How One Becomes What One Is:

The Case for a Nietzschean Conception of Character Development

by Mark Alfano, University of Oregon

**1 Introduction**

Gone are the heady days when Bernard Williams (1993) could get away with saying that “Nietzsche is not a source of philosophical theories” (p. 4). The last two decades have witnessed a flowering of research that aims to interpret, elucidate, and defend Nietzsche’s theories about science, the mind, and morality. This paper is one more blossom in that efflorescence. What I want to argue is that Nietzsche theorized three important and surprising moral psychological insights that have been born out by contemporary empirical psychology. The first Nietzschean insight is the *disunity of the self*. The second, connected, Nietzschean insight is the *primacy of affect*. This primacy is expressed by what I have called elsewhere (Alfano 2010, 2013a) the *tenacity of the intentional*, and what Nietzsche calls the *Socratic equation* (TI Socrates 4, 10; WP 2:432-3). The third major Nietzschean insight is the *social construction of character*, which presupposes a wild *diversity within the extensions of trait-terms* and the *dual direction of fit of character trait attributions*. This last point is somewhat in tension with the only other published defense of the empirical credentials of Nietzsche’s moral psychology (Knobe & Leiter 2007), so I will make a few remarks about the contrast between my view and theirs.

**2 Three Nietzschean insights**

*2.1 The disunity of the self*

When Nietzsche talks about the self, he often seems to contradict himself. For each passage in which he explicitly denies the existence of the self, there is another where he affirms it. Compare these two passages, for instance:

The subject (or, to use a more popular expression, the *soul*) has perhaps been believed in hitherto more firmly than anything else on earth because it makes possible to the majority of mortals, the weak and oppressed of every kind, the sublime self-deception that interprets weakness as freedom. (*GM* I:13)[[1]](#footnote-1)

At the bottom of us, really ‘deep down,’ there is […] some granite of spiritual *fatum* [….] Whenever a cardinal problem is at stake, there speaks an unchangeable ‘this is I’.(*BGE* 231)

Or contrast these two:

‘The subject’ is the fiction that many similar states in us are the effect of one substratum: but it is we who first created the ‘similarity’ of these states; our adjusting them and making them similar is the fact, not their similarity (– which ought rather to be denied –). (*WP* 269)

All instincts that do not discharge themselves outwardly *turn inward* – this is what I call the *internalization* of man: thus it was that man first developed what was later called his ‘soul’ [which one might at least be tempted to identify with the subject]. (*GM* II:16)

Or these:

Through thought the ego is posited; but hitherto one believed [...] that in ‘I think’ there was something of immediate certainty, and that this ‘I’ was the given cause of thought [….] However habitual and indispensable this fiction may have become by now – that in itself proves nothing against its imaginary origin. (*WP* 268)

Behind your thoughts and feelings, my brother, stands a mighty commander, an unknown sage – he is called Self. He lives in your body, he is your body. (*Z* I.4)

The tension between these pairs of quotations cannot be explained away by positing a change of heart or a failure to recognize inconsistency. Instead, the self whose existence Nietzsche denies differs so dramatically from the self he accepts that the inconsistency of his view is merely apparent. While he rejects the traditional notion of the self, he accepts a novel conception. Moreover, this novel conception has two levels: a descriptive level, at which a self is just a body, and a normative or honorific level at which a self is a body whose affects, desires, and drives are organized in a particular way. On this view, the extent to which one is a self is a matter of degree, not an all-or-nothing affair. The more wholehearted one is, the more one really is a self. For Nietzsche, wholeheartedness is a healthy, flourishing way to be.

Nietzsche opposes the conception of the self as a “pure will-less, painless, timeless subject of knowledge” (*GM* III.12), a conception he attributes with some justice to Kant. As Nietzsche understands the Kantian picture of the self, a person is essentially a unified, practically rational will that makes choices on the basis of known principles. Though a person may have all manner of desires, sentiments, and preferences, they are strictly irrelevant to who and what she is. The person herself stands behind these impulses, possessed of an autonomous capacity for rational choice.

Nietzsche denies this Kantian conception of the self (Janaway 2007, p. 220; Katsafanas 2011; Risse 2007; Richardson 1996, p. 46). But even if the self is not a unified, conscious, knowing, rational monad (GS 333, 354, 357), it may nevertheless exist as a teeming, diverse, non-conscious, desiring, irrational swarm. This interpretation of the Nietzschean self is supported by D 119, where he remarks that a person’s “*drives*” (note the plural) “constitute his being.”

Many contemporary interpretations of Nietzsche cluster around this view. According to Richardson, for instance, a person is not “a simple will for Nietzsche but an organized complex of numerous drives of various strengths” (1996, p. 46). Janaway says that “the great bulk of the real self” for Nietzsche is “composed of many competing drives” (2007, p. 18-19) and constituted by a “literal multiplicity of affects” (2008, p. 127). On his interpretation, the Nietzschean self is “a composite of hierarchically related drives” (2007, p. 263). For Risse (2007, p. 77), “What unites those drives and affects is their joint presence in one body with shared memories and cognitive abilities.”

