CAMBRIDGE
UNIVERSITY PRESS

RESPONSE

On the limited policy relevance of evolutionary explanations

Cass R. Sunstein[†]

Law School, Harvard Law School, Cambridge, MA, USA
E-mail: csunstei@law.harvard.edu

Abstract

Evolutionary explanations for behavioral findings are often both fascinating and plausible. But even so, they do not establish that people are acting rationally, that they are not making mistakes, or that their decisions are promoting their welfare. For example, present bias, optimistic overconfidence, and use of the availability heuristic can produce terrible mistakes and serious welfare losses, and this is so even if they have evolutionary foundations. There might well be evolutionary explanations for certain kinds of in-group favoritism, and also for certain male attitudes and actions toward women, and also for human mistreatment of and cruelty toward nonhuman animals. But those explanations would not justify anything at all. It is not clear that in Darwinia (a nation in which departures from perfect rationality have an evolutionary explanation), policymakers should behave very differently from Durkheimian policymakers (a nation in which departures from perfect rationality have a cultural explanation).

Keywords: heuristics; biases; evolutionary explanations

Suppose that we agree that human beings show 'present bias,' in the sense that they give a great deal of weight to short-term effects on their welfare, and relatively little weight to long-term effects. Suppose that we also agree that present bias leads, in important cases, to serious welfare losses, especially when people impose losses on their future selves (the problem of 'internalities'). Now suppose that a plausible evolutionary account is offered for present bias: In the early stages of human evolution, today and tomorrow greatly mattered, and it was not important to focus on what might happen in two decades, or three, or four. For policy purposes, the evolutionary account might be neither here nor there. It would not suggest that present bias is not a problem.

Now suppose we agree that human beings show 'optimistic overconfidence,' and that optimistic overconfidence helps explain a number of behavioral findings, including the planning fallacy. Suppose that we also agree that optimistic overconfidence has a plausible evolutionary explanation: In the early stages of human evolution,

---

[†]Robert Walmsley University Professor, Harvard University.

optimistic overconfidence could improve people's performance, and often save their lives. For policy purposes, the evolutionary account might be neither here nor there. It would not suggest that optimistic overconfidence is not a problem.

Suppose, finally, that we agree that human beings use the availability heuristic, in the sense that their judgments about risks are affected by whether relevant events come readily to mind. Suppose we agree that the availability heuristic can lead to erroneous judgments about risks, including excessive fear and excessive complacency. Suppose that an evolutionary explanation of the availability heuristic is plausible: In the early stages of human evolution, the use of that heuristic worked well, or well enough, and it is fast and frugal. For policy purposes, the evolutionary account might be neither here nor there. It would not suggest that the use of the availability heuristic is not a problem.

These brisk claims raise many questions. One of the most fundamental is what it means, exactly, to say that an evolutionary explanation for a behavioral bias is 'plausible' [for a valuable discussion in the context of strong reciprocity, see Fehr and Henrich (2003); for a similarly valuable discussion in the context of altruism, see Fehr and Fischbacher (2003)]. For many people, evolutionary explanations seem to produce a kind of cognitive 'click,' and also a kind of pleasure, even if those explanations are speculative and *ad hoc*. It is especially concerning if an evolutionary explanation for the opposite bias, or for no bias at all, would be equally plausible. Suppose, for example, that human beings were unrealistically pessimistic. If so, it would be easy to produce a speculative, *ex post* evolutionary explanation of that non-existent bias: Unrealistic pessimism might lead to salutary precautions. Exasperated by the speculative nature of (some) evolutionary accounts, Amos Tversky once exclaimed, 'Listen to evolutionary psychologists long enough, and you'll stop believing in evolution!' To be made convincing, an evolutionary explanation for a cognitive bias would have to be thickened in multiple ways [see Henrich (2020) for some impressive, relevant detective work]. No one should doubt that such thickening is possible (see Henrich, 2020), and there are fascinating evolutionary explanations for an assortment of behaviors, including altruistic punishment, cooperation, in-group favoritism, and spite (on some of these points, see Jones & Goldsmith, 2005; Henrich, 2020).

But the force of evolutionary explanations is not my topic here. For purposes of public policy, present bias, optimistic overconfidence, and the availability heuristic have evident importance. Their existence might help explain why certain interventions will not work. They might also suggest interventions that might be necessary, and that might succeed. Such interventions might take the form of taxes; for example, taxes on sugar-sweetened beverages might counteract present bias. They might take the form of mandates and bans; for example, vaccine mandates might be necessary to counteract optimistic overconfidence. They might take the form of nudges, either educative or architectural; for example, automatic enrollment in savings programs might counteract, at once, present bias, optimistic overconfidence, and use of the availability heuristic.

