this sense, without *the agent* playing any role in the production of the action.

agency.[35,36] Nietzsche was familiar with Schiller's work, and, like Schiller, took Kant's model of agency as a target and a point of departure. So it stands to reason that Nietzsche drew from Schiller's model.

Schiller begins by accepting a version of the Kantian distinction between reason and sensibility. The individual, Schiller tells us, has two aspects: a rational nature, manifest in judgment and self-conscious thought, and a sensible nature, manifest in sense perception and affects.[37] These two aspects of human nature can be related in three different ways.[38]

First, the individual might be dominated by his sensible nature, merely acting on whatever inclination happens to arise. "Prey to desire," he "lets natural impulse rule him unrestrainedly" (NA 280/147). Schiller terms such an individual *ochlocratic* (i.e. ruled by a mob). Schiller denounces this type of individual, claiming that he is analogous to a failed state in which citizens do not acknowledge the legitimacy of their sovereign (NA 282/148).

Second, the individual might be dominated by his rational nature. Schiller takes Kant to endorse this state of the soul (NA 282–85/148–50). Schiller terms such an individual *monarchic*; his rational nature rules his sensible nature with "strict surveillance" (NA 281–82/148). Schiller claims that the monarchic agent is better off than the ochlocratic agent, for his actions will be in accordance with the balance of reasons, and will have moral worth.

Although the monarchic agent is superior to the ochlocratic agent, Schiller finds something problematic about *both* of these agents: namely, the fact that one *part* of the individual dominates the other part. "This much is clear: that neither the will … nor the affect … ought to use *force*" (NA 279/146). He endorses a third state: harmony between the rational and sensible parts of the soul. A harmonious individual would have affects that incline her to pursue the very same ends that rational thought inclines her to pursue. Like the monarchic agent, her actions would be in accordance with the balance of reasons. But unlike the monarchic agent, there would be no struggle, no antagonism, in the soul of this agent. Her whole being would incline her in one direction:

> It is only when he gathers, so to speak, *his entire humanity* together, and his ethical way of thinking becomes the result of the united effect of both principles [e.g. Reason

---

[35]Schiller develops these ideas in several works, including his essay "On Grace and Dignity" and the *Letters on the Aesthetic Education of Man*. For the sake of brevity, I will confine my discussion to "On Grace and Dignity." References to this essay are in the following format: NA followed by page number refers to the pagination in Schiller's *Nationalausgabe*; the second page number refers to the English translation of the essay in Jane Curran and Christopher Fricker, *Schiller's "On Grace and Dignity" in Its Cultural Context* (Rochester, NY: Camden House, 2005), 123–70.

[36]Kant responds to Schiller in a footnote to *Religion within the Limits of Mere Reason* (6:23–24n). For a helpful discussion of what is at issue in the debate between Schiller and Kant, see A.M. Baxley, "The Beautiful Soul and the Autocratic Agent: Schiller's and Kant's 'Children of the House,'" *Journal of the History of Philosophy* 41 (2003): 493–514.

[37]Schiller claims that the human being's "purely intellectual nature is accompanied by a sensuous one" (NA 284/149). He discusses this point at NA 257/128–29, NA 262/132, NA 266/135–36, NA 272–73/140–41, and NA 276–78/144–45.

[38]"One can think of three ways altogether in which a human can relate to himself, that is, in which the sensuous part can relate to the rational" (NA 280/147).

and Sensibility], when it *has become his nature*, it is then only that it is secure. (NA 284/150)[39]

> The human being has been set the task of promoting a sincere accord between his two natures, of always being a harmonious whole, and of acting with his whole harmonious humanity. (NA 289/154)

Schiller calls the harmonious individual the beautiful soul. Her actions have not only dignity, but *grace* (*Anmut*).

Thus, Schiller holds that an agent is unified when the two aspects of the soul—rational nature and affective nature—are harmonious, directing the agent toward the same ends. Disunity arises when there is a conflict between the rational and the affective, which takes the form of reason being out of accordance with the affects. In short: unity obtains when the agent's reflective judgments and affects incline her in the same direction. Thus,

- Schiller's Parts: The self is divided into two parts: the rational and the sensible
- Schiller's Relation: The self is unified when the rational and sensible parts incline the agent toward the same ends.

There are two interesting parallels between Schiller's account and Nietzsche's account. Like Nietzsche, Schiller denies that unity can be achieved via one part's dominance of the other part(s). Moreover, Nietzsche describes his paradigm of unity (Goethe) in terms that are reminiscent of Schiller:

> What [Goethe] wanted was *totality*; he fought against the separation of reason, sensation, feeling, and will [*das Auseinander von Venunft, Sinnlichkeit, Gefühl, Wille*] (preached with the most abhorrent scholasticism by *Kant*, Goethe's antipode); he disciplined himself to wholeness. (*TI* IX.49)

Nietzsche tells us that Goethe fought against the "separation" [*das Auseinander*] of reason, sensation, feeling, and will. The reference to Kant suggests that Nietzsche is making Schiller's point: Kant allegedly thought that reason should dominate passion. Nietzsche, with Schiller, conceives a harmony between the various aspects of the soul. Goethe exemplifies that state.