This apparent consensus, however, masks a crucial question. Is the Nietzschean self merely a congeries of desires, drives, and affects that inhere in a given body, or is there something more to it? The correct response, I contend, is that both answers capture an aspect of Nietzsche’s full position. He thinks of the self as minimally the collection of desires, affects, and drives inherent in a given body, but he reserves a special, honorific status for selves that exhibit wholeheartedness.

This interpretation reconciles pronouncements like Z I.4 that the self just is the body, with more sophisticated claims elsewhere that the self is some kind of organization of one’s internal life. For instance, in BGE 6, Nietzsche equates the question who someone really is with the question “by what rank order the innermost drives of his nature are related to each other,” implying that the self is not a mere collection of drives but (at least in the ideal case) a structured system of drives.

This Nietzschean theory of the self further suggests that being a self is not a binary, all-or-nothing affair. Instead, one is more or less a self in the honorific sense in direct proportion to one’s wholeheartedness. If this is right, it helps explain the enigmatic subtitle of *Ecce Homo*: “How One Becomes What One Is.”[[2]](#footnote-2) One essentially is one’s desires, which are thrust upon one by fate, but one becomes what one is by endorsing those desires, by exercising what Nietzsche sometimes calls *amor fati*.

The main pieces of this picture are: 1) the human self is importantly embodied; 2) it is not unified by consciousness or the rational will; 3) it *may* become unified through harmony among the affects, desires, and drives; and 4) this sort of harmony contributes to (and may even constitute) well-being.[[3]](#footnote-3)

Empirical support for these theses is not far to seek. The implications of embodiment are one of the main areas of exploration of the psychology of the last few decades, which have focused on the embodied nature of affect and cognition. The philosophical debate has certainly not been settled, but I at least side with those who think that the disunity of the mind is better supported than the opposite view (Prinz 2013). But what I really want to focus on is the potential for harmony among conative states and dispositions, and the impact of this harmony on well-being. In this discussion, I draw primarily on the work of Daniel Kahneman, which is beautifully summarized in his recent capstone monograph, *Thinking, Fast and Slow* (2011).

The picture of human agency that has emerged from decades of research by Kahneman and other psychologists is deeply and fundamentally disunified. A crucial distinction in his work is between two “systems,” which he characterizes as follows: “*System 1* operates automatically and quickly, with little or no effort and no sense of conscious control” (p. 20). Its processes are often unconscious and affect-laden, and may be unamenable to introspection. Typically, we experience only the *deliverances* of System 1 as intuitions and impulses, which often seem inevitable, obvious, and right. In contrast, “*System 2* allocates attention to the effortful mental activities that demand it, including complex computations. The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration” (p. 21). Though System 2 is often assumed to be the true self, Kahneman contends that System 1 should wear that crown, and for several reasons. First, most of our cognitive and affective states and behavior are the products of System 1; the automatic, affective system runs the show. Second, System 2 is “lazy” (p. 39), unreliable, and often a mere bystander.

As Kahneman realizes, talk of two systems is prone to lead to reification, but he is resolute in his commitment to the distinction between these two forms of processing. Though System 2 may have some degree of unity, System 1 is better seen as a loose collection of – sometimes cooperating, sometimes competing – semi-autonomous processes. The affect heuristic, which I discuss in more detail below, is one such process, but Kahneman and his colleagues have investigated many others, especially the availability and representativeness heuristics.

Thus, there are at least two levels of disunity within this picture. First, in many instances, System 1 and System 2 will “disagree” with each other, and since the processes in System 1 tend to be unconscious, there will be no easy way to resolve their conflict. Second, System 1 itself is only a loose collection of processes, states, and dispositions, so different parts of System 1 will sometimes conflict with each other. Moreover, since System 1 runs the show, disputes between it and System 2 (and disputes within System 1) are registered affectively, as decrements in well-being. In particular, since System 2 processes tend to be conscious, they tend to be second-order (Rosenthal 2006), so when System 1 and System 2 disagree, this will typically be manifested as a lack of second-order identification or satisfaction with the deliverances of System 1 – exactly the failure of wholeheartedness that Nietzsche thinks is prone to undermine selfhood in the normative or honorific sense.[[4]](#footnote-4)

*2.2 The primacy of affect*

The primacy of affect refers to the idea that “hot” mental states, dispositions, and processes (affect, emotion, desire) – as opposed to “cold” states, dispositions, and processes (consciousness, cognition, and reasons) – are fundamental. This might seem like a restatement of the old Humean saw that reason is the slave of the passions, but Nietzsche puts his own distinctive spin on the idea, relying on the embodied nature of the self.[[5]](#footnote-5) Two examples of the primacy of affect help to clarify his position.

