UC Irvine
UC Irvine Previously Published Works

Title
Rationalization may improve predictability rather than accuracy

Permalink
https://escholarship.org/uc/item/2qs8312t

Authors
Stanford, P Kyle
Thomas, Ashley J
Sarnecka, Barbara W

Publication Date
2020

DOI
10.1017/s0140525x19002279

Peer reviewed

Rationalization may improve predictability rather than accuracy

P. Kyle Stanford
stanford@uci.edu
949-824-1520
Dept. of Logic and Philosophy of Science
University of California, Irvine
3151 School of Social Sciences-Plaza A
Irvine, CA 92697-5100
https://faculty.sites.uci.edu/pkylestanford/

Ashley J. Thomas
athomas@g.harvard.edu
805-701-5276
Department of Psychology
Harvard University
https://ashleyjothomas.wixsite.com/mysite

Barbara W. Sarnecka
sarnecka@uci.edu
949-824-8495
University of California, Irvine
Dept. of Cognitive Sciences
2201 SBSG
Irvine, CA 92697-5100
https://sites.google.com/uci.edu/sarneckalab

Word Counts
Abstract: 72
Main text: 849
References: 112
Entire text: 1,110

Abstract:
We present a theoretical and an empirical challenge to Cushman's claim that rationalization is adaptive because it allows humans to extract more accurate beliefs from our non-rational motivations for behavior. Rationalization sometimes generates more adaptive decisions by making our beliefs about the world less accurate. We suggest that the most important adaptive advantage of rationalization is instead that it increases our predictability (and therefore attractiveness) as potential partners in cooperative social interactions.

Commentary:

Cushman makes a compelling case that rationalization is a form of what he calls "representation generalization" and that such representation generalization is itself a crucial feature of action-guiding cognition. We agree with Cushman that these processes of representation generalization, including rationalization, are broadly adaptive. We are less convinced, however, by his claim that rationalization is adaptively advantageous because it allows humans to extract valuable information ("true beliefs and useful desires") from the highly evolved inclinations of our non-rational action-guiding systems. We see two problems with this proposal, one theoretical and one empirical.

We begin with the theoretical problem. Cushman recognizes that our actions are generated by a complex combination of rational (beliefs, desires) and non-rational (instincts, habits, norm-compliance) impulses, and he suggests that rationalization "constructs new beliefs and desires where none had existed, in order to extract information from the *non-rational* processes that influence our behavior." Thus, if I must decide which route to take as I walk my dog, and the complex product of the non-rational and rational influences on my behavior leads me to walk along the road rather than along the river, I might rationalize that decision by deciding that I am afraid of the river route. On Cushman's account, I will then come to accept and adopt that rationalization (my fear of the river route or a belief about its danger) as a further, conscious motivation for avoiding the river. But creating this new, additional motivation will not change the impact of those same non-rational action-guiding impulses on my behavior. Thus, if I face the same (or a similar) decision tomorrow, the very same non-rational action-guiding impulses will remain in place and continue to influence my behavior, ensuring that my degree of aversion to the river route is not explained by the magnitude of my new consciously-accessible fear (or my beliefs about its danger) and triggering a new round of rationalization that strengthens or intensifies that fear in response. (Indeed, if these non-rational impulses fully preserve their influence, the degree of mismatch between my actions and my consciously accessible motivations should remain just as large as it was initially.). It seems that this process of repeated rationalization will continue as long as the relevant behavior is generated by a combination of non-rational and rational impulses, ultimately ensuring that the accumulated wisdom of my non-rational action-guiding impulses is substantially *overrepresented* in our actual (complex, multisystem) decisions about what to do (e.g., leaving me absolutely terrified of taking the river route). This would not be cause for concern if we thought the guidance of our non-rational action-guiding systems were always correct, but in that case there would be no evolutionary advantage in making such decisions accessible to rational influence in the first place.

The empirical problem is that even if rationalization invariably produces more fitness-maximizing choices and decisions, it sometimes does so by giving us less accurate beliefs about the world. One example is Liu and Ditto's work (2012) on moral coherence: Briefly, subjects who are induced by argument to shift their views concerning the *moral defensibility* of the death penalty will also shift their views about the extent to which the death penalty is *practically effective* in deterring crime. And we ourselves have shown that subjects will judge a child left alone in precisely the same circumstances to be in significantly more danger if the parent leaves for a morally unacceptable reason (e.g. an adulterous affair) than a morally neutral reason such as going to work (Thomas, Stanford, and Sarnecka, 2016) In these studies, it seems that representation generalization is inducing subjects to modify their factual beliefs in ways that do not increase their accuracy but instead make them better cohere with the subjects' own moral

judgments. This may well ensure more adaptive responses (ones that better reflect the community's normative views, for example), but if so this is achieved by making the subject's beliefs *less* accurate or responsive to relevant evidence.

The challenges we have presented also suggest an alternative to Cushman's hypothesis concerning the adaptive benefits of rationalization itself. Even if it does so at some cost to the accuracy of our beliefs, the one thing representation generalization undeniably increases is the coherence of our motivations for action and (therefore) the predictability of our behavior to others. There is reason to think that behaving predictably has adaptive value in its own right, as emphasized by those evolutionary theorists who appeal to mechanisms of partner choice, reputation management, and the virtues of predictable social partners more generally in seeking to understand the evolution of human ultrasociality (e.g. Baumard, et.al. (2013), Tomasello (forthcoming in BBS), Stanford (2018)). Thus, while we agree with Cushman that representation generalization is broadly adaptive and will sometimes increase the accuracy of our beliefs, we doubt that this is the only or even the most important way in which rationalization in particular increases our fitness. We suggest that rationalization in humans is adaptive in large part because it renders us more predictable (and therefore more attractive) partners for one another in the sorts of hypercooperative social structures on which human societies depend.

References

Baumard, N., André, J-B. & Sperber, D. (2013) A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences* 36(2): 59-122.

Liu, B.S., and Ditto, P.H. (2012) What dilemma? Moral evaluation shapes factual belief. *Social Psychological and Personality Science* 4(3): 316-323.

Stanford, P. K. (2018) The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation. *Behavioral and Brain Sciences* 41, E95. doi: 10.1017/S0140525X17001911.

Thomas, A.J., Stanford, P.K., and Sarnecka, B.W. (2016) No child left alone: Moral judgments about parents affect estimates of risk to children. *Collabra* 2: 10, doi: http://doi.org/10.1525/collabra.33.

Tomasello, M. (forthcoming) The moral psychology of obligation. *Behavioral and Brain Sciences.*