So Nietzsche seems to accept a version of Schiller's idea, that unity is attained when the parts relate to each other in a harmonious way, rather than when one part dominates the other parts. In addition, I will suggest that Nietzsche accepts and develops Schiller's basic point: the agent is unified when there is a harmonious relationship between the agent's reflective thought and the agent's affects. Notice that, in the passage quoted above, Nietzsche is *not* referring to a unity among drives alone. Rather, he speaks of attaining unity among *reason, sensation, feeling,* and *will.* Now, I think it would be a mistake to put too much weight on the particular terms that Nietzsche is using. For example, we should not interpret Nietzsche as claiming that the self has exactly four discrete parts that must be unified, for Nietzsche elsewhere rejects claims of this form. However, it is interesting that Nietzsche speaks of a unity not among drives alone, but among drives and

---

[39]Here I have departed from the translation in Curran and Fricker, which seems to me to obscure Schiller's point. In the next passage, I have also made some minor modifications to the translation.

other parts of the self. This helps us to make sense of Nietzsche's claim that unity is compatible with conflict among drives (see section 2.2). If Nietzschean unity were unity among drives, then Nietzsche's praise of conflict among drives would be odd. However, if Nietzschean unity were unity between, say, drives and reason, then conflict among drives need not be relevant. Take an analogy: Plato does not worry about the fact that the part labeled Appetite contains, within itself, many conflicting particular appetites. *That* kind of conflict does not endanger unity, because Platonic unity is unity between Appetite, Reason, and Spirit. Just so, I suggest, Nietzsche is not inconsistent when he praises individuals with conflicting drives, for Nietzschean unity is *not* unity between particular drives, but unity between drives and other parts of the individual.[40]

What are these parts? I suggest that they are roughly the ones that Schiller focuses upon: Reason and Sensibility. Of course, Nietzsche would not characterize the parts in these terms, nor would he agree with Schiller on the exact nature of these parts. However, Nietzsche draws attention to a closely related state. This is the state in which the agent experiences a particular kind of discrepancy or discord between his reflective judgments (Reason), on the one hand, and his drives and affects (Sensibility), on the other hand. To explain this point, I am going to spend some time examining one of Nietzsche's paradigms of disunity: the priests of the *Genealogy*.

4. AN ANALYSIS OF NIETZSCHEAN UNITY

In this section, I argue that we can grasp Nietzsche's concept of unity by examining a paradigm of disunity: the *Genealogy's* ascetic priest.[41] I argue that Nietzschean unity consists in a certain relation between an agent's reflective and unreflective aspects at the time of action. Although many commentators have assumed that Nietzschean unity is a unity between drives, I argue that there are persuasive textual grounds for interpreting Nietzschean unity as a unity between drives and self-conscious thought.[42]

---

[40]I do not mean to suggest that proponents of the predominance model have no way of explaining Nietzsche's praise of conflict among drives. On the contrary, they do have ways of doing so. For example, one could argue that Nietzsche wants *both* the predominance of one drive *and* constant struggle among the dominated drives for predominance. There is nothing incoherent about this proposed solution. However, my suggested interpretation has the advantage of rendering these sorts of proposed solutions completely unnecessary. For, on my interpretation, there is no *problem* to solve: Nietzsche's praise of conflict among drives only looks puzzling and problematic if we assume that the unity he desires is a unity among the drives. So my proposed solution has the advantage of *dissolving* the apparent problem, thereby rendering a quest for solutions unnecessary. Notice that if my reading is correct, it has the added benefit of explaining why Nietzsche himself never seems to detect any potential tension between his endorsement of agential unity and his valorization of conflict among the drives. (Thanks to an anonymous referee for pressing me on this point.)

[41]Nietzsche frequently emphasizes the priest's disunity. The ascetic priest is characterized as "a self-contradictory type" (*GM* III.11), manifesting "inner corruption" (*GM* III.14) and "ill-constitutedness" (*GM* III.13). Nietzsche further criticizes the priest for being "discordant," a "self-contradiction" (*GM* III.13). He is the "shepherd" of the "ill-constituted" (*GM* III.13); he seeks to make the "well-constituted" suffer (*GM* III.14), he "walk[s] among us as embodied reproaches, as warnings to us—as if health, well-constitutedness, strength … were in themselves necessarily vicious things" (*GM* III.14).