2.2.1 The tenacity of the intentional

One natural assumption about desire-like states is that satisfying them also extinguishes them. If you want to eat some guacamole, and then you do so, you will presumably no longer want to eat guacamole. What about when a desire-like state loses its object? For instance, someone madly in love is abandoned by his paramour. Or someone’s resented father dies. Or a committed Christian ceases to find the existence of God credible. It might seem that, in these cases as well, since the intentional object of the mental state is unavailable, the state itself would dissipate. Against this, Nietzsche subscribes to a view I have elsewhere dubbed the *tenacity of the intentional* (Alfano 2010, 2013a): When an intentional state with a subpropositional object loses its object, the affective component of the state often persists without a corresponding object, and that affect will generally be redeployed in a state with a distinct object. More precisely, the probability of redeployment is positively correlated with the strength of the affect.

Nietzsche makes claims that rely on such tenacitythroughout his writings. A few examples:

However much one thinks he has lost the habit of religion, he has not lost it to the degree that he would not enjoy encountering religious feelings and moods without any conceptual content. (*HH* 131)

[Buddhism and Christianity flourished because] they encountered a situation in which the will had become diseased, giving rise to a demand that had become utterly desperate for some ‘thou shalt.’ (*GS* 347)

Under peaceful conditions a warlike man sets upon himself. (*BGE* 76)

One *has* to repay good and ill – but why precisely to the one who has done us good or ill? (*BGE* 159)

[The need for obedience] is now innate in the average man, as a kind of *formal conscience* that commands: ‘thou shalt unconditionally do something, unconditionally not do something else,’ in short, ‘thou shalt.’ […] This need seeks to satisfy itself and to fill its form with some content. According to its strength, impatience, and tension, it seizes upon things as a rude appetite, rather indiscriminately, and accepts whatever is shouted into its ears by someone who issues commands. (*BGE* 199)

All instincts that do not discharge themselves outwardly *turn inward* [against] their possessors. (*GM* II:16)

I do not have the space to argue here for the correctness of my interpretation, but I can point to Paul Katsafanas’s (2012, 2013a, 2013b, 2014) work on drives as requiring ever new objects on which to vent themselves, and to Bernard Reginster’s (2006) work on will-to-power as a second-order desire to overcome resistance, and which therefore makes one insatiable because it spins off ever new first-order desires. In addition, the tenacity hypothesis bears striking similarities to Sigmund Freud’s conception of *displacement*, one of the so-called *primary processes* (along with condensation, reaction-formation, and splitting) that explain the psycho-dynamics of dreams, neuroses, and other apparently irrational mental phenomena (Freud 1899 / 1998).[[6]](#footnote-6) Since it’s obvious that Freud ripped off but also developed further many of Nietzsche’s ideas, this similarity serves as corroborating evidence. I plan to develop these ideas further in my next monograph, *Nietzsche’s Socio-Moral Psychology* (forthcoming b).

Some of the best contemporary evidence for something very like the tenacity of the intentional is due to Jon Haidt and his colleagues (Schnall, Haidt, Clore, & Jordan 2006), who have found over and over again that when an affective attitude is induced in someone, that person becomes much more likely to take up and to express the same type of attitude towards some other object. Disgust, for instance, is especially promiscuous. People in whom disgust has been elicited (by the surreptitious use of “fart spray” or by having them sit at a desk on which a used tissue has been discarded, etc.) are significantly more likely to feel disgusted by morally questionable behavior. What seems to happen is that they transfer their attitude of disgust from its initial target onto whatever is ready-to-hand – in these experiments, other people. The man who farted, the woman who left the used tissue on the desk, the ostensible target of the affect of disgust – these are unavailable, so the affect is instead discharged on whatever is available at the moment.

Another telling piece of evidence for the tenacity of the intentional is a disturbing study by Dan Ariely (2008), who found that male participants were significantly more likely to say that they would engage in a variety of disgusting, illegal, and immoral activities while they were masturbating to a pornographic movie than while they were in a “cold” state. Questions included:

* Can you imagine being attracted to a 12-year-old girl?
* Could it be fun to have sex with someone who is extremely fat?
* Could you enjoy sex with someone you hated?
* Can you imagine getting sexually excited by contact with an animal?
* Would you slip a woman a drug to increase the chance that she would have sex with you?

As with Haidt’s experiments, one plausible interpretation of this study is that sexual desire evolves in accordance with the tenacity hypothesis: since the object of desire for someone masturbating to a video is unavailable, his lust finds a new object – even if that object would otherwise be unattractive or repellant.[[7]](#footnote-7)

Suppose for the moment that Nietzsche does endorse the tenacity hypothesis, and that this hypothesis or something like it is supported by the empirical psychology I’ve just mentioned. What explains the truth of the tenacity hypothesis? The answer, I think, depends on the fact that humans are embodied cognitive-affective systems. To harbor a given disposition, to experience a given emotion, to be filled with a given affect – these are bodily states. They are states that take intentional objects, to be sure, but the bodily characteristics of someone who loves *x* and someone who loves *y* differ much less than the bodily characteristics of someone who loves *x* and someone who does not love anything. When the object of an intentional state becomes unavailable, a new one is often found to fill the gap. It’s for this reason, I think, that when someone is in an affective state with a given intentional object, but is unable to express that state towards the object, a new object – any object – is typically found.

