

Katsafanas, Paul. *The Nietzschean Self: Moral Psychology, Agency and the Unconscious*. Oxford: Oxford University Press, 2016. Pp. 272. \$74.00 (cloth).

In his latest offering Paul Katsafanas, a leading Nietzsche scholar, provides a detailed account of Nietzsche's moral psychology. We are told that like Plato, Aristotle, and Hume, Nietzsche belongs to a tradition in which "a specification of human nature feeds into an articulation of an ethical theory" (2) and that "Nietzsche addresses a host of topics that are today grouped under the rubric of moral psychology," which Katsafanas glosses as "the study of human nature, especially the aspects of human nature that are relevant for assessing the justificatory status of normative claims and determining what happens when people act on the basis of those claims" (4).

Although the book reads as a sustained argument on this theme, it is fruitfully divided into two sections. The first five chapters provide the building blocks of Nietzsche's philosophical psychology; here we find discussions of the conscious/unconscious divide, drives, affects, and values. In the subsequent chapters (6–10), the theses from earlier chapters feed into an account of the Nietzschean self, as concerned with willing, unified agency, reflection, self-knowledge, and freedom. The project is to provide an account of Nietzsche's philosophical psychology which either leaves space for or actively complements the positive account of the Nietzschean self that is offered in the second half of the book. So, when Katsafanas asks, "given these accounts of human psychology [the ones we find in chaps. 1–5] . . . can we still speak of conscious willing, agency, reflection and the like?" (134), his answer is yes.

The book is ambitious in scope, so given space limitations my detailed criticisms will be of specific discussions from the first half (chaps. 1–5). However, the second half deserves mention as well, so I will briefly survey its contents. Chapter 6, "Willing without a Will," articulates a nuanced 'vector model of the will', explaining how Nietzsche retains a role for conscious thought, as one force among many in the production of action. The resulting view is one which Katsafanas argues is more plausible than the standard Kantian model which is committed to the 'Suspension' claim, namely, that "when an agent reflects on her motives for A-ing, she suspends the influence of the motives upon which she is reflecting" (144); a claim which Katsafanas persuasively argues Nietzsche rejects. Chapter 7 discusses Nietzschean unity, criticizing the stock interpretations which see unity as a formal relation among the drives (see main discussion for more on drives), and suggests a form of self-knowledge as an alternative criterion. These criticisms of formal approaches carry over into chapter 8, which stresses the connection between the way Nietzsche uses the 'self' as an honorific term and the nature of cultural situatedness,

arguing convincingly that individual greatness for Nietzsche cannot involve merely abandoning society and social norms but rather “what makes them great is that their revaluations have extensive influence on their cultures” (209). Finally, after a discussion of Nietzschean freedom in chapter 9, the volume closes with a comparison between the Nietzschean moral psychology developed over the course of the book and its competitors, as found in Kant, Hume, and Aristotle. Here Nietzschean Moral Psychology is argued to be preferable in various respects, both empirically and philosophically, although the most significant point is the enlarged role that Katsafanas argues Nietzsche gives to unconscious processes in motivation (279). If this is the central contrasting and positive contribution of Nietzschean moral psychology, then it is crucial that the Nietzschean ‘specification of human nature’, from the first half of the book, which “feeds into an articulation of an ethical theory” (2) as found in the second, is both exegetically and independently plausible. In what follows I provide some criticisms and queries about this philosophical psychology.

Katsafanas begins the first half of the book by discussing the relation between unconscious and conscious mental states for Nietzsche. In contrast to Descartes, Locke, Kant, and Hegel, who “bestow an overriding importance on consciousness” (14), Nietzsche claims that “consciousness is characteristic of only a small portion of our mental economies; it issues from or expresses something deeper [namely, the unconscious]; it is a potentially dangerous regression, it is superficial; and it falsifies” (16). Katsafanas documents Nietzsche’s commitment to these claims through textual evidence and then motivates them by attributing to him a novel way of carving up our mental economy, namely, that “he understands unconscious states as those with nonconceptual content, and conscious states as those with conceptual content” (16). As we will see, this claim is problematic.