[42]This fact helps to explain why Nietzsche treats disunity as a condition peculiar to self-conscious animals. A non-self-conscious animal, on Nietzsche's view, cannot be disunified.

4.1 The character type "priest"

The First Essay of the *Genealogy* introduces three "character types": the slave, the noble, and the priest. Although my goal is simply to characterize the psychic state of the priest, describing this psychic state requires a brief reconstruction of the main argument in the First Essay.[43,44] I should note that the First Essay's argument is highly complex, and there are a number of controversies surrounding the structure of the argument. For our purposes, nothing important hangs on the precise way in which the pieces of the argument fit together, so I will pass over these controversies, simply mentioning them in the footnotes.

Nietzsche begins by juxtaposing two types of individuals: the heroic, strong, healthy nobles, and the weak, craven, downtrodden slaves. Initially, both types of individuals accept a system of values that posits health, strength, and beauty as the chief goods, and sickness, weakness, and ugliness as the chief forms of badness. By the light of these values, the slaves are bad, and the nobles good. Thus, the slaves are committed to regarding themselves as bad, unfortunate wretches. Moreover, they do not see this as a temporary or surmountable state of affairs; rather, they view themselves as irredeemably bad (*GM* I.6).

Nietzsche next introduces a third type of individual, the priest (*GM* I.6–7, I.16). Like the slaves, the priests are weak and unhealthy. Unlike the slaves, the priests are not content to resign themselves to this state of affairs. On the contrary, the priests have a "lust to rule," an ardent desire to occupy positions of power and influence (*GM* I.6). Thus, they are engaged in a struggle for dominance with the nobles (*GM* I.6–10). However, the priests have none of the traits that are regarded as valuable: they are not physically strong, healthy, and so forth (*GM* I.6–7).

Frustrated by their inability to attain dominance, the priests come to bear an intensely negative affect toward the nobles, an affect that Nietzsche calls *ressentiment. Ressentiment* has several distinctive features. First, and most obviously, *ressentiment* involves a negative affect of hatred and vengefulness (*GM* I.10, III.14). This hatred is directed at the nobles (*GM* I.10). It is crucial that the priest hates the noble for a particular reason: he wants to possess the characteristic traits of the noble—he wants to be beautiful, strong, wealthy, and healthy—but finds himself completely incapable of doing so. In other words, the priest finds himself unable to realize the form of life that he regards as most valuable, and is able to live only in ways that he regards as disvaluable. Confronted with those who enjoy the form of life that he values and ardently desires, the priest comes to hate the nobles. But notice that this intense hatred is directed at those who, by the priest's own lights, are living the good life; moreover, the priest hates them precisely *because* they are leading the good life.

So the priest has an odd response to the noble: he values their way of life, and for that reason it would make sense for him to regard the nobles with admiration

---

[43]The slave also provides an excellent example of disunity. However, I have chosen to focus on the priest, for the psychic tensions that I wish to highlight are somewhat clearer and more pronounced in the priest.

[44]My account of the first essay draws on Bernard Reginster, "Nietzsche on *Ressentiment* and Valuation," *Philosophy and Phenomenological Research* 57 (1997): 281–305, and Jay Wallace, "*Ressentiment, Value, and Self-Vindication,*" in Leiter and Sinhababu, *Nietzsche and Morality*.

or respect; but, instead, the priest hates the nobles. This creates a discord between the priest's reflective evaluations and his emotions.

Typically, emotions and evaluative judgments are reciprocally related. For emotions involve a claim to rationality: they have standards of intelligibility.[45] Consider an example. Hating a person commits one to viewing the person as in some respect bad. There are borderline cases: you can, I suppose, hate someone for no reason, though this borders on unintelligibility. But it does not make sense to hate someone whom you regard as having wholly positive attributes, for hating a person involves judging the person to be in some respect disvaluable.

The priests violate this standard. The nobles embody all the traits that the priests reflectively evaluate as good, yet the priests have an intensely negative affect directed at the nobles.

This kind of discord between affect and evaluative judgment can be eliminated in two ways. The agent can attempt to change his affects, bringing his affects into accordance with his evaluative judgments. Or the agent can attempt to change his evaluative judgments, bringing his evaluative judgments into accordance with his affects. The priests take the second route. As Jay Wallace puts it,

> The powerless find themselves in a conceptual situation in which the negative affect that dominates their lives is directed at individuals whom they themselves seem compelled to regard as exemplars of value and worthy of admiration. This is a highly unstable combination of attitudes. … In my view, the slave revolt should be understood as a response on the part of the slavish to this psychic tension. The weak are subject to attitudes that color their experience of the social world, in ways that cannot be reconciled with the dominant ethical ideology that they themselves have so far accepted. So they come to embrace a new and more congenial scheme of values.[46]