2.2.2 The Socratic equation

Another manifestation of the primacy of affect is what Nietzsche sometimes calls the “Socratic equation of reason = virtue = happiness” (*TI* Socrates 4, 10; *WP* 2:432-3, 437). As with many things Nietzsche says, it’s not absolutely clear what he means, but I understand his denunciation of the Socratic equation as a rejection of the idea that all values point in the same direction. Nietzsche finds it preposterous to think that if something is intellectually valuable, then it is morally, prudentially, and aesthetically valuable; he finds it preposterous to think that if something is morally valuable, then it is prudentially and aesthetically valuable. And so on.

There are two, interconnected strands to this idea. The first is this: when someone has a positive (negative) affective attitude towards some property of an object, she will tend to experience positive (negative) affective attitudes towards other properties of that same object, even if these latter attitudes are unwarranted. If something is seen as good in some way (e.g., it is virtuous), it tends to be seen as good in all ways (e.g., it produces happiness); if something is seen as bad in some way (e.g., it is unreasonable), it tends to be seen as bad in all ways (e.g., it is vicious). As Keats put it, “’Beauty is truth, truth beauty,’ – that is all / Ye know on earth, and all ye need to know.” Holding constant the valence of properties, attribution of a positive (negative) property induces attribution of further positive (negative) properties.

The second strand of the Socratic equation is this: when two properties are (perceived to be) co-instantiated, and one has a clear valence, the other will tend to be seen as having the same valence. If *x* is *F* and *G*, and *F* is seen as good (bad), then *G* will tend to be seen as good (bad) as well.

Nietzsche recognizes, of course, that the Socratic equation is indefensible, but he thinks that it has infected humanity. For him, the ability to harbor ambivalence – even deep ambivalence – is a mark of nobility and objectivity (*BGE* 283-4, *GM* III:12).[[8]](#footnote-8) The “Tartuffery of morals” (*BGE* 24) demands that everything be simplified, that objects of desire also be objects of admiration, love, and respect, that objects of aversion also be objects of contempt, resentment, and scorn. Against this, Nietzsche holds that, for the most part, values are multiple, disjoint, even opposed.

Empirical support for the tendency to fall for the Socratic equation derives from work on the so-called affect heuristic. The affect heuristic was discovered by Paul Slovic and his colleagues, and refers to the tendency of an affective attitude towards one aspect of an object to spill over onto other aspects of the object. For instance, if people’s “feelings toward an activity,” such as using pesticides, “are favorable, they tend to judge the risks as low and the benefits as high; if their feelings toward the activity are unfavorable, they tend to make the opposite judgment – high risk and low benefit” (Finucane, Alhakami, Slovic, & Johnson 2000; see also Alhakami & Slovic 1994). In the real world, risk and reward are positively correlated; when people use the affect heuristic, they treat them as if they were negatively correlated. And it’s very hard to avoid using the affect heuristic because affective responses are often automatic, occurring prior to cognitive appraisal (Zajonc 1984).

A closely related phenomenon – the halo effect (Thorndike 1920) – occurs when someone who is seen as having one positive characteristic (e.g., beauty) is automatically seen as having other positive characteristics (e.g., intelligence). And the opposite can also occur (Nisbett & Wilson 1977): when someone who is seen as having one negative characteristic (e.g., emotional coldness), he is automatically seen as having other negative characteristics (e.g., irritating mannerisms). The same processes occur at the level of groups, rather than individuals. Social stereotypes have a kind of affective magnetism that trumps the beliefs people have about the members of groups. Indeed, emotions are better predictors of stereotyping discrimination than beliefs (Talaska, Fiske, & Chaiken 2008).

As with the tenacity of the intentional, the Socratic equation is deeply rooted our embodied nature. People use affect as information (Damasio 1994; Slovic, Finucane, Peters, & MacGregor 2002). In evaluating the properties of an object, person, or group, felt affective valence often trumps cognitive processes. Because emotions and affects are states of the body, they may differentially tend to trigger one another. For instance the embodied state of disgust is more similar to the embodied state of horror than it is to the embodied state of elevation. If you feel disgust towards *x*, it will be easier to transition to a state where you also feel horror at *x* than to a state where you feel awe at *x*. If you feel admiration for one element of the Socratic equation, you’ll more easily feel admiration for other elements of the Socratic equation.

*2.3 The social construction of character*

The third and final Nietzschean insight that I’d like to discuss is the social construction of character. This view of character development presupposes broad diversity within the extensions of trait terms, and depends on the dual direction-of-fit of character trait attributions – both surprising phenomena that Nietzsche diagnosed. In this, I am somewhat in disagreement with the only other published defense of the empirical credentials of Nietzsche’s moral psychology (Knobe & Leiter 2007), which I use here as a stalking horse.

