

How actions create – not just reveal – preferences

Dan Ariely¹ and Michael I. Norton²

Forthcoming in TRENDS in Cognitive Sciences Vol.12 No.1 (2008)

¹ Duke University, One Towerview Road, Durham, NC 27708, USA ² Harvard Business School, Soldiers Field Road, Boston, MA 02163, USA

How actions create – not just reveal – preferences

Abstract

The neo-classical economics view that behavior is driven by – and reflective of – hedonic utility is challenged by psychologists' demonstrations of cases in which actions do not merely reveal preferences but rather create them. In this view, preferences are frequently constructed in the moment and are susceptible to fleeting situational factors; problematically, individuals are insensitive to the impact of such factors on their behavior, misattributing utility caused by these irrelevant factors to stable underlying preferences. Consequently, subsequent behavior might reflect not hedonic utility but rather this erroneously imputed utility that lingers in memory. Here we review the roles of these streams of utility in shaping preferences, and discuss how neuroimaging offers unique possibilities for disentangling their independent contributions to behavior.

Introduction

Both economists and psychologists place significant importance on the construct of utility. Although this construct takes different forms within each discipline (as well as in subfields within each discipline), the idea of utility as a central driving force in human motivation and behavior is well accepted. However, despite the importance of utility for both disciplines, psychologists and economists have substantially different views about the nature of utility. A central aspect of utility for economics is hedonic utility, the units of pleasure and pain that the brain assigns to each and every event in the world [1,2]. In this view the brain functions as a hedonic meter, providing an individual with a preview of the pain or pleasure she can expect to get from a trip to the dentist or a vacation in Belize. These hedonic expectations need not turn out to be accurate (one can decide to see a movie on the expectation that it will be enjoyable only to find it horribly boring), but positive and negative utilities serve as the input for decision-making, and as a result shape people's behavior. According to this view, the actions that people take reveal the underlying utility of that behavior [3–5] (Figure 1a).

Although economists view utility as the imputed sum of the positive and negative aspects of some course of action, psychologists have argued that people do not have well-formulated preferences and often construct utilities in the moment [6–9]. In some sense, the well-documented finding that attitudes (psychologists' preferred term for the sum of positive and negative aspects) are imperfect predictors of behavior [10,11] is a clear indicator of the less-than-perfect relationship between utilities and actions. In our view, individuals' lack of stable preferences has two key consequences. First, people's decisions can be highly sensitive to situational factors, even when such factors are unrelated to the actual utility of that course of action. Worse still, people seem to be relatively insensitive to the impact of these situational factors in shaping their behavior, and instead misattribute the behavior caused by these fleeting factors to stable underlying preferences, inferring that their past actions are good indicators of their utilities [12]. In short, from the psychological point of view, actions can create, rather than reflect, preferences. The second important consequence of the ill-defined nature of utilities is that

when individuals assess their own utilities when called upon to make a choice, they rely not only on stable hedonic utilities but also on their memories of utility for their own past behaviors. Of course, relying on memory as a source of utility is not problematic in the many cases in which memories truly reflect the hedonic experience. We suggest, however, that when memories of the utility of some action have been biased by irrelevant situational factors present when they were enacted, subsequent behavior might not reflect hedonic utility but rather be driven in part by traces of erroneously imputed utility that linger in memory (Figure 1b).

Here we review evidence of cases in which people's actions shape utilities rather than reflect them in both the short- and long-term, discuss the implications of this bidirectionality for the measurement of utility and finally suggest ways in which a neuroeconomics approach might be uniquely suited for teasing apart the manner in which different streams of utility interact to impact preferences and behavior.

Psychological evidence for inferring utilities from actions Festinger and Carlsmith's [13] classic experiment is one of the most famous demonstrations of the influence of actions on utilities: participants were paid a small or large sum for completing a boring task and then were asked to pretend to another person that they enjoyed this task. This setup pitted the negative utility of the task against a behavior that implied a belief that the task had positive utility. The result that paying people less to lie about the enjoyment level of a boring task made them like that task more suggests the impact of actions on utility: When people were not paid enough to justify lying about that task, their action led them to infer that they derived utility from this objectively boring task, although of course the source of that utility was unrelated to the task itself. A second example of the influence of behavior on utilities arises from an investigation of 'coherent arbitrariness' [14]. In one experiment, participants were asked to indicate whether they would pay a given price – arbitrarily set by the last two digits of their social security number – for a range of products, after which they bid for those items in an auction. Despite the arbitrary nature of these price anchors (the utility of a bottle of wine likely has little to do with one's social security number) people with higher social security

numbers bid more for bottles of wine. Nor are these the only examples of the impact of situational factors on preferences, such as the many demonstrations of the impact of experimentally- induced emotions and arousal on preferences [15,16].