This new set of values labels health, strength, and beauty as evil, and sickness, weakness, and ugliness as good (*GM* I.7). In effect, it inverts the older system of values. By the lights of this new set of values, the noble is evil. So the priest's *ressentiment* makes sense: if the noble is evil, then the priest is justified in bearing hatred or *ressentiment* toward the noble.[47]

---

[45]See, for example, *D* 119, 279; *GS* 14, 39; *Z* I.15; *BGE* 187, 220, 230, 260, 268, 284; *GM* II.16, III.12; *TI* IX.20. For a very helpful discussion of Nietzsche's views on this point, see Peter Poellner, "Affect, Value, and Objectivity," in Leiter and Sinhababu, *Nietzsche and Morality.*

[46]Wallace, "*Ressentiment,* Value, and Self-Vindication," 220. Wallace's language is misleading: Nietzsche certainly would not be troubled by the fact that an agent bears "a highly unstable combination of attitudes." After all, Nietzsche explicitly praises those who bear contradictory, unstable affective states. For example, he writes that the individual with "contradictory drives" has "a great method of acquiring knowledge: he feels many pros and cons. … The wisest man would be the one richest in contradictions" (*WP* 259; cf. *HH* 618, *GM* III.12, and *BGE* 284). Thus, Wallace's concerns about unstable combinations of attitudes seem out of place. However, I think the idea that Wallace is attempting to express with his claims about "unstable attitudes" is exactly right: as I will explain below, the problem with the priests is that they exhibit a discrepancy between affects and evaluative judgments.

[47]The values preached by the priests come to be accepted by the slaves. It is easy to see why: so long as the slaves embrace the older system of values, they are committed to regarding themselves as bad. Thus, they will experience shame and frustration at their current condition. If they accept the new system of values, they will regard themselves as good. Under the influence of a desire to think well of themselves, the slaves come to accept the priestly values. This is what Nietzsche terms the "slave revolt" in values: the new system of values takes root among the slaves.

Let us summarize the argument so far. The priests come to bear a negative affect of *ressentiment* toward the noble. This creates a state of discord between affect and evaluative judgment; the priests have an intensely negative affect directed at those who epitomize their own values. The priests respond to this state of tension by altering their evaluative judgments.[48]

There is one more important complication. Recall that the priest came to resent the noble because the priest was engaged in a struggle for dominance with the noble. That is, the priest wanted to occupy positions of political power, wanted to dominate, and so forth. After the revaluation, the priest judges traits such as the desire for dominance to be evil. However, Nietzsche makes it clear that the priest continues to desire to be dominant. Indeed, his very process of revaluing values was undertaken because he ardently desired dominance. As Bernard Reginster puts it, "The priests who so vehemently condemn the thirst for 'spoil and victory' of the noble 'blond beast' (*GM* I.11) are in fact pursuing the very same 'goals … victory, spoil, and seduction'" (*GM* I.8).[49]

So although the revaluation eliminates one form of discord between affect and evaluative judgment, it gives rise to a *second* form of discord between affect and evaluative judgment: the priest ardently desires that which he judges evil. Confronted with the discrepancy between affect and evaluative judgment, the priest resorts to self-deception:

> The man of *ressentiment* is neither upright nor naïve nor honest and straightforward with himself. His soul *squints*. (*GM* I.10)

> [The man of *ressentiment*] has, thanks to the counterfeit and self-deception of impotence, clad itself in the ostentatious garb of the virtue of quiet, calm resignation, just as if the weakness of the weak … were a voluntary achievement, willed, chosen, a *deed*, a *meritorious* act. (*GM* I.13; cf. I.14)

Through self-deception, the priest hides from himself the fact that he desires that which he condemns as evil.

4.2 *Important features of the character type*

Notice that the priests exhibit *precisely* the structure that Schiller identifies with disharmony: their reflective judgments are out of accordance with their affects; the two parts direct the agent toward opposing ends. As Schiller would put it, the rational and the sensible parts of the self are in conflict.

---

[48]This interpretation is somewhat controversial. A different, and perhaps more familiar, interpretation claims that the priests engage in revaluation not in order to eliminate their own psychic tension, but in order to succeed in their struggle for dominance with the nobles. On this interpretation, the priests believe that if they can convince the slave class that the nobles are evil, then the priests will be able to assume a dominant position. So the priests engage in revaluation in order to fulfill their desire for dominance. I think this interpretation faces a number of difficulties, which are discussed by Wallace in "*Ressentiment,* Value, and Self-Vindication." However, nothing that I am about to say about the psychic state of the priest hinges on this point. Readers who prefer the second interpretation of the priestly revaluation can substitute this interpretation for the one offered in the text, above. For my purposes, the only important point is that the priests have a form of discord between affect and evaluative judgment. It does not matter whether this discord motivates the revaluation, or plays a less substantial role.