According to Leiter, Nietzsche accepts the “doctrine of types,” according to which “Each person has a fixed psycho-physical constitution, which defines him as a particular *type* of person” (2002, p. 8). These type-facts are meant to be “*physiological* facts about the person, or facts about the person’s unconscious drives or affects” (Knobe & Leiter 2007), and they largely determine both what a person *can* do and what a person *should* do from the point of view of his own well-being. To support this interpretation, Knobe and Leiter cite Nietzsche’s claim that a

well-turned out human being […] *must* perform certain actions and shrinks instinctively from other actions; he carries the order, which he represents physiologically, into his relations with other human beings and things. (TI “Errors” 2)

Furthermore, they claim that, according to Nietzsche, type facts “radically circumscribe possible developmental trajectories,” but that they do not “determine *everything*.” Much of the remainder is determined by the “*moral* environment.” My view focuses on this other aspect of character development, which I think receives short shrift from Knobe and Leiter.[[9]](#footnote-9)

How does the moral environment partially determine a person’s character? How does one become what one is? The answer is manifold, of course, but I want to concentrate on what I take to be Nietzsche’s novel insight: one becomes what one is by becoming what one is called. There is a feedback loop connecting character trait attributions and traits themselves. The dispositions a person ends up with are the ones that she is taken to have – either by herself, if she exhibits an active, masterly, creative tendency, or by others, if she exhibits a reactive, slavish, mimetic tendency. Someone who thinks she is honest will tend to behave in honest ways; someone whom others consider dishonest will tend to behave in dishonest ways. This view can be found throughout Nietzsche’s writings, but especially in the following passages:

If someone obstinately and for a long time wants to *appear* something it is in the end hard for him to *be* anything else. (HH 51)

The reputation, name, and appearance […] of a thing […] nearly always becomes its essence and *effectively acts* as its essence. (GS 58)

[S]ince time immemorial, in all somehow dependent social strata the common man *was* only what he was *considered*: not at all used to positing values himself, he also attached no other value to himself than his masters attached to him. (*BGE* 261)

[T]he ordinary man still always *waits* for an opinion about himself and then instinctively submits to that – but by no means only a ‘good’ opinion; also a bad and unfair one. (*BGE* 261; see also GM I)

I have argued elsewhere that the best way to interpret such passages is to say that Nietzsche believes in the social construction of character (Alfano 2015; see also Hacking 2006); I’ve also argued that the empirical evidence suggests that character actually is socially constructed (Alfano 2013b, 2014). Here, I want to bring these two ideas together: despite the textual support for a weakened version of Knobe and Leiter’s doctrine of types, Nietzsche did in fact discover the social construction of character.

It may be hard to square the passages that support the doctrine of types with those that support the social construction of character, but here’s a try: Nietzsche thinks that many people have the *precise* character traits they do because they have been labeled with those traits. The idea is that type-facts limit the palette or menu of traits that someone could end up with, but do not uniquely determine how his character will develop. From that menu, social pressures select and shape the character that results.[[10]](#footnote-10)

This view presupposes a wild diversity in the extensions of trait terms.[[11]](#footnote-11) There is nearly as much diversity in the kinds of dispositions we are willing to call courage as there is between those dispositions and the kinds of dispositions we call by other names. Becoming what one is by becoming what one is called is possible because there are so many ways to be what one is called. This is an idea we see in various passages in Nietzsche, including these two:

Courage before the enemy is one thing: it does not prevent one from being a coward and indecisive scatterbrain. (GS 169)

Courage as cold valourousness and intrepidity, and courage as hotheaded, half-blind bravery – both are called by the same name!  Yet how different from one another are the *cold virtues* and the *hot*! (D 277)[[12]](#footnote-12)

Combined with this diversity is the idea of the dual direction-of-fit of character trait ascriptions. Labeling functions as a self-fulfilling prophecy, which gives it world-to-word direction-of-fit. But labeling also has word-to-world direction of fit: you can’t get away with calling just anyone courageous, even if courage is multifarious. In particular, it seems to me that *developmentally* trait ascriptions have world-to-word direction-of-fit, but *in the moment* they have word-to-world direction of fit. If this is right, the doctrine of types must be supplemented with the social construction of character. One’s type may, as Knobe and Leiter say, partially determine one’s character, but room must also be made for the influence of society.[[13]](#footnote-13)

This striking hypothesis receives empirical support from decades of social and personality psychology. In a seminal study, Miller, Brickman, & Bolen (1975) compared the effects of labeling with those of moral exhortation on the behavior of fifth graders. Participants in the exhortation group were asked repeatedly by the principal, the teachers, and the janitor to keep their classroom tidy. The labeling group, by contrast, heard congratulatory (false) announcements of their above-average tidiness over the course of eight days. On Day 1, the teacher praised them for being ‘ecology minded’ and mentioned that the janitor had commented that theirs was one of the cleanest classrooms in the school. On Day 2, the teacher noticed some litter on the floor but explained, “our class is clean and would not do that.” On Day 4, the principal visited the class and commended their orderliness; after he left, the students actually complained that the *teacher’s* desk was not as neat as theirs. On Day 8, the janitors washed the room and left a note thanking the students for making their job so easy. After a brief improvement in their behavior, the exhortation group settled back into its old routine, but the labeling group exhibited higher levels of tidiness over an extended period.