It is important to add that we might agree that some behavioral phenomenon has a plausible evolutionary explanation while also insisting that it leads to severe and systematic errors, and also to serious welfare losses. Present bias, optimistic overconfidence, and use of the availability heuristic can produce real trouble, and this is so

even if they have evolutionary foundations. There might well be evolutionary explanations for certain kinds of in-group favoritism, and also for certain male attitudes and actions toward women, and also for human mistreatment of and cruelty toward non-human animals. But those explanations would not justify anything at all (to be sure, they might have some significant implications for policy, see Jones & Goldsmith, 2005).

To clarify the point, imagine two worlds: Darwinia and Durkheimia. In Darwinia, human beings evolved to show present bias and optimistic overconfidence, and use of the availability heuristic. In Durkheimia, human beings did not evolve in the same way, but for many generations, the Durkheimian culture transmitted present bias, optimistic overconfidence, and use of the availability heuristic. In both Darwinia and Durkheimia, present bias, optimistic overconfidence, and use of the availability heuristic can cause serious welfare losses. Should Darwinian policymakers behave differently from Durkheimian policymakers? That is not at all clear. It is true that Darwinian policymakers might learn that (for example) present bias is quite robust and therefore hard to counteract through (let us say) educative nudges. But if present bias has an evolutionary explanation, it need not follow that it is quite robust and hard to counteract – any more than it follows that if present bias has a cultural explanation, it is not robust and easy to counteract.

It is also important to note that some evolutionary accounts appear to help explain differences across both regions and time, in a way that does have policy relevance (Henrich, 2020). Suppose, for example, that anthropologists find that present bias and optimistic overconfidence are not universal, and that some cultures show more of both than others, and that some cultures show neither. If so, we might have a 'behavioral market failure' in some places and not others, and some interventions might be highly successful in some places and wildly unsuccessful in others. Cultural differences can matter to behavioral public policy, whether we are speaking of poverty reduction, environmental harm, or pandemic response. Some explanations for those differences might have evolutionary components (Henrich, 2020; Henrich & Muthukrishna, 2021).

Turn in this light to Adam Oliver's provocative exploration (Oliver, 2021) of prospect theory, the reflection effect, and the four-fold pattern of risk preferences (Tversky & Kahneman, 1992), in accordance with which people tend to be:

- risk-averse for high-probability gains (and hence might prefer a sure gain of $100 to an 80% chance of winning $140, even though the latter has a higher expected value);
- risk-seeking for high-probability losses (and hence might prefer an 80% chance of losing $140 to a sure loss of $100);
- risk-seeking for low-probability gains (and hence might prefer a 10% chance of winning $500 to a sure gain of $60, even though the latter has a higher expected value); and
- risk-averse for low-probability losses (and hence might prefer a sure loss of $60 to a 10% chance of losing $500).

It would be interesting to know whether different nations and different cultures show differences on these counts, and whether any such differences might have

evolutionary explanations. Some evidence suggests, for example, that Chinese subjects are risk-seeking in all four contexts (Brumagim & Xianhua, 2005). One international survey finds that while 'most people worldwide follow in their behavior the typical features that prospect theory captures,' there are also significant *cultural* differences (Rieger *et al.*, 2011). The same survey finds that risk aversion for gains is higher in the Middle East and Eastern Europe, while loss aversion is higher in Latin American and Germanic and Nordic countries. Another such survey finds that prospect theory does replicate across 19 nations, but also that there is some heterogeneity among countries (Ruggeri *et al.*, 2020).

Oliver is not focused on international differences. For gambles that involve both money and health, Oliver assembles evidence that is consistent with the reflection effect. But his main aim is to show that the evidence does not necessarily support prospect theory. He urges instead that the 'results may be driven by evolved responses to circumstances that provoke perceptions of scarcity and abundance.' Oliver questions the widespread idea that the relevant patterns are 'erroneous or biased.' To be sure, they are not consistent with rational choice theory (because people are favoring options with a lower expected value), but that inconsistency does not show that people are making mistakes.