[49]Reginster, "Nietzsche on *Ressentiment* and Valuation," 291.

This discord between affect and evaluative judgment shows up at two crucial junctures of the *Genealogy*. First, when the priest begins to bear *ressentiment* toward the noble, the discord between affect and evaluative judgment motivates the priest's revaluation of values. Second, after the revaluation, the priest experiences discord between his ardent desire for dominance and power, on the one hand, and his reflective condemnation of those traits, on the other hand. This discord motivates the priest's self-deception.

Given the centrality of this notion, we might be tempted to identify it with Nietzschean disunity. That is, we might claim that an agent is disunified iff the agent's reflective judgments conflict with the agent's drives and affects, in the sense that they direct the agent toward conflicting ends. I think this is almost, but not quite, correct. There is an additional facet to Nietzsche's account of disunity: discord between reflective judgment and affect is not sufficient to engender disunity. It is only *certain forms* of discord between reflective judgment and affect that lead to disunity.

There are two pieces of textual evidence indicating that certain forms of discord between affect and reflective judgment are compatible with agential unity. First, Nietzsche often enjoins us to put ourselves in states of conflict between affect and evaluative judgment. He claims that conflicts between affects and evaluative judgments are an excellent opportunity for acquiring a deeper understanding of our values. He writes that the individual with "contradictory drives" has "a great method of acquiring knowledge: he feels many pros and cons. … The wisest man would be the one richest in contradictions" (*WP* 259; cf. *HH* 618, *GM* III.12, and *BGE* 284). Indeed, Nietzsche describes his genealogies as operating in precisely this way: they are meant to generate an emotional reaction to the value under investigation, thereby creating a discord between the affective response to the valued object and the reflective evaluation of the object.[50] For example, a genealogy of pity might proceed by fostering a negative emotional response to pity, in order to generate a conflict with our positive evaluation of pity. This state of tension leads us to reassess the value (of course, this reassessment need not culminate in the rejection of the value).

Now, the fact that Nietzsche enjoins us to enter these discordant states does not by itself indicate that conflict between affect and evaluative judgment is compatible with agential unity. After all, we could identify disunity with conflict between affect and evaluative judgment, and interpret Nietzsche as claiming that disunity is instrumentally valuable, as a means to some valued end. But recall that unity is meant to be an account of agential activity. It is hard to believe that the individual who engages in a Nietzschean genealogy, and discovers formerly hidden aspects of her psychic states and values, thereby becomes *less* active in the production of her actions.[51]

---

[50]Nietzsche writes, "The inquiry into the origin of our evaluations and tables of the good is in absolutely no way identical with a critique of them, as is so often believed: even though the insight into some *pudendo origo* certainly brings with it a feeling of diminution in the value of the thing that originated thus and prepares the way to a critical mood and attitude toward it" (*WP* 254).

[51]And indeed, Nietzsche seems to associate increasing self-knowledge with increasing activity. For an illuminating discussion of this point, see John Richardson, *Nietzsche's New Darwinism* (New York: Oxford University Press, 2004).

This brings us to the second piece of textual evidence. Some of the individuals whom Nietzsche praises—especially Goethe and Nietzsche himself—are praised for harboring and tolerating inconsistent values.

> Human beings have in their bodies the heritage of multiple origins, that is, opposite, and often not merely opposite, drives and value standards that fight each other and rarely permit each other any rest. Such human beings of late cultures and refracted lights will on the average be weaker human beings: their most profound desire is that the war they *are* should come to an end. … But when the opposition and war in such a nature have the effect of one more charm and incentive of life—and if, moreover, in addition to his powerful and irreconcilable drives, a real mastery and subtlety in waging war against oneself, in other words, self-control, self-outwitting, has been inherited or cultivated too—then those magical, incomprehensible, and unfathomable ones arise, those enigmatic men predestined for victory and seduction, whose most beautiful expression is found in Alcibiades and Caesar … and among artists perhaps Leonardo da Vinci. They appear in precisely the same ages when that weaker type with its desire for rest comes to the fore: both types belong together and owe their origin to the same cause. (*BGE* 200; cf. *BGE* 224)

Nietzsche claims that having conflicting drives and "value standards" often *leads to* disunity: it leads to "war," "opposition," and so on. However, certain individuals remain unified despite experiencing these forms of conflict: these individuals manage to achieve "self-control." In other words, disunity is a frequent result of having discordant affects and evaluative judgments, but disunity is not identical to having discordant affects and evaluative judgments.