Other experiments have corroborated the tidiness study with other trait attributions. Jensen & Moore (1977), for instance, found that children labeled as charitable donated more than those who were subjected to moral suasion. Grusec, Kuczynski, Rushton, & Simutis (1978) announced to experimental participants that a questionairre they had completed indicated either that they were competitive or that they were cooperative, inducing congruent behavior in a subsequent game. Grusec & Redler (1980) found that ten-year-olds who helped once and were then labeled (“You know, you certainly are a nice person. I bet you’re someone who is helpful whenever possible.”) contributed 350% more in a subsequent trial than students whose actions were praised after helping (“You know, that was certainly a nice thing to do. It was good that you helped me with my work here today.”). More recently, in a longitudinal study, Srivastava et al. (2006) showed that optimists tend to have better romantic relationships – a relationship mediated by perceived support from romantic partners. In other words, if you systematically attribute supportiveness to your partner, you’ll be more likely to perceive support, which in turn will make your relationship more supportive: your partner *is* more supportive because *take him to be* more supportive. These are just a few examples, but the point should be clear: praise and exhortation are worse ways to elicit trait-congruent behavior than attribution. People become what they are by becoming what they are called.

Thus, while I agree with Knobe and Leiter both that Nietzsche accepts the doctrine of types and that there is strong empirical support for the doctrine of types, I think that they overlook Nietzsche’s insight into the social construction of character and the empirical support for the social construction of character. These two positions may seem to be in tension, but ultimately I think that both can be accommodated. Types are diverse. What counts as agreeableness, neuroticism, extraversion, and so on differs from case to case and person to person. Character traits develop through the interaction of types and social influence, and even they exhibit a great deal of diversity. In a recent book, Adams (2009, p. 182) argues that courage should be divided up into “modules” that include physical courage, social courage, financial courage, and vicarious courage (the courage not to be overprotective or paternalistic). This echoes Nietzsche’s own distinctions between different types of courage.

**3 Conclusion**

One might wonder how Nietzsche – a loner who had no access to the findings of contemporary experimental psychology and who seemed to delight in expressing stupid, reactionary views about race and gender (though he is occasionally surprisingly insightful about the latter, e.g., *D* 346) – could have achieved the insights I attribute to him in this paper. He attempts to answer this question himself in various works, especially *BGE* and *EH*. For instance, in *BGE* he goes out of his way to stress that his virtues are intellectual virtues, in particular, curiosity, (intellectual) courage, honesty (with himself), and cruelty (towards his own tender feelings). He seems to think that the insights to which he is privy are available to most people, but that most people are too intellectually lazy, too afraid of really seeing (and smelling) unpleasant truths about human nature, too inclined to self-deception, too merciful towards themselves to really appreciate these truths.[[14]](#footnote-14) And in *EH* he claims to be especially attuned to his own visceral and other bodily states, which in turn allows him to detect subtle cognitive and affective phenomena that others simply glide over. Whether these are accurate explanations of Nietzsche’s insights or mere post hoc confabulations is hard to say, but a question worthy of further pursuit.[[15]](#footnote-15)

**References**

Adams, R. M. (2009). *A Theory of Virtue*. Oxford UP.

Alhakami, A. & Slovic, P. (1994). A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk Analysis*, 14, 1085-1096.

Alfano, M. (forthcoming a). An enchanting abundance of types: Nietzsche’s modest unity of virtue thesis. *Journal of Value Inquiry*.

Alfano, M. (forthcoming b). *Nietzsche’s Socio-Moral Psychology*. Cambridge University Press.

Alfano, M. (2015). How one becomes what one is called: On the relation between traits and trait-terms in Nietzsche. *Journal of Nietzsche Studies*.

Alfano, M. (2014). What are the bearers of virtues. In H. Sarkissian & J. Wright (eds.), *Advances in Moral Psychology*. New York: Continuum.

Alfano, M. (2013a). Nietzsche, naturalism, and the tenacity of the intentional. *Journal of Nietzsche Studies*, 44:3, 457-464.

Alfano, M. (2013b). *Character as Moral Fiction*. Cambridge UP.

Alfano, M. (2013c). The most agreeable of all vices: Nietzsche as virtue epistemologist. *British Journal for the History of Philosophy*, 21(4): 767-90.

Alfano, M. (2010). The tenacity of the intentional prior to the *Genealogy*. *Journal of Nietzsche Studies*, 30, 123-140.