Oliver's argument on this count is relatively simple. Under conditions of abundance, he urges, people are risk-averse; under conditions of extreme scarcity, he also urges, people are risk-seeking. To support these conclusions, he points to research suggesting that in an uncertain environment, animals will rationally 'switch to risk seeking when a nutritional need has to be fulfilled and which cannot be met with a low risk option,' which suggests that 'the evolved emphasis – which may still today be rational in many circumstances – is on attempting to meet one's basic needs in the moment.' If this is correct, it would seem to follow that wealthy people will be risk-averse and that poor people will be risk-seeking, and that wealthy communities will be risk-averse and poor ones risk-seeking – hypotheses that would be worth testing and on which we have some evidence. But prospect theory does not make any prediction of that kind. Much of its interest lies in the claim that whether people are risk-averse or instead risk-seeking 'flips' depends on whether we are speaking of losses or gains, and on whether we are speaking of high probabilities or instead low ones. What – it is natural to ask – does that have to do with abundance or scarcity?

Oliver's intriguing answer is that probability can serve 'as a proxy for prior (and thus expected) frequency of success in the respondents' psychologies as they process the risky decisions that they face.' In Oliver's account, that means that high and low probabilities of a *gain* 'may respectively provoke perceptions of abundance and scarcity. … With abundance, people may be quite satisfied with an implicitly risk averse strategy – accepting a guaranteed amount that is less than the expected value of the gamble suffices.' By contrast, high and low probabilities of a *loss* can be seen to 'suggest scarcity and abundance', respectively, thus producing 'patterns of risk attitude that are consistent with the full reflection effect.'

I am not sure that I fully understand this argument, and for two different reasons. First: Is it really true that under conditions of abundance, people are risk-averse, and that under conditions of scarcity, people are risk-seeking? Are the Elon Musks of the world risk-averse, and are indigent people risk-seeking? Are wealthy communities

risk-averse, and are indigent communities risk-seeking? These are interesting empirical questions. We do not have definitive answers. Some research does find that in many (but not all) domains, low GDP is correlated with risk-seeking preferences (Vieider *et al.*, 2015). But other research fails to find that poor people or poor communities are distinctly risk-seeking (Orozco, 2010; Galliera & Ruström, 2019), and still other research finds that poor people are risk-averse, not risk-seeking (Fafchamps & Pender, 1997; Nielsen, 2001). Among large groups of rich or poor people, we should expect a great deal of heterogeneity. Nor is it at all clear that under conditions of scarcity, it is reasonable or rational to be risk-seeking. Everything depends on the nature and extent of the scarcity and the nature of the gamble. For those who are barely able to survive, it might well make sense to be risk-averse. (Note, again, that prospect theory does not predict risk aversion or risk-seeking in general; everything depends on whether we are speaking of gains or losses, and of low or high probabilities.)

Second and perhaps more fundamentally: Is it really true that high and low probabilities can 'provoke perceptions of abundance and scarcity,' in such a way as to explain the reflection effect? How, exactly, do probabilities 'provoke' such 'perceptions'? Consider an example. Would you rather (1) receive $100 for sure or (2) receive an 80% chance of obtaining $140? If you choose (1), you are risk-averse in the domain of high-probability gains, but it is not so clear that the stated (high) probability triggers 'perceptions of abundance.' Whether it does so is a testable hypothesis; is there supportive evidence? Or consider another example. Would you rather (1) lose $100 for sure or (2) face a 10% chance of losing $900? If you choose (1), you are risk-averse with respect to low-probability losses, but does the stated (low) probability trigger perceptions of scarcity? Maybe. But does any evidence suggest that it does so?

Arguing for 'genuine explanatory depth' for behavioral findings, Oliver urges that the four-fold pattern 'may be perfectly reasonable responses given the circumstances and the environment' in which people find themselves. Perhaps so. What matters is welfare, not money. It is not unreasonable to purchase insurance, even if the expected (monetary) value of the purchase is higher than the expected (monetary) value of the loss. One reason is that once we speak in terms of welfare rather than money, the purchase might turn out to be entirely rational. But when we are speaking of monetary gambles, the pattern of risk preferences identified by prospect theory is challenging to defend as 'perfectly reasonable.' If you keep choosing $100 over an 80% chance to win $140, you will lose a lot of money over your lifetime. And even if we were able to identify a plausible evolutionary explanation for the four-fold pattern, it would not follow that those who show that pattern are reasonable (or rational).

Oliver also thinks that his claims are relevant to policy. It is tempting, for example, to think that certain forms of gambling behavior come from errors and biases, some of them connected with prospect theory, and that the relevant tendencies can be exploited by those who seek to encourage people to gamble and to take gamblers' money. Oliver responds that departures from 'standard notions of rational choice,' including by gamblers, 'are not necessarily irrational in a more inclusive sense of the term.' Gamblers might believe that gambling 'is the only possible way for them to escape a particular financial predicament. If their situation is already dire in the absence of gambling, then attempting to regulate their gambling behaviors in some way without other supporting mechanisms may do little to assist them.'