I think these two pieces of textual evidence indicate that Nietzsche treats discord between affect and evaluative judgment as a condition that frequently results in, but is not identical to, disunity. Setting the textual evidence aside, though, there is also a compelling philosophical reason for distinguishing disunity from discord between affect and evaluative judgment. Recall that the concept of unity is designed to play a particular philosophical role in Nietzsche's account of agency: when an agent is unified, the agent plays an active role in producing her action; when she is disunified, her action is produced independently of her participation. Nietzsche quite rightly points out that having conflicting affects does not undermine an action's attributability to the agent.[52] After all, it is possible for an agent to harbor massively conflicting sets of affects, while remaining in control of her actions. Just so, the mere fact that an agent has conflicting affects and evaluative judgments does not indicate that the agent plays no role in the production of her action. For it is possible, albeit difficult, to harbor conflicting affects and evaluative judgments while remaining in control of one's actions.

Consider an example of an individual who experiences a conflict between his affects and his values. Earlier, we considered an alcoholic who ardently craves a drink, but judges that drinking would be disvaluable. Suppose this agent experiences frequent and powerful desires for alcohol. Yet he values sobriety and disvalues alcohol. His affects and values are discordant, for the desire for alcohol conflicts with the disvaluation of alcohol. This agent is certainly in an unfortunate state; it will be hard for him to control his actions, and he may end up succumbing to

---

[52]See especially *WP* 966 and *BGE* 212, quoted above.

the desire for alcohol. But suppose he does manage to avoid drinking. Rather than drinking the alcohol, he pours it down the sink. That action—pouring the alcohol down the sink—seems attributable to the agent's activity. The fact that the action is rendered difficult by opposing affects, which tempt the agent to refrain from pouring out the alcohol, does not seem to show that the agent is passive in the production of his action. If anything, the opposite is the case. Put simply, the discordant affects make it hard for the agent to control the action, but do not render the action not his own.

4.3 Nietzschean Unity

In the prior section, I pointed out that the priests exhibit a conflict between their reflective judgment and their affects. I argued that this feature does not, by itself, engender disunity: certain agents display the same type of conflict, but remain unified. So what differentiates the priests from these unified individuals?

Notice that the conflict between reflective judgment and affect takes a special form in the priests: it is *hidden*. I will argue that this fact is important.

Of course, there is nothing unusual about being ignorant of various aspects of one's actions. Nietzsche disparages the idea that agents are generally cognizant of their actions, calling it "the universal madness." For "the opposite is precisely the naked reality demonstrated daily and hourly from time immemorial! … Actions are *never* what they appear to be … all actions are essentially unknown" (*D* 116; cf. *D* 119).

However, the priests have a distinctive form of self-ignorance. I think we can capture their condition as follows:

> (Disunity) The agent currently approves of his *A*-ing. However, if he knew more about the drives and affects that figure in *A*'s etiology, he would *not* approve of his *A*-ing.

This notion of disunity is meant to characterize a peculiar psychological state, in which an agent approves of his action, but would reverse the attitude, were he to know more about the action's causal history. (The particular elements in the causal history that occupy Nietzsche's attention are the agent's drives and affects, though in principle there is no reason for restricting it to these types of motives.)

Notice that disunity, so defined, constitutes a kind of dissatisfaction with one's action. If an agent is disunified, he would cease to approve of his action, were he to know more about its etiology. So we might state the definition of disunity as follows: an agent is disunified in performing an action *A* if, were he to know more about the drives and affects that are causing him to *A*, he would *not affirm A*-ing.

With this in mind, let us offer a characterization of the contrary of disunity, namely *unity*:

> (Unity) The agent A's, and affirms his A-ing. Further knowledge of the drives and affects that figure in A's etiology would not undermine this affirmation of A-ing.

The account should be understood as applied to agents, holding all else constant except giving the agent further information about the drives and affects figuring in the etiology of the action under consideration. In particular, we do not want to consider cases in which the agent changes his values. For example, I would now disapprove of many of the actions that I performed, approvingly, as a child; but

this does not show that I was disunified in performing the actions, for at the time of action I may have wholeheartedly approved of them.[53]

Additionally, notice that the only factor that we are changing here is how much information the agent has about the *etiology* of the action. Agents sometimes disapprove of a past action, not because they learn more about the act's etiology, but because they learn more about the act's *consequences*. I did not realize that my innocent, offhand remark would hurt Sarah's feelings. Now, seeing her upset, I regret the remark, and wish that I had not made it. But this does not show that I was disunified in acting.

Notice, finally, that a disunified agent would not necessarily want to *act* differently. She might be dissatisfied with her actions, not because she disapproves of what she has done, but because she disapproves of her *motives* for doing what she has done. For example, suppose that Sally volunteers in a soup kitchen. She believes she is volunteering out of a desire to aid the impoverished beneficiaries. Yet a psychologically adept observer, well acquainted with Sally's character, would describe things differently: Sally takes satisfaction in feeling superior to the impoverished recipients, and her volunteering is in part motivated by this desire. Suppose Sally comes to realize that one of the desires motivating her action is the desire to feel superior. She finds this desire reprehensible, and she is no longer able to view her action of volunteering with approval. Thus, she is disunified. However, it would be inaccurate to say that she wants not to volunteer. Rather, she still wants to volunteer, but she wants to volunteer out of beneficent motives, rather than self-serving ones. So she is disunified, not because she wants to act differently, but because she wants to act out of different motives.

