

Nina Strohminger, Joshua Knobe, and George Newman's compelling and thought-provoking piece on the "True Self" presents an original theoretical intervention into the vast body of literature on the self, which spans across several different disciplinary and epistemological traditions. Aside from making an important contribution to the existing theories of the self, the true self concept also opens up an avenue for raising a number of interesting questions in the domain of moral cognition. One such question that merits further exploration concerns the relationship between the perception of agency and the attribution of moral qualities. The possibility of moral evaluations and attribution of the true (and the "superficial") self is commonly contingent upon the perception of agency (mental states and intentions), but does this relationship go the other way around? In other words, does the perception of agency immediately triggers the attribution of a self (and a true self, assuming the two always come in one package)? If the attribution of the true self is indeed independent of any culturally specific notions of human nature, then there is no reason for this phenomenon to be restricted to human agents only and one would expect such attributions to extend to any perceived agents, including animals and supernatural beings. The moral taxonomy of the latter, however, poses a challenge for the notion of the inherently good true self. The rich body of ethnographic research on supernatural agents might indeed supply a good chunk of supporting evidence for the case of moral essentialism, in so far as the essence of many supernatural agents is often described through a number of core defining character traits. Nonetheless, these ascribed core traits are far from always being morally good. The supernatural cosmologies of many cultures feature creatures, whose "deeper" self is described as either inherently ambiguous and volatile (all sorts of trickster characters) or as plainly and purely evil (the Devil and all kinds of demons).

To give just one example, consider the case of *exotiká* – the monstrous creatures and malicious spirits described by the anthropologist Charles Stewart in his *Demons and the Devil* (1991), an ethnography of popular supernatural beliefs in modern Greece. According to Stewart, *exotiká*, whose origin can sometimes be traced back to Greek antiquity, not only co-exist with but have been successfully integrated into the Christian Orthodox cosmology, within which they assume the role of mediating the Church's rather abstract and philosophical notion of evil and personifying the demonic forces. Not all *exotiká* are seen as unambiguously evil. In fact, they are often represented as dangerous trickster characters, an encounter with whom might on a very rare occasion yield some benefit for a person, but in majority of cases would be harmful or deadly (p. 175). This ambiguity, however, leaves little space for the attribution of a morally good true self. Rather, the occasional out-of-character manifestations of benevolence on the part of *exotiká* are seen as a devilish trick – an attempt at disguising their demonic nature in order to deceive, seduce, and manipulate their unsuspecting victims.

Provided that one accepts that people, who believe in supernatural agents, indeed process some of them as inherently bad, what are the implications of the example of the *exotiká* and similar cases for the theory of the true self? One possibility would be to treat such an example as anecdotal counterevidence against the moral goodness of the attributed true self. The other possible interpretation would be to view this case as evidence for the fact that the phenomenon of the true self attribution is robustly manifested in relation to the human agents only. But then, what is it about the specifics of processing of human and non-human agents that makes the attribution of the morally good true self invariable in the former case and optional in the latter? Another related question is whether it is for some reason easier for people to ascribe the bad deeper self to non-human, rather than to human agents? Although there is presently no experimental data on the attribution of moral traits to supernatural agents (especially to the malicious ones), the very fact of cross-cultural and historical persistence of beliefs in evil spirits is suggestive of the ease and readiness with which people are prepared to imagine agents, who are essentially (perhaps even ontologically) bad. The concept of the true self thus inevitably generates its own version of the proverbial problem of evil, which yet remains to be addressed.