5-17-2018

Greenbeards and Signaling: Why Morality isn't Indispensable

Toby Handfield
Monash University

John Thrasher
Chapman University, thrasheriv@chapman.edu

Julian García
Monash University

Recommended Citation

Handfield, Toby, John Thrasher, and Julian García. "Greenbeards and Signaling: Why Morality isn't Indispensable." *Behavioral and Brain Sciences*, vol. 41, 2018, e103. DOI: 10.1017/S0140525X18000080

Greenbeards and Signaling: Why Morality isn't Indispensable

Comments

Copyright

Target Article: The Difference Between Ice Cream and Nazis: Moral Externalization and the Evolution of Human Cooperation

Word Counts

Abstract 53

Main Text 998

References 214

Total 1327

Greenbeards and Signaling: Why Morality isn't Indispensable

Toby Handfield
Monash University
SOPHIS, Monash University,
Victoria 3800, Australia
+61 3 9905 3202
toby.handfield@monash.edu
tobyhandfield.com
ORCID: 0000-0003-2995-2067

John Thrasher
SOPHIS, Monash University,
Victoria 3800, Australia
+61 3 9902 0016
John.Thrsaher@Monash.edu
JohnJThrasher.com
ORCId: 0000-0001-9094-5283

Julian García
Monash University
Faculty of Information
Technology, Monash
University, Victoria 3800,
Australia
+61 3 9905 3654
julian.garcia@monash.edu
http://garciajulian.com

Abstract We argue that while moral objectivist moral attitudes may facilitate cooperation, they are not necessary for the high levels of cooperation in humans. This is implied by evolutionary models that articulate a mechanism underlying Stanford's account, and is also suggested by the ability of merely conventional social norms to explain extreme human behaviors.

The lead paper argues that the distinctive psychology that regards moral properties as objective, external features of the world is adaptive because it allows us to engage in beneficial cooperation through correlated interactions. Implicit in the account are two theses: (1) the *correlation thesis* that moral commitment serves as a correlating device, allowing fellow norm-followers to

associate with each other and to ostracize defectors; and (2) the *indispensability thesis* that externalized moral norms of this sort are necessary to achieve pro-social cooperation, at least at the high rate seen in humans.

We argue that the first thesis is plausible, but that it undermines the second thesis. If the correlation thesis is true, there is good reason to think that externalized moral attitudes are not indispensable for achieving cooperation, but are merely one possible solution among many.

There is already a well understood set of mechanisms by which agents may ensure correlated interaction with fellow biological altruists. It is not clear how Stanford's account is supposed to relate to this menu of options. Is it a type of reciprocity, costly signaling, group selection, or something else? We suggest there are three possibilities (not mutually exclusive) that are especially promising.

Costly signaling models require a diversity of types in the population (e.g. cooperators and defectors), who – at least in simple cases – face differential signaling costs, or who stand to make differential gains from being believed. Some types can afford to send signal that are uneconomic for other type and, hence, any sufficiently costly signal is credible. Costly signaling mechanisms have been proposed to explain phenomena like unconditional sharing, costly punishment, apology, and guilt (Gintis, Smith, & Bowles, 2001; Jordan, Hoffman, Bloom, & Rand, 2016; Ohtsubo & Watanabe, 2009) Could externalizing moral psychology be a form of costly signal also?

It is conceivable. By rigid commitment to moral attitudes, an agent might incur a cost that a defector would not be willing to pay. But for this hypothesis to be plausible, we need an explanation for why externalized moral attitudes are especially burdensome or costly. Prima facie, it is hard to see that they are. Compared to better established examples of costly signals, including rituals such as fasting, bodily mutilation, and animal sacrifice, it does not seem like having the belief that moral requirements are objective features of the world is especially costly at all.

The *greenbeard hypothesis* is that altruistic traits are genetically linked to a distinctive phenotypic trait. If only altruists can develop green beards, then a strategy of cooperating with fellow green beards is a plausible evolutionary outcome. A genetic barrier to green-bearded defectors keeps the world safe for cooperation. Recent work (cf. Gardner & West, 2010) has shown that green beard mechanisms can sustain cooperation even if the link between beard and altruism is not strict, but permits some plasticity. The result is an unstable dynamic, in which the population cycles through different beard colors, with bursts of cooperation in the beginning of each beard cycle, followed by invasion by defectors, followed by development of a new beard color (Jansen & van Baalen, 2006; Traulsen & Nowak, 2007). Perhaps the best elaborated account of how these sorts of dynamic might explain actual human cooperation is the case of accents, which are certainly hard to fake, and are also not tightly linked to any particular genes (Cohen, 2012).

