The Nietzschean precedent for anti-reflective, dialogical agency

Mark Alfano, Delft University of Technology & Australian Catholic University

John Doris and Friedrich Nietzsche have a lot in common. In addition to being provocative and humorous writers in their native idioms, they share a conception of human agency. It can be tiresome to point out the priority claims of an earlier philosopher, so I should say at the outset that I do so not to smugly insist that my guy got there first but to showcase a closely-allied perspective that may shed additional light and offer glimpses around blind corners. In particular, I argue that Nietzsche anticipates both the anti-reflective and the dialogical aspects of Doris's theory of agency.

Doris's primary target is *reflectivism*, according to which the exercise of human agency consists in judgment and behavior ordered by (accurate enough) reflection about what to think and do. As Paul Katsafanas (2013, 2016) and I (Alfano 2010, 2013b, 2016c) have argued, Nietzsche's doctrine of will to power presupposes that human agency does not require reflection. Instead, agency emerges from the interaction of drives, affects, and emotions that are sometimes in harmony and sometimes in conflict (*Daybreak* 119; *Gay Science* 333, 354, 357; *Beyond Good and Evil* 6, 12, 200, 224; *Genealogy of Morals* I:13, II:16).[1] While affects and emotions receive much attention in contemporary dual-process psychology, drives – which Nietzsche construes as standing motivational dispositions to engage in particular action-types – have largely been ignored. The Nietzschean perspective thus expands the class of mental processes that can lead to incongruent parallel processing, though empirical research is of course needed to determine the role played by drives in our mental economies.

Like Doris, Nietzsche is concerned that incongruent parallel processing may undermine agency. Doris focuses on cases in which someone would not endorse the causes of her own behavior because those causes are what I have dubbed "non-reasons" (Alfano 2013a, pp. 43-45). While Nietzsche does not rule out such arational influences on behavior, his concern is directed to cases in which someone would outright reject an accurate description of the causes of her behavior (*Genealogy of Morals* I:10-11, II:11, III:15). Nietzsche's solution to the problem of incongruent parallel processing also resembles Doris's. According to Doris, someone's behavior constitutes an exercise of agency just in case it expresses a subset of the agent's values: namely, those values that are sufficiently longstanding, strong, and accepted by the agent as justificatory. The key move in this account is to allow for agency despite self-ignorance. The agent needn't know that she is willing to assign a justificatory role to the values she expresses in action, nor need she realize that her exercise of agency expresses one or more of her values. Similarly, for Nietzsche (*Daybreak* 109; *Beyond Good and Evil* 3), someone's behavior constitutes an exercise of agency just in case it expresses a subset of the agent's drives, affects, and emotions: namely, those that, were the agent to learn about them, would not lead her to disapprove of her own action (Katsafanas 2013, p. 138; 2016, ch. 7). Like Doris, then, Nietzsche allows for agency despite substantial self-ignorance. The agent needn't know that she would endorse her own action after learning more about its etiology, nor need she know which drives, affects, and emotions led to her action in the first place.

---

[1] I refer to Nietzsche's texts using the canonical section numbering rather than page numbers.

Lowering the bar of accurate reflection in this way enables Doris and Nietzsche to countenance and even make use of a common human foible: the tendency to *post hoc* confabulation. When we reflect on why we did what we did or what sorts of people we are, we often enough tell ourselves flattering stories that enhance our own rationality, moral rectitude, or agency – turning every "it was" into a "thus I willed it" (*Thus Spoke Zarathustra* II.Redemption; see also *Gay Science* 277). Without disputing the inaccuracy of such confabulation, Doris argues that it can lend someone the courage to move confidently into the future, to undertake ambitious projects, and to make demanding commitments. Tactically deployed fictions about ourselves can become facts. Indeed, in Alfano (2013a ch. 4; see also Alfano 2016a, ch. 4 and 2016b) I argue that they can lead to factitious traits of character, where someone becomes, for example, honest because she first falsely attributes honesty to herself.

To understand how people manage this trick, we must turn to the dialogical aspect of Doris's and Nietzsche's conception of agency. Doris's secondary target is individualism, according to which optimal human decision-making is exemplified by individual thinkers. He argues instead for collaborativism, according to which human decision-making and agency are socially embedded and perhaps even extended through dialogical processes. The factitiously honest person is (typically) not honest in splendid isolation. Instead, she engages in social interactions that "hold" her in a network of narratives that represent what she and those close to her consider her most important actions, passions, traits, roles, relationships, and values (cf. Lindemann 2014).

There are several ways in which such narratives help to build and stabilize human persons. They can change someone's self-concept by representing them as embodying values or drives that they previously did not self-attribute. They can reassure someone of the accuracy of their self-concept by providing social proof. They can lend someone a communal identity by representing them as one of "us." And they can help to "shape and crystallize" someone's traits and values by "making more determinate tendencies and impulses […] that are in some degree inchoate" (Wong 2006, p. 136). If this is right, then talking about our mental lives resembles not so much describing the weather as negotiating a cooperative agreement.