Damasio, A. R. (1994). *Descartes’ error: Emotion, reason, and the human brain*. New York: Avon.

Finucane, M., Alhakami, A., Slovic, P. & Johnson, S. (2000). The affective heuristic in judgments of risk and benefits. *Journal of Behavioral Decision Making*, 13, 1-17.

Freud, S. (1899 / 1998). *The Interpretation of Dreams*. Translated by James Strachey. New York: Harper Collins.

Grusec, J. & Redler, E. (1980). Attribution, reinforcement, and altruism: A developmental analysis. *Developmental Psychology*, 16:5, 525-534.

Grusec, J., Kuczynski, L., Rushton, J., & Simutis, Z. (1978). Modeling, direct instruction, and attributions: Effects on altruism. *Developmental Psychology*, 14, 51-57.

Hacking, I. (2006). Making up people. *London Review of Books*, 28:16, 23-26.

Janaway, C. (2007). *Beyond Selflessness: Reading Nietzsche’s* Genealogy*.* Oxford: Oxford University Press.

Jensen, A. & Moore, S. (1977). The effect of attribute statements on cooperativeness and competitiveness in school-age boys. *Child Development*, 48, 305-307.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Strauss, and Giroux.

Katsafanas, P. (2013a). Nietzsche’s philosophical psychology. In J. Richardson & K. Gemes (eds.), *Oxford Handbook of Nietzsche*. Oxford UP.

Katsafanas, P. (2013b). Philosophical psychology as a basis for ethics. *Journal of Nietzsche Studies*, 44, 297-314.

Katsafanas, P. (2014). Value, affect, and drive. In P. Kail & M. Dries (eds.), *Nietzsche on Mind and Nature*. Oxford UP.

Katsafanas, P. (2012). Nietzsche on agency and self-ignorance. *Journal of Nietzsche Studies*, 43:1, 5-17.

Katsafanas, P. (2011). The concept of unified agency in Nietzsche, Plato, and Schiller. *Journal of the History of Philosophy*, 49:1, 87-113.

Knobe, J. & Leiter, B. (2007). The case for Nietzschean moral psychology*.* In B. Leiter & N. Sinhababu (eds.), *Nietzsche and Morality*. Oxford UP.

Leiter, L. (2002). *Nietzsche on Morality*. London: Routledge.

Miller, R., Brickman, P., & Bolen. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology*, 31:3, 430-441.

Nisbett, R. & Wilson, T. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35:4, 250-256.

Prinz, J. (2013). Attention, atomism, and the disunity of consciousness. *Philosophy and Phenomenological Research*, 86:1, 215-222.

Reginster, B. (2006). *The Affirmation of Life: Nietzsche on Overcoming Nihilism*. Harvard UP.

Richardson, J. (forthcoming). Nietzsche, language, community. In J. Young (ed.), *Nietzsche and Community*. Cambridge University Press.

Richardson, J. (1996). *Nietzsche’s System*. Oxford: Oxford University Press.

Risse, M. (2007). Nietzschean “animal psychology” versus Kantian ethics, in Leiter & Sinhababu (eds.) *Nietzsche & Morality*, 57-82. Oxford: Oxford University Press.

Rosenthal, D. (2006). *Consciousness and Mind*. Oxford: Oxford UP.

Schnall, S., Haidt, J., Clore., G., & Jordan, A. (2006). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, 34:8, 1096-1109.

Shargel, D. (2013). Genuinely embodied emotions. Unpublished dissertation manuscript.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Grifﬁn, & D. Kahneman (eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York, NY: Cambridge University Press.

Srivastava, S., Richards, J., McGonigal, K., Butler, E., & Gross, J. (2006). Optimism in close relationships: How seeing things in a positive light makes them so. *Personality Processes and Individual Differences*, 91:1, 143-153.

Talaska, C. A., Fiske, S. T., & Chaiken, S. (2008). Legitimating racial discrimination: A metaanalysis of the racial attitude-behavior literature shows that emotions, not beliefs, best predict discrimination. *Social Justice Research: Social Power in Action*, 21, 263-296.

Thorndike, E. (1920). A constant error in psychological ratings. *Journal of Applied Psychology* 4:1, 25–29.

Williams, B. (1993). Nietzsche’s minimalist moral psychology. *European Journal of Philosophy*, 1:1, 4-14.

Zajonc, R. (1984). On the primacy of affect. *American Psychologist*, 39:2, 117-123.