Both costly signaling and greenbeard accounts undermine the indispensability thesis, however, because both imply that the cooperative correlating mechanism is arbitrary. Stanford emphasizes that the moral psychology which he seeks to explain appears to be cross-culturally robust. This makes it quite unlike rituals or accents, which vary dramatically across time and space.

Finally, the most promising mechanism to underlie Stanford's account is *social selection*. Suppose that a competitive mating environment exists in which fitness is enhanced by finding reliable long-term cooperative partners. In such a market, it is adaptive to have a reputation for being reliable. So, we predict adaptations that make one sensitive to reputation (cf. Haley & Fessler, 2005). Cooperative, altruistic behaviors may then be adaptive because they enhance one's reputation—even if those behaviors are done cynically, for reputation enhancing reasons. The competitiveness of the mating market may then drive this process so that ever better demonstrations of reliability are required in order to obtain a mate. In this setting, it may be more cost effective to *be reliable* than to merely *appear reliable* (Sperber & Baumard, 2012). One particular way to be reliable is to take moral facts as external, objective demands. Notably, this process explains the development of 'high' quality types as emerging from a competition among lower quality types who are more self-serving in their pursuit of reputation.

This better explains the adaptive function of externalizing psychology in particular, but without more explicit modeling, it remains open that a population in equilibrium may have only a small minority who display this phenotype. The main mechanism –reputation sensitive coordination with the prevailing norms—may explain most observed cooperation. Indeed, this is plausible. Consider the evidence of social norms prevailing over moral commitments, revealed both in history and in experimental settings (e.g., Haney, Banks, & Zimbardo, 1972). Externalizing psychology is apt to fascinate philosophers, who are in the grip of related meta-ethical puzzles dating back to Plato's *Euthyphro*, but this is not yet evidence that it plays a significant role in achieving actual cooperation.

Debates about the genesis of human cooperation are unlikely to make significant advances without comparison of models "in the field"—using disciplines such as archaeology, ethnography, genetics, and experimental economics. This work, however, requires models that can deliver testable predictions. We await with interest to see whether Stanford's proposal constitutes a novel mechanism, with novel predictions, or if it can be assimilated to existing mechanisms. We have suggested here that if it is assimilated to existing mechanisms, there is little hope for the indispensability thesis. The "categorical imperative" nature of moral commitments may be a contingent artefact of our idiosyncratic evolutionary history.

References

Cohen, E. (2012). The Evolution of Tag-Based Cooperation in Humans: The Case for Accent. *Current Anthropology*, *53*(5), 588–616. https://doi.org/10.1086/667654

Gardner, A., & West, S. A. (2010). Greenbeards. *Evolution*, *64*(1), 25–38. https://doi.org/10.1111/j.1558-5646.2009.00842.x

Gintis, H., Smith, E., & Bowles, S. (2001). Costly Signaling and Cooperation. *Journal of Theoretical Biology*, *213*(1), 103–119. https://doi.org/10.1006/jtbi.2001.2406

Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, *26*(3), 245–256. https://doi.org/10.1016/j.evolhumbehav.2005.01.002

Haney, C., Banks, C., & Zimbardo, P. (1972). Interpersonal Dynamics in a Simulated Prison. *International Journal of Criminology and Penology*, *1*, 69–97.

Jansen, V. A. A., & van Baalen, M. (2006). Altruism through beard chromodynamics. *Nature*, *440*(7084), 663–666. https://doi.org/10.1038/nature04387

Jordan, J. J., Hoffman, M., Bloom, P., & Rand, D. G. (2016). Third-party punishment as a costly signal of trustworthiness. *Nature*, *530*(7591), 473–476. https://doi.org/10.1038/nature16981

Ohtsubo, Y., & Watanabe, E. (2009). Do sincere apologies need to be costly? Test of a costly signaling model of apology. *Evolution and Human Behavior*, *30*(2), 114–123. https://doi.org/10.1016/j.evolhumbehav.2008.09.004

Sperber, D., & Baumard, N. (2012). Moral Reputation: An Evolutionary and Cognitive Perspective. *Mind & Language*, *27*(5), 495–518. https://doi.org/10.1111/mila.12000

Traulsen, A., & Nowak, M. A. (2007). Chromodynamics of Cooperation in Finite Populations. *PLOS ONE*, *2*(3), e270. https://doi.org/10.1371/journal.pone.0000270