After distinguishing the way in which Nietzsche rejects the idea of consciousness as a substantial faculty, while still allowing for consciousness as a property of mental states (21–23), Katsafanas begins his defense of the above thesis by reference to *Gay Science* 354. There Nietzsche writes: “We could think, feel, will, remember, and also ‘act’ in every sense of the term, and yet none of all this would have to ‘enter our consciousness’ (as one says figuratively). All of life would be possible without, as it were, seeing itself in a mirror; and still today, the predominant part of our lives actually unfolds without this mirroring—of course including our thinking, willing, and feeling lives” (Friedrich Nietzsche, *The Gay Science* [Cambridge: Cambridge University Press, 2001], 354). Katsafanas interprets the reference to ‘mirroring’ as suggesting that Nietzsche “associates consciousness with some form of introspective awareness” (24). The link to introspection is puzzling though, since the idea does not seem to be that either our episodic

conscious states or any putative unconscious states can be distinguished from separate acts of reflective introspection on them. Rather, Nietzsche seems to be animated by the idea later criticized by Gilbert Ryle, namely, that “the contents of the mental world were thought of as being self-luminous or refulgent” (Gilbert Ryle, *The Concept of Mind* [London: Penguin, 1990], 152–53). In other words, that in ‘thinking, willing, and feeling’ I am not just aware of whatever object it is I am f-ing about, but I am synchronously aware of myself being aware of that object. If absence of this kind of immediate self-awareness is what distinguishes the unconscious from consciousness, then we can already see that Nietzsche has an idiosyncratic way of carving up our mental economy.

Katsafanas then discusses later sections of *Gay Science* 354, in which Nietzsche links consciousness to communication and the development of language. We are told that “the central claim in this passage is that conscious thinking, and only conscious thinking, occurs in words” (25), which after some supporting passages associating concepts with words (although see below) is glossed as the claim that “conscious thinking is conceptually articulated” (25). The main thesis is then put forward as following from this: “Further, since Nietzsche claims that conscious states, and only conscious states, have conceptual content, it follows that unconscious mental states do not have conceptual content; unconscious states must have a type of nonconceptual content. Accordingly, the distinction between conscious and unconscious states is coextensive with the distinction between mental states with conceptual content and those with nonconceptual content” (26).

However, there is a non sequitur here. Granting that *Gay Science* 354 can be read as claiming that conscious thinking, and only conscious thinking, is conceptually articulated through linguistic symbols, this only establishes a point about thinking and also that if there is unconscious thinking then it is nonconceptual (this is an unhappy consequence, since there are good reasons for thinking that the notion of a nonconceptual belief is highly problematic). What it does not establish is that consciousness and conceptualization are coextensive. Even if Nietzsche could be read as claiming that all conscious states involve thought, and that all thought is conceptually structured via linguistic symbols, then it is possible that we have mixed conscious states which involve both a thought which is conceptually structured in this way and some other component that is not. However, the motivation behind the non sequitur is obvious enough, since the claim that conscious thinking is conceptually structured via linguistic symbols might result from its having propositional content (and so be of limited interest). The question that has animated philosophers of the mind is not whether thinking is conceptual in some way but rather whether perceptual experiences, which seem different from beliefs and judgments, have, either partly or exclusively,

conceptual content.

Katsafanas, to be fair, does discuss perceptual experiences. According to him Nietzsche's view involves "two different forms of perception: one involving non- conceptual contents, the other involving conceptual contents. The former are un- conscious, the latter conscious" (31). Yet, Katsafanas claims that Nietzsche does not think of such unconscious perceptions as unstructured 'mere data', but rather they have "definite, structured content—just not conceptually structured or articulated content" (31). Again, this way of framing Nietzsche's view sounds odd when compared to debates about whether it is possible to have personal-level representational states with partly or exclusively nonconceptual content, which are rarely framed in terms of the conscious/unconscious distinction.

Moreover, such remarks highlight problems with the earlier identification of conceptual content with linguistic content, since what is typically motivating the conceptualist (someone who claims that perceptual experience is exclusively conceptual) is finding a way of thinking about perceptual experience as conceptual while making the concession that at least some of it is too finely grained to be structured in terms of abstract general concepts (i.e., via linguistic symbols). That a perceptual experience is not conceptually structured in terms of linguistic symbols does not mean that it is nonconceptual; it just means we need a more nuanced theory of what it means for a perceptual experience, as contrasted with a belief or judgment, to be conceptually structured. Nietzsche will be blind to this point if, as Katsafanas has it, conceptual content is (wrongly) equated with linguistic articulation. In fact, Nietzsche seems alive to this distinction in one of the passages Katsafanas quotes, since after saying that "words are acoustic signs for concepts" Nietzsche tells us that "concepts, though, are more or less determinate pictorial signs for sensations that occur together and recur frequently, for groups of sensations" (Friedrich Nietzsche, *Beyond Good and Evil* [Cambridge: Cambridge University Press, 2002], 268), therefore distinguishing between concepts and the words that are used to articulate them.