1. I cite Nietzsche’s texts using the standard English-language acronyms. Translations are by Maudemarie Clark and Alan Swenson, R. J. Hollingdale, and/or Walter Kaufmann. [↑](#footnote-ref-1)
2. I return to the interpretation of this subtitle in the next section. [↑](#footnote-ref-2)
3. Note that this notion of honorific selfhood cross-cuts any notion of being a morally good person. Nietzsche does not think that someone who has a self in the honorific sense is bound (or even likely) to be morally good. It’s consistent with his view that the whole-hearted Nazi is a self in the honorific sense. Indeed, one might think that this helps to explain what’s so terrifying about such a person. Thanks to Iskra Fileva for asking me to point this out. [↑](#footnote-ref-3)
4. There are further potential ways for System 1 and System 2 to conflict. For instance, someone might be inclined to endorse, through System 2, conflicting intuitions or impulses delivered by System 1. Or someone might be inclined to endorse a System 1 intuition or impulse for some reasons and reject it for others. Thanks to Iskra Fileva for pointing out these possibilities. [↑](#footnote-ref-4)
5. Leunissen (this volume) discusses Aristotle on embodiment and character. It’s pretty clear, though, as she would surely admit, that Aristotle’s notion of different types of blood (hot/cold, wet/dry, etc.) is not empirically supportable. [↑](#footnote-ref-5)
6. I should note that, for Freud, there are two kinds of displacement: displacement of *content*, which corresponds to the tenacity hypothesis, and displacement of *cathexis*. ‘Cathexis’ is a technical term for Freud; it refers, in contemporary terminology, to the amount of affective arousal (positively- or negatively-valenced) attached to a particular mental construct (attitude, concept, representation, what have you). Freud conceives of mental constructs as a functional system of interconnected nodes, each of which is uniquely identified by an ordered triple: its content, its level of cathexis, and the set of (weighted) connections it bears to other nodes. Displacement of content occurs when a sub-propositional content is replaced by one of the contents to which it is connected; displacement of cathexis occurs when a node’s cathexis is transferred to an adjacent node. Both displacement of content and displacement of cathexis are understood to be potentially interative processes. Thanks to Iskra Fileva for inviting me to draw the connection between Nietzsche and Freud. [↑](#footnote-ref-6)
7. These are of course just a couple of examples. A host of others are collected in a terrific dissertation by Daniel Shargel (2013). [↑](#footnote-ref-7)
8. It might seem that this claim is inconsistent with my remarks about Nietzsche’s honorific notion of selfhood above, but it isn’t. Ambivalence comes in many varieties, only some of which undermine selfhood. Firm disapproval of one’s own first-order affective, emotional, or evaluative states undermines selfhood. First-order drives are often *potentially* in conflict, but can be harmonized in various ways. When I say that, for Nietzsche, harboring deep ambivalence is a mark of nobility, what I mean is that he thinks that, the more diverse one’s drives, the more of an achievement it is (and hence the more a mark of nobility it is) to bring them into harmony. [↑](#footnote-ref-8)
9. In Alfano (forthcoming a) I discuss at length the problems with Leiter’s strong version of the doctrine of types, arguing that while a much weaker version of it is textually supported, the strong version he advances is not a reasonable interpretation of Nietzsche. [↑](#footnote-ref-9)
10. There may be further ways in which social expectations mold and shape character. For instance, perhaps some social pressures are strong or persistent enough to result in character traits that were initially outside the “menu.” So-called “adaptive preferences” sometimes seem to fit this description. Thanks to Iskra Fileva for pointing this out. [↑](#footnote-ref-10)
11. This view shares some similarities with Christian Miller’s “Mixed Trait” account of character (this volume), but diverges in a number of ways. Miller and I agree that the traits that people tend to embody don’t easily fit into the categories that have traditionally been identified by philosophers as virtues and vices. We seem to diverge on a number of other points, though. First, Miller, like most philosophers who talk about virtue, thinks of virtue attributions solely as assertions, whereas I think of them as sometimes assertions but sometimes directives (when uttered in the second-person present tense), commissives (when uttered in the first-person present tense), or declarations. Second, from what I can tell, I’m more willing to call a trait a virtue or a vice than Miller is; he imposes rather stringent requirements on a trait’s being a virtue or vice. Third, Miller is more concerned with the distribution of traits in the normal adult population, whereas I am more concerned with how traits are acquired developmentally. Fourth, he prefers to understand traits as clusters of (dispositional) beliefs, desires, and emotions, whereas I think of them as more or less calibrated drives. [↑](#footnote-ref-11)
12. Richardson (forthcoming) discusses further evidence for this interpretation in a fascinating paper. [↑](#footnote-ref-12)
13. As I discuss in Alfano (2015), this view presupposes that many people do have a kind of global character: a *second-order* disposition, which Nietzsche calls “slavishness” to acquire or simulate whatever personality dispositions are attributed to one. [↑](#footnote-ref-13)
14. I explore the connections among these virtues, at least for Nietzsche and people like him, in Alfano (2013c). [↑](#footnote-ref-14)
15. I am grateful to Iskra Fileva for inviting me to contribute this chapter to her collection on virtues, as well as for her detailed and constructive criticism. I benefited in developing and formulating the views expressed in this paper from the feedback of Neil Sinhababu, Paul Katsafanas, Walter Sinnott-Armstrong, Rebecca Bamford, and Christine Daigle. The research presented here was supported by a grant from the Oregon Humanities Center. [↑](#footnote-ref-15)