For public officials who are concerned about gambling, how helpful is that? Oliver is surely correct about some gamblers. But consider four propositions. (1) Many people who are in dire situations do not gamble. (2) Many people who are in dire situations hate the idea of gambling. (3) Many people who are not in dire situations do gamble. (4) Many people end up in dire situations *because they gamble*. To be sure, some people gamble because they are desperate for money, and under certain assumptions, it is rational for them to do that. But it is not entirely clear that we need evolutionary theory to know that we ought to help people who gamble because they are in dire situations.

Behavioral findings, including but not limited to the reflection effect, help explain why gambling occurs. But my larger points lie elsewhere. Even when evolutionary explanations for behavioral findings are plausible, they do not establish that people are acting rationally, that they are not making mistakes, or that their decisions are promoting their welfare.

References

Brumagim, A. L. and W. Xianhua (2005), 'An examination of cross-cultural differences in attitudes towards risk: Testing prospect theory in the People's Republic of China", *Multinational Business Review*, **13**(3): 67–86.

Fafchamps, M. and J. Pender (1997), 'Precautionary saving, credit constraints, and irreversible investment: Theory and evidence from semiarid India', *Journal of Business and Economic Statistics*, **15**: 180–94.

Fehr, E. and U. Fischbacher (2003), 'The nature of human altruism', *Nature*, **425**(6960): 785–91. doi:10.1038/nature02043.

Fehr, E. and J. Henrich (2003), 'Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism', in P. Hammerstein (ed.), *Genetic and cultural evolution of cooperation*, Cambridge: MIT Press, 55–82.

Galliera, A. and E. E. Rutsröm (2019), "*Grabbing the opportunity*": *Risk attitudes among poor households in the US*. GSU Center for the Economic Analysis of Risk. Working Paper No. 2019–06, 1–31.

Henrich, J. (2020), *The WEIRDest people in the world*. New York: Allen Lane.

Henrich, J. and M. Muthukrishna (2021), 'The origins and psychology of human cooperation', *Annual Review of Psychology*, **72**(1): 207–40.

Jones, O. and T. H. Goldsmith 2005), 'Law and behavioral biology,' *Columbia Law Review*, **105**: 405–501.

Nielsen, U. (2001), *Poverty and attitude towards time and risk experimental evidence from Madagascar*. Royal Veterinary and Agricultural University of Denmark, Department of Economics and Natural Resources, Working Paper.

Oliver, A. (2021), 'Reflecting on reflection: Prospect theory, our behaviours, and our environment', *Behavioural Public Policy*, 1–11.

Orozco, G. A. C. (2010), *Risk preferences under extreme poverty: A field experiment*. Universidad de los Andes Centro de Estudios sobre Desarrollo Economico, 1–39.

Rieger, M. O., M. Wang and T. Hens (2011), *Prospect theory around the world*. NHH Department of Finance & Management Science. Discussion Paper No. 2011/19, 1–36.

Ruggeri, K., S. Alí, M. L. Berge, G. Bertoldo, L. D. Bjørndal, A. Cortijos-Bernabeu, C. Davison, E. Demić, C. Esteban-Serna, M. Friedemann, S. P. Gibson, H. Jarke, R. Karakasheva, P. R. Khorrami, J. Kveder, T. L. Andersen, I. S. Lofthus, L. McGill, A. E Nieto, J. Pérez, S. K. Quail, C. Rutherford, F. L. Tavera, N. Tomat, C. V. Reyn, B. Većkalov, K. Wang, A. Yosifova, F. Papa, E. Rubaltelli, S. Linden and T. Folk (2020), 'Replicating patterns of prospect theory for decision under risk', *Nature Human Behavior*, **4**: 622–33.

Tversky, A. and D. Kahneman (1992), 'Advances in prospect theory: Cumulative representation of uncertainty', *Journal of Risk and Uncertainty*, **5**(4): 297–323.

Vieider, F. M., M. Lefebvre, R. Bouchouicha, T. Chmura, R. Hakimov, M. Krawczyk and P. Martinsson (2015), 'Common components of risk and uncertainty attitudes across contexts and domains: Evidence from 30 countries', *Journal of the European Economic Association*, **13**: 421–52.