With these points in mind, we can see that disunity constitutes a form of psychic conflict. An agent acts and approves of his action. However, this approval is contingent upon ignorance of the drives and affects that are actually leading him to act. So there is a conflict between the agent's attitude toward the action as he takes it to be, and the agent's attitude toward the action as it is. Moreover, disunity implies that one has affects and drives that are moving one in ways that one would disavow. Thus, there is an interesting form of conflict between the agent's reflective and unreflective aspects at the time of action.

The notion of disunity can be used to characterize the psychic state of the priests. The priests are ignorant of aspects of their actions; they do not realize that they lust after the very states of affairs that they reflectively condemn. To adopt Nietzsche's pithy formulation, "The *motives* of this morality stand opposed to its *principle*" (*GS* 21). The last section argued that this discord between affect and reflective judgment is compatible with the agent's playing an active role in the production of his action. The struggling alcoholic who successfully overcomes his cravings for alcohol plays an active role in the production of his action, despite his experiences of discord. However, notice that the priest's discord is *hidden*. Rather than

---

[53]Nietzschean unity bears a resemblance to Harry Frankfurt's notion of wholeheartedness. Roughly, Frankfurtian wholeheartedness obtains when the agent bears a higher-order attitude of acceptance or approval toward his lower-order desires (see Frankfurt, *The Reasons of Love*). Nietzschean unity is more demanding: the approval in question must be stable in the face of further information about the action's etiology.

struggling with this discord in the open, or attempting to resolve the discord, the priest hides the discord from himself: the priest is "neither upright nor naïve nor honest and straightforward with himself" (*GM* I.10). This combination of discord and self-ignorance has an interesting result: it renders the priest disunified.

Consider an action that a priest performs. Say, he inveighs against the noble for being powerful. The priest's moral system gives power a negative evaluation. Thus, the priest will approve of his philippic. However, the priest has a "lust to rule"; he, too, seeks power. If he were to discover that this attitude were present in the etiology of his action, the priest would not be able to approve of his action. Thus, he is disunified.

In sum, then, I submit that Nietzsche's analysis of unity is as follows.
- Nietzschean Parts: the self is divided into drives and affects, on the one hand, and reflective thought, on the other hand.
- Nietzschean Relation: the self is unified when the reflective and unreflective parts of the agent are harmonious, in the following sense: the agent *A*'s, affirms his *A*-ing, and further knowledge of the drives and affects that figure in *A*'s etiology would not undermine this affirmation of *A*-ing.

When the agent manifests this form of unity, she produces a genuine action.

5. Is Nietzschean unity an adequate account of genuine agency?

We now have an analysis of what Nietzsche's conception of unity is. But we still need to ask whether the concept fulfills its philosophical role, namely, distinguishing genuine actions from their lesser relatives. To answer this question, it will help to examine a case in which an agent comes to recognize his own disunity. Consider a famous quotation from Augustine, in which Augustine describes his effort to extirpate his attraction to pride:

> Even when I reproach myself for it, the love of praise tempts me. There is temptation in the very process of self-reproach, for often, by priding himself on his contempt for vainglory, a man is guilty of even emptier pride.[54]

Augustine is endeavoring reflectively to revalue pride. Formerly, he valued pride; now, he regards it as evil. Yet the revaluation is no easy task: despite his efforts, he finds himself entangled in lingering affective traces of the old valuations. His desire to reproach himself for instances of pride is motivated, in part, by his persistent attraction to pride. He prides himself on reproaching himself for being prideful. So the renounced value lives on, in the sense that Augustine continues to have positive affective responses to pride.[55]

Here we have an individual who harbors a kind of psychic tension: the original evaluation of a state of affairs lingers on in a motivational tendency that opposes the motivational tendency springing from the new and different evaluation of the state of affairs. We can put the point this way: Augustine reflectively revalues pride,

---

[54]Augustine, *Confessions*, trans. R.S. Pine-Coffin (New York: Penguin Classics, 1961), Book X, 38.
[55]Reginster discusses the relevance of this passage for Nietzsche in "Nietzsche on *Ressentiment* and Valuation," 293.

but his affects lag behind; he continues to bear the affects that were associated with the valuation of pride (e.g. attraction to pride, delight in instances of pride, etc.), and these affects mingle with and influence the affects that are associated with disvaluation of pride (e.g. shame at displays of pride, aversion to pride, etc.). Thus, Augustine manages to take delight in his shame at displays of pride, thereby manifesting a mixture of the old and new affects.