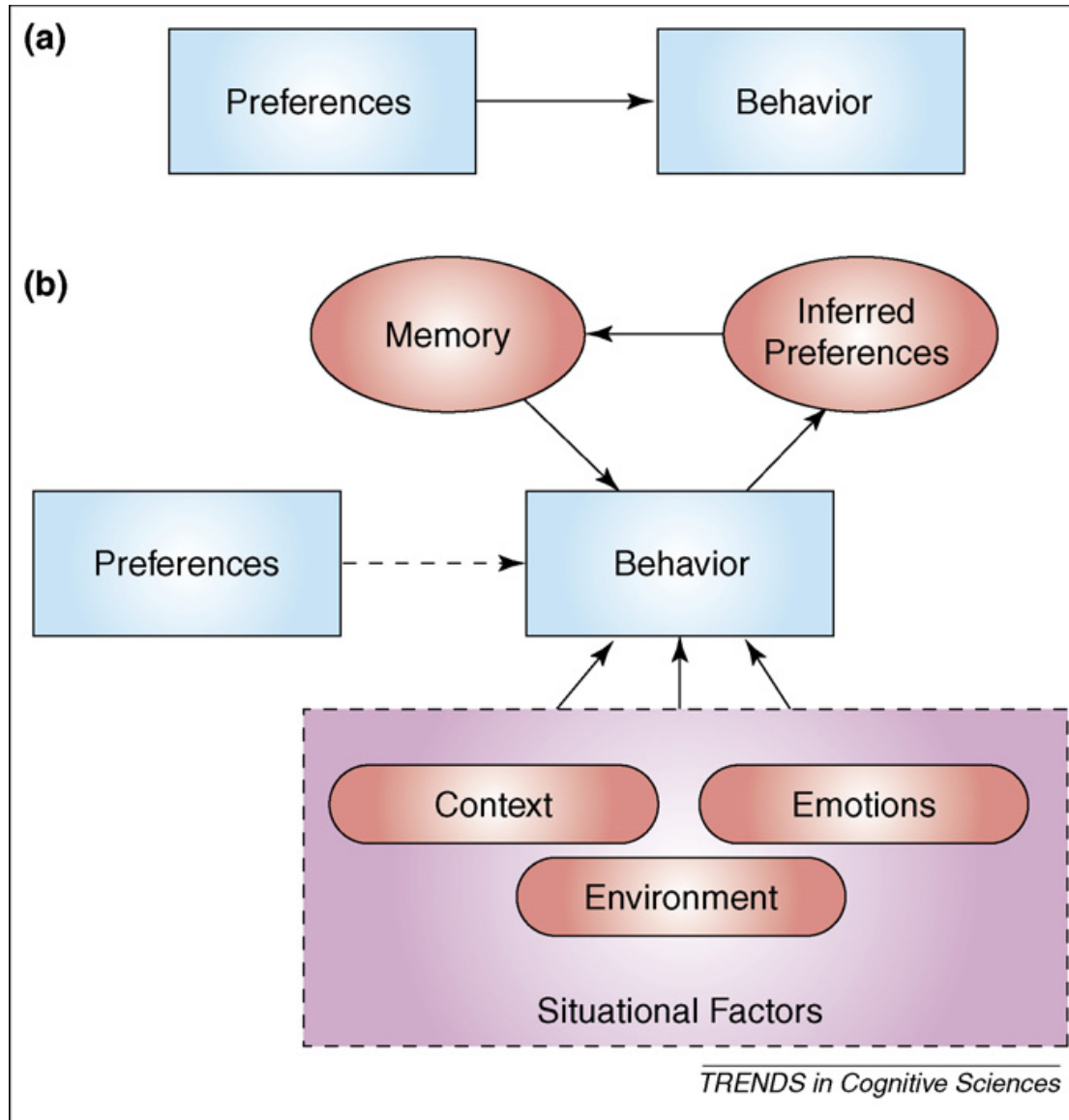


Figure 1. Two views of the relationship between preferences and actions. (a) This captures the view from neo-classical economics, in which a calculation of hedonic utility underlies behavior. In this view, preferences cause actions and actions reveal preferences. (b) This captures the authors' account, in which behavior is driven not just by hedonic utility, but also by situational factors, causing individuals to erroneously infer stable preferences from their actions. Moreover, later behavior is also driven in part by memories for these inferred preferences. In this view, actions can create preferences and biased memories then reinforce these preferences.