As I argue in Alfano (2015, 2016c), Nietzsche identifies two main channels through which such dialogical processes flow (*Human, All-Too-Human* 51; *Daybreak* 105, 201, 248; Gay *Science* 21, 40, 58; *Beyond Good and Evil* 42, 44, 261; *Genealogy of Morals* I:2, I:6). On the one hand, sometimes a person announces what she is (i.e., what her values, motives, concerns, or drives are), and that announcement meets with social acceptance; on the other hand, sometimes someone else announces what the person is, and she accepts the attribution. Such bid-and-accept patterns can be iterated. X could describe herself has embodying value V, to which Y responds by pointing to evidence (e.g., in her past behavior) that she actually embodies value V*, to which X responds by pointing to evidence that she actually embodies V†, and so on. Moreover, the negotiation needn't be so explicit. X could instead tell a story that represents herself as embodying V, to which Y responds by asking a question that presupposes that she embodies V*, to which X responds by telling another story that represents herself as embodying V†. The kind of person or self that emerges from such feedback loops is reflected or echoed rather than reflective and transparent.

If this is right, then much of human agency is constitutively social. Moreover, it suggests that a novel class of dispositions – namely, the dispositions associated with being a good echo –

must be recognized and theorized by philosophers and social scientists.[2] A virtuous echoer may not be entirely accurate and comprehensive. Instead, I contend, a virtuous echoer filters and perhaps even distorts to some extent; in this way, the echoer modulates existing dispositions without inventing them whole cloth. By contrast, a vicious echo may negotiate their partner into accepting an unflattering self-description or even an incoherent self-description, undermining their agency.

I conclude with a speculation about the future of dialogical agency in the wake of self-tracking technologies and predictive analytics (cf. Selke 2016). If it's true that someone's agency and self are constructed by the stories she tells and is told about herself, and that these stories need be neither wholly accurate nor unfiltered, then such technologies and analytics may curtail one's capacity to – in Nietzsche's phrase – become what one is. Self-tracking is liable to make it more difficult to enjoy self-enhancing illusions. If I tell a story about how I once did this or saw that, it may become possible for me or someone else to verify or falsify my narrative. In addition, predictive analytics is liable to make it more difficult to have inflated confidence in one's capabilities. If positive illusions about my own potential enable me to undertake ambitious plans and commitments, then well-evidenced predictions that I am likely to fail may keep from trying in the first place. There may be a steep prudential price to be paid for the epistemic benefits of these innovations.

References:

Alfano, M. (2010). The tenacity of the intentional prior to the *Genealogy*. *Journal of Nietzsche Studies*, 40: 123-40.

Alfano, M. (2013a). *Character as Moral Fiction*. Cambridge University Press.

Alfano, M. (2013b). Nietzsche, naturalism, and the tenacity of the intentional. *Journal of Nietzsche Studies*, 44(3): 457-64.

Alfano, M. (2015). How one becomes what one is called: On the relation between traits and trait-terms in Nietzsche. *Journal of Nietzsche Studies*, 46(1): 261-9.

Alfano, M. (2016a). *Moral Psychology: An Introduction*. Polity.

Alfano, M. (2016b). Friendship and the structure of trust. In A. Masala & J. Webber (eds.), *From Personality to Virtue*. Oxford University Press.

Alfano, M. (2016c). How one becomes what one is: The case for a Nietzschean conception of character development. In I. Fileva (ed.), *Perspectives on Character*. Oxford University Press.

Dixon, R. & Gould, O. (1996). Adults telling and retelling stories collaboratively. In P. Baltes & U. Staudinger (eds.), *Interactive Minds: Life-Span Perspectives on the Social Foundation of Cognition*. Cambridge University Press.

Katsafanas, P. (2013). *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford University Press.

Katsafanas, P. (2016). *The Nietzschean Self: Moral Psychology, Agency, and the Unconscious*. Oxford University Press.

Lindemann, H. (2014). *Holding and Letting Go: The Social Practice of Personal Identities*. Oxford University Press.

---

[2] Arguably, the literature on transactive memories (e.g., Dixon & Gould 1996) already does this to some extent. Thanks to Alessandra Tanesini for pointing this out.

Selke, S. (2016). *Lifelogging: Digital Self-tracking and Lifelogging: Between Disruptive Technology and Cultural Transformation*. Springer.

Nietzsche, F. (1878/1996). *Human All-Too-Human*. Translated by R. J. Hollingdale. Cambridge University Press.

Nietzsche, F. (1881/1997). *Daybreak: Thoughts on the Prejudices of Morality*. Translated by R. J. Hollingdale. Edited by M. Clark & B. Leiter. Cambridge University Press.

Nietzsche, F. (1882/2001). *The Gay Science*. Translated by J. Nauckhoff. Edited by B. Williams. Cambridge University Press.

Nietzsche, F. (1883/2006) *Thus Spoke Zarathustra*. Translated by A. Del Caro. Edited by A. Del Caro & R. Pippin. Cambridge University Press.

Nietzsche, F. (1886/2001). *Beyond Good and Evil*. Translated by J. Norman. Edited by R.-P. Horstmann. Cambridge University Press.

Nietzsche, F. (1887/2006). *Genealogy of Morals*. Translated by C. Diethe. Edited by K. Ansell-Pearson. Cambridge University Press.

Wong, D. (2006). *Natural Moralities*. Oxford University Press.