Like the *Genealogy's* priest, Augustine experiences discord between his affects and his reflective judgments. Unlike the priest, Augustine strives to recognize and eliminate the manifestations of this discord. He often fails: above, he describes his discovery that his reflective assessments of his own motives are being influenced by the very motives that he is attempting to disavow. That is, he discovers that his reflective disavowal of pride is motivated by pride. So, in reflectively disavowing pride, he is disunified: once he learns more about the etiology of these disavowals, he comes to disapprove of them. For he can no longer see these disavowals of pride as expressive of his disvaluation of pride; on the contrary, he sees that these disavowals of pride are being motivated by his persistent attraction to pride.

With this discovery, Augustine is revealed to have been less than fully active in the production of his past actions. For his choices were being influenced in a fashion that he did not recognize, and which he would have found objectionable. This shows that Augustine's action was not wholly the product of his own activity.

Like the *Genealogy's* priest, Augustine is strongly attracted to that which he reflectively condemns. Moreover, Augustine does not initially notice this attraction, so he is initially disunified. And the problem that Augustine's disunity produces is that there is no clear answer to the question of where the agent stands. He has chosen to denounce pride, and this choice is consistent with his evaluative judgments. However, the denunciation is motivated by an attitude that Augustine cannot condone: pride itself. So Augustine's condemnation of pride no longer appears to be a manifestation of agential activity. His reflective thought seems to be buffeted about by inconsistent drives and affects, in a way that renders him a passive conduit for forces within.

As this example indicates, Nietzschean unity is at least a necessary condition for agential activity. Unity seems to offer a characterization of the conditions under which an agent can be said to be in control of her action.[56] The agent acts, approves of the act, and further knowledge of the action would not undermine this approval. To speak metaphorically, the agent's whole being is behind the action. I conclude that Nietzsche's distinction between unity and disunity adequately captures the distinction genuine action and mere behavior.

Moreover, notice that Nietzsche's account of unity gives us a way of distinguishing genuine cases of agency from their lesser relatives without committing ourselves to concepts of the will or the self that Nietzsche would regard as problematic. In order to manifest unity, an individual need not be pictured as standing apart from her drives and exerting a controlling influence on them. Nor must we conceive of the agent as having and exercising a will that is wholly independent from her

---

[56]Notice that an individual who is unified cannot act akratically. If an agent performs an action of which he disapproves, he is not unified.

drives. Rather, the account of unity merely requires that agents have conscious thoughts and engage in episodes of deliberation and choice, on the one hand, and also have drives and affects, on the other. The conscious thoughts and the capacity for choice are pervasively and inescapably influenced by drives. Yet they are distinct from drives, and therein arises the potential for disunity.[57]

6. CONCLUSION

Nietzsche's account of agential unity is designed to offer a way of distinguishing genuine manifestations of agency from acts that are produced merely by the agent's drives and affects. In other words, the account of unity distinguishes genuine actions from mere behaviors. The account of unity enables us to draw this distinction without relying on notions of the will or the self that Nietzsche would regard as problematic. Unity simply requires a certain kind of harmony between the agent's reflective and unreflective aspects at the time of action. In particular, an agent is unified—or, equivalently, the agent is active in the production of her action—when she approves of her action, and further knowledge of the action's etiology would not undermine this approval.

Moreover, Nietzschean unity is an analysis of what it is for the agent to determine her action through choice. Everyone will agree that *one* way of failing to determine one's action by choice is for one's action not to conform to one's choice: I decide not to drink at the party, but end up drinking after all. But Nietzsche draws our attention to another way that one can fail to determine one's action: one's *choice* can be determined by one's motives, in a way that one would disavow were one to recognize it. Thus, Nietzsche points out that in addition to examining the connection between choice and action, we must examine the connection between *the agent* and *choice*.[58]

---

[57]My interpretation does commit Nietzsche to two claims that are somewhat controversial: conscious thoughts must be both *causally efficacious* and *distinct from drives*. Some commentators read Nietzsche as rejecting one or both of these claims. For example, Brian Leiter argues that for Nietzsche, conscious thoughts are causally inert and perhaps not even distinguishable from drives; see "The Paradox of Fatalism and Self-Creation in Nietzsche," in Richardson and Leiter, *Nietzsche*, 281–321. I defend the idea that Nietzsche views conscious thoughts as causally efficacious and distinct from drives in my "Nietzsche's Theory of Mind," *European Journal of Philosophy* 13 (2005): 1–31, and "Nietzsche's Philosophical Psychology."

[58]For extremely helpful discussions of material in this essay, I owe great thanks to Lanier Anderson, Christine Korsgaard, Richard Moran, Bernard Reginster, John Richardson, and Danielle Slevens. I am also grateful to two referees for their illuminating comments.