This mistake generates further problems when we are told that for Nietzsche "unconscious perceptions have nonconceptual content, in the sense that they represent their objects in a definite way, but do not represent them as instantiating concepts" (32). One worry is that an 'unconscious perception', so described, seems close to a conscious perception involving an experience of an object expressible by way of a demonstrative concept, where the experience is structured in terms of a 'definite way' or even what Katsafanas later calls a 'discriminatory ability' (34). Indeed, when Katsafanas claims that for Nietzsche "the movement from an unconscious state to a conscious state is the process of conceptualization" (31)—keeping in mind that conceptualization is being thought of as linguistic articulation—we might seem to have a

movement from a personal-level conscious state which presents an object in a 'definitive way', what some would be happy to call the exercise of a conceptual 'discriminatory ability' in the experience, to an explicit belief or judgment formed on its basis. A sophisticated conceptualist will insist that all of this should, on pain of making rational transitions in the epistemology of perception mysterious, be described as both conscious and conceptual, thus making the stipulation of the original perceptual state as both unconscious and nonconceptual a confusing mislabeling.

At the end of the chapter, Katsafanas addresses the worry that the way he has carved up Nietzsche's philosophy of mind will seem odd, using terms in ways that will strike contemporary readers as idiosyncratic. Katsafanas appeals to the disparate senses of the term 'unconscious' in play, suggesting that if Nietzsche is, at least at times, using the conscious/unconscious distinction to "mark a difference not in degrees of awareness, but in types of mental state" (44), then the idiosyncrasy might be alleviated. Nevertheless, the confusion turns on carving up the mental realm according to the unconscious/conscious distinction first and then affixing a restricted version of the conceptual versus nonconceptual distinction onto this. Moreover, a theory of mind in which anything other than introspectively accessible, linguistically articulated mental states turns out to be unconscious, and so nonconceptual, will seem to have drawn the conscious/unconscious distinction in the wrong place, and in doing so it will occlude one of the interesting questions in these debates, namely, whether there are personal-level conscious states with nonconceptual content. It seems that on Katsafanas's reading Nietzsche is terminologically crippled from providing anything other than a flat-footed negative answer to this question. What parts of the view are Nietzsche's philosophy of mind is a more complex matter, although some of the interpretative leaps from the structure of thought to that of experience, and the equating of linguistic and conceptual content, are not supported by the passages cited.

One other central discussion in Katsafanas's book is Nietzsche's philosophy of value and specifically the notion of a drive, which is central insofar as Nietzsche "frequently claims that drives include evaluations and that drives explain reflective judgements" (109). Chapter 4 persuasively argues against homunculi readings of drives (as found in Maudemarie Clark and David Dudrick, *The Soul of Nietzsche's "Beyond Good and Evil"* [Cambridge: Cambridge University Press, 2012]) and drives as merely dispositions. Katsafanas's own preferred reading, that a "Nietzschean drive is a disposition that induces an affective orientation" (106), which makes better sense of Nietzsche's evaluative and agential language when talking of drives, still needs further clarification, since talk of dispositions 'inducing' anything sounds strange. A dispositional property like being soluble does not 'induce' its predicated object to 'do' anything; rather, dispositions are realized in occurrent states and are therefore expressible in law-like hypothetical

statements. Perhaps talk of Nietzschean drives as multitrack affective dispositions (akin to temperaments) would be another option and avoid the implication that drives are mysterious paramechanical inner forces that 'induce' things.

Chapter 5 uses the above understanding of drives to make sense of Nietzsche's philosophy of value. Here Katsafanas suggests that an "agent values X iff (i) the agent has a drive-induced affective orientation toward X and (ii) the agent does not disapprove of this affective orientation" (108). Katsafanas is right to stress the distinctiveness of Nietzsche's position, which eschews understanding values as akin to what we reflectively endorse and self-consciously act on, noting that he "more often focuses upon valuations inherent in pre-reflective psychological phenomena: our drives" (109). However, there is another prereflective psychological phenomena which makes as strong a claim to be Nietzsche's focus, namely, what he calls 'affects'.