The above examples of actions influencing preferences owing to factors unrelated to the actual utility of those preferences are laboratory-based demonstrations; one critique of

such cases is that in the laboratory, it is easy to create exogenous variables that impact preferences, but that in the real world, contexts are endogenous to decisions [17]^{a, b}, making these demonstrations irrelevant to understanding utility. However, the world seems quite capable of producing exogenous factors that people erroneously incorporate into their calculation of utility. For example, in one recent investigation, enrollment rates at a given college were dramatically influenced by whether students happened to visit that campus on a sunny day instead of a rainy day, suggesting that students were incorporating this exogenous source of utility into a stable preference for that college^c. If unrelated situational factors can impact predicted utility and behavior even for momentous life decisions such as which college to attend, it might be likely that such factors also impact more mundane decisions, though more research using field rather than laboratory data is warranted.

The impact of inferring utilities from actions on memory and subsequent behavior even if real-world preferences are impacted by unrelated factors, it is possible that people correct these preferences over time (or that the market forces such correction). Here, too, however, some evidence suggests that people might not do so. Those participants whose values for bottles of wine had been set arbitrarily by their social security numbers, for example, made bids for subsequent bottles of wine that followed in a coherent manner, such that better bottles of wine fetched higher prices, and inferior bottles lower prices [14]. Even in Festinger and Carlsmith's experiment [13], those participants who reported liking the task – having misattributed their display of positive utility to a stable preference – reported being more eager to return to participate in a similar experiment, suggesting a longer- term impact of their initially biased preferences.

These results demonstrate a kind of 'self-herding', in which people observe their

^a Lazear, E.P. et al. (2006) Sorting in experiments with application to social preferences. NBER Working Paper No. W12041. Available at SSRN: <http://ssrn.com/abstract=883090>.

^b Levitt, S.D. and List, J.A. (2006) What do laboratory experiments tell us about the real world? Working paper, University of Chicago.

^c Simonsohn, U. (2007) Weather to go to college. Working paper, University of Pennsylvania. Available at SSRN: <http://ssrn.com/abstract=964091>.

past behavior, infer some amount of utility and act in accordance with that inference of utility, despite the fact that this behavior can be based not on an initial choice driven by hedonic utility but on any host of trivial situational factors that impacted that first decision. Unlike a reinforcement learning model in which past behavior enables people to ‘fine-tune’ their predictions of the hedonic utility they will receive from different courses of action, our account holds that past behavior gives people an arbitrary set point, not necessarily related to hedonic utility, which people then treat as a meaningful input into their subsequent calculation of utilities. Most importantly for our discussion of the implications of this account for economic theory, this type of self-herding might be particularly pronounced when subsequent choices are related to initial choices by observable economic variables (size, delivery time etc.) because these variables make it easy for individuals to adjust carefully upward and downward from their arbitrarily formed preference.

Somewhat surprisingly, even laypeople seem to believe at some level that actions change utilities, as evidenced by the curious case of self-signaling, in which people behave in a certain way not because some action offers utility in itself but because engaging in that action convinces them that they are the kind of person who derives utility from that action, such as behaving altruistically [18]^d. In sum, the psychological literature challenges the notion of utility in neo-classical economics by providing a wide range of evidence that actions are not merely the consequence but also the cause of preferences.

Implications for measuring utility

The notion that the relationship between utilities and actions is bidirectional raises a fundamental issue for the measurement of utility: do actions influence hedonic utility itself or do they change some other input, such as erroneous memories? Consider again the case of arbitrarily determined prices for bottles of wine. In the standard economic framework, individuals have an expected hedonic utility for the pleasure they expect to receive from consuming a particular wine, which translates directly to their actions (e.g. to the amount they are willing to pay for that wine). This hedonic utility would obviously

vary according to certain states of the individual – someone wanting to make a toast might have greater utility for a bottle of wine – and these situational changes in utility are reflected in actions – this person will pay more for wine. However, in the neo-classical view, when the person no longer needs to make a toast, that increased utility will not be factored into subsequent decisions.

We suggest that, in addition to hedonic utility, decisions people make about paying for wine also involve the residue of their past decisions, even when these past decisions are irrelevant to their current state. Paying more for a bottle of wine because one's social security number is higher changes the valuation of that wine in the moment, but also impacts behavior in the longer term because that higher price point might change their valuations for that kind of wine (such that they continue to treat that brand as offering more utility than other brands) and for other wines (bottles from nearby vineyards might also be valued more). The key point here is that this entire process can be set into motion randomly, but people's reliance on their past behavior – and their insensitivity to the impact of situational factors on that behavior – makes them behave as though these wines have different hedonic utilities.

Using neuroeconomics to measure utility

This example illustrates two key challenges in the measurement of utility. First, if past actions change individuals' estimation of utility, it is also the case that actions cannot be taken as a simple measure of utilities, leaving the neo-classical economist in a bind. Second, the manner in which past actions influence utility is unclear: is it the case that paying more for wine in the past increases the actual hedonic utility of that wine, or alternatively is it the case that the hedonic utility is the same but the person's estimation of this utility is now biased? Most importantly, how might we measure these potentially distinct – and at times even contradictory – streams of utility? In markets, we can observe only people's actions, which (as reviewed in the previous section) might not reflect people's underlying utility; in experiments we can observe actions and ask people to report on the reasons for their decisions, but given the limitations of such reports [19–23] this method is also suspect. The emerging field of neuroeconomics [24–26] offers some

promise for assessing utility directly without the usual mediating mechanism of relying on people's behavior or their stated preferences to infer utility, and it thus offers a possible means for determining how actions impact utilities. Indeed, many researchers have been engaged in attempting to localize the source of utilities to specific brain regions tied to reward. Across a host of stimuli – from food to money to sports cars to human faces [27–34] – researchers have implicated a common network of regions involved in reward, including the orbitofrontal cortex (OFC), ventral striatum, and ventromedial prefrontal cortex (vmPFC).

Most relevant to our account, some evidence suggests that even when experiences with stimuli are altered by some manipulation, a trace of 'true' hedonic utility might still remain. For instance, although labeling wine with higher prices or labeling sodas with popular brands does not cause differences in recruitment of gustatory regions, suggesting that underlying taste utilities might be pre-served, such manipulations do impact activity in reward regions (mPFC and OFC), as well as regions implicated in memory [35]^e suggesting the possible involvement of memories for past behaviors in these shifts in utility.

Our account – in which actions lead to inferred preferences, memories for which then influence future behavior (Figure 1b) – is one that unfolds over time; experiments in which people's self-reports of their utilities (a measure of their inferred preferences) as well as the response from different regions of the brain involved in memory and reward are used as simultaneous predictors of future behavior would thus be revealing in determining the relative contributions of each. Indeed, some efforts to elucidate how different aspects of the decision-making process (from predicting to anticipating to experiencing utility) recruit different regions related to reward are underway [36–39], offering promising avenues for future work.

Conclusion

In conclusion, we suggest that rather than being driven by hedonic utility, behavior is based in part on observations of past actions, actions that have been influenced by

essentially random situational factors – such as the weather – but that people interpret as reflective of their stable preferences. In short, people have some sense of the hedonic utilities of different options, but what determines their actions is not only the hedonic utility of those options but also their (potentially biased) memories for their past actions, which can shape their future utilities and thus future actions. It is possible that memory for past behaviors will have an important role in determining actions early in a stream of similar decisions, but that over time the effect of memory will be substituted with learned hedonic utility. At the same time, it is also possible that memory for past behaviors will have an increasing impact later in a sequence of decisions because its accessibility – and therefore influence – might increase. We believe that neuroimaging methods can offer insight into the central dilemma highlighted here – the relative roles of these different streams of utilities in impacting preferences and subsequent behavior. By using all possible sources of utility as predictors of subsequent choices, and using imaging data to parse the relative contributions of each by localizing these streams to distinct brain regions, the inter- play of actions and utility, and ultimately the formation of stable preferences, could be better captured.

^e Plassmann, H. et al. The modulation of marketing actions on the neural representation of experienced utility. Working paper, California Institute of Technology.

Acknowledgements

We thank Malia Mason, Uri Simonsohn and two anonymous reviewers for their helpful suggestions.

References

1. Bentham, J. (1948) *An Introduction to the Principle of Morals and Legislations* (1789), Blackwell
2. Kahneman, D. et al. (1997) Back to Bentham? Explorations of experienced utility. *Q. J. Econ.* 112, 375–405
3. Samuelson, P.A. (1938) A note on the pure theory of consumers' behaviour. *Economica* 5, 61–71
4. Stigler, G.G. (1950) The development of utility theory, I,II. *J. Pol. Econ.* 43, 307–327 373–396
5. Varian, H. (1992) *Microeconomic Analysis*, (3rd edn), Norton
6. Kahneman, D. and Snell, J. (1992) Predicting a changing taste: do people know what they will like? *J. Behav. Decision Making* 5, 187–200
7. Payne, J.W. et al. (1993) *The Adaptive Decision Maker*, Cambridge University Press
8. Shafir, E. et al. (1993) Reason-based choice. *Cognition* 49, 11–36
9. Slovic, P. (1995) The construction of preference. *Am. Psychol.* 50, 364– 371
10. Eagly, A. and Chaiken, S. (1993) *The Psychology of Attitudes*, Harcourt Brace Jovanovich
11. Fazio, R.H. (1990) Multiple processes by which attitudes guide behavior – the MODE model as an integrative framework. *Adv. Exp. Soc. Psychol.* 23, 75–109
12. Bem, D.J. (1972) Self-perception theory. *Adv. Exp. Soc. Psychol.* 6, 1–62
13. Festinger, L. and Carlsmith, J.M. (1959) Cognitive consequences of forced compliance. *J. Abnorm. Soc. Psychol.* 58, 203–211
14. Ariely, D. et al. (2003) Coherent arbitrariness: stable demand curves without stable preferences. *Q. J. Econ.* 118, 73–105
15. Schachter, S. and Singer, J.E. (1962) Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.* 69, 379–399
16. Strack, F. et al. (1988) Inhibiting and facilitating condition of facial expressions: a non-obtrusive test of the facial feedback hypothesis. *J. Pers. Soc. Psychol.* 54, 768–777
17. Glaeser, E.L. (2004) Psychology and the market. *Am. Econ. Rev.* 94, 408–413
18. Bodner, R. and Prelec, D. (2002) Self-signaling and diagnostic utility in everyday decision making. In *Collected Essays in Psychology and Economics* (Brocas, I. and Carillo, J., eds), Oxford University Press

19. Johansson, P. et al. (2005) Failure to detect mismatches between intention and outcome in a simple decision task. *Science* 310, 166–199
20. Libet, B. (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav. Brain Sci.* 8, 529–566
21. Nisbett, R.E. and Wilson, T.D. (1977) Telling more than we can know. *Psychol. Rev.* 84, 231–259
22. Norton, M.I. et al. (2004) Casuistry and social category bias. *J. Pers. Soc. Psychol.* 87, 817–831
23. Wegner, D.M. (2002) *The illusion of conscious will*, MIT Press
24. Camerer, C. et al. (2005) Neuroeconomics: how neuroscience can inform economics. *J. Econ. Lit.* 43, 9–64
25. Glimcher, P.W. and Rustichini, A. (2004) Neuroeconomics: the confluence of brain and decision. *Science* 306, 447–452
26. Sanfey, A.G. et al. (2006) Neuroeconomics: cross-currents in research on decision-making. *Trends Cogn. Sci.* 10, 108–116
27. Aharon, I. et al. (2001) Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron* 32, 537–551
28. Breiter, H.C. et al. (2001) Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30, 619–639
29. Erk, S. et al. (2002) Cultural objects modulate reward circuitry. *Neuroreport* 13, 2499–2503
30. Knutson, B. et al. (2003) A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage* 18, 263–272
31. Kringelbach, M.L. (2005) The human orbitofrontal cortex: linking reward to hedonic experience. *Nat. Rev. Neurosci.* 6, 691–702
32. Montague, P.R. et al. (2006) Imaging valuation models in human choice. *Annu. Rev. Neurosci.* 29, 417–448
33. O’Doherty, J.P. (2004) Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14, 769–776
34. Padoa-Schioppa, C. and Assad, J.