After persuasively showing why the interpretation of values as the "aims of drives" (114) encounters problems, Katsafanas considers versions of the view that Nietzschean evaluations are based in affects; here I will examine his criticism of Peter Poellner's view that values are coconstituted by affects (Peter Poellner, "Affect, Value and Objectivity," in *Nietzsche and Morality*, ed. Brian Leiter and Neil Sinhababu [Oxford: Oxford University Press, 2007], 227–62). Roughly, Poellner's 'perceptual emotions' view is as follows: Nietzsche's affects sometimes possess a distinctive kind of affective intentionality, where the evaluative content of such affects, as disclosed through a felt phenomenology of attraction or repulsion, constitutively requires a reference to evaluative features as characterizing how that object itself is, as a "registering of the object's nature" (*ibid.*, 232). This view aims to capture a distinctive feature of the phenomenology of affective experiences (as conscious emotions), namely, that we sometimes experience our emotions as not merely contingently caused by their object (the Humean picture) but an appropriate or merited uptake of what are taken to be the object's evaluative features. Katsafanas's formulations of this view are clumsy, insofar as he says that it involves "the thought that this attraction is merited by X's features" (116). This makes the view overly cognitivist, whereas the normative component is supposed to be a constitutive, and synchronous, part of the affective experience.

His main criticism of the view is as follows:

It seems implausible to identify values with affects-regarded-as-justified for a simple reason: I can have passing, fleeting affects that I regarded as warranted but that do not count as values. Suppose that once in my life, I have an attraction to skydiving. I have affective experience of approval towards the thought of skydiving, and I regard this affect as justified

by the belief that the experience will be exhilarating. Nonetheless, the fancy passes, and I never give it a second thought. It would hardly be right to say that I value skydiving, even at the moment when I have that affective response. A “value” held only for a moment is no value at all. (117)

Katsafanas’s choice of an example is odd, since the affective experience of attraction he describes is indirect, insofar as it mediated by a thought of skydiving (rather than an experience of skydiving) and then justified in the light of a belief. A fairer example would be as follows: “Suppose that once in my life, I go sky-diving. I have affective experience of approval (‘attraction’) during the skydiving, and I experience this affect as warranted insofar as the skydiving experience seems genuinely exhilarating and so of positive value. Nonetheless, I do not go skydiving again.” The original conclusion should also be rephrased: “It would hardly be right to say that I experience the value of skydiving, even at the moment when I have that affective response. A ‘value’ experienced only for a moment is no value at all.” Yet, when rephrased the objection falls flat.

Underlying the above is a confusion about different uses of the term value. We need to distinguish between (1) an affective experience of value and (2) a value as something that exerts a structuring effect on an individual’s life. Some of the things that I affectively experience as of value might come to be ‘life-orientating’ values, but not all will. The revised skydiving example highlights this: I had an experience of the value of skydiving, yet it was merely a ‘fancy’, something I did not come to take all that seriously. This distinction is not foreign to Nietzsche either, since his appeal to prereflective affective experiences of value highlights the disparity that can emerge between (1) our experience of value and (2) our firmly rooted ‘life-orientating’ values. While there is much of interest in the chapters on Drives and Values (chaps. 4 and 5), the strong connection that Katsafanas argues for between values and drives simplifies the role that affects play for Nietzsche. One does not get the impression from reading his texts that ‘affects’ are mere drive conduits, always to be explained in terms of a ‘deeper’ psychology.

In closing, a number of general features are worth highlighting. Throughout the volume the writing is clear and intelligent, and Katsafanas is impressive in drawing links between theses advanced earlier in the book and later discussions. Although, if one has substantive disagreements about some of those theses, for example, the equating of conceptual content with linguistic articulation, then the dependence of later claims on those earlier theses raises problems. For example, Katsafanas’s discussion of the causal role of interpretations on our motives in chapter 6 has to say that interpretations are always conscious, since interpretations are said to have conceptual content and a “conscious state . . . is a conceptualized state” (155). Katsafanas

should also be praised for his sympathetic reconstruction of the views of other Nietzsche scholars, and, in all but the one case noted, these accurately reflect those positions and draw out what is appealing in them. Moreover, the volume includes several carefully presented discussions of historical philosophers, most notably of Schopenhauer on nonconceptual content (27–30) and drives (92–124) and Kant and Hegel on freedom (235–43). It is also a testament to the impressive scope of the book that it covers many of the important, and still contested, debates in Nietzsche studies. Even if one does not agree with the book's significant claims, careful attention has to be paid to Katsafanas's arguments to see why. Therefore, despite specific misgivings, this volume deserves to be read seriously by all those with an interest in Nietzsche and moral psychology.

Jonathan Mitchell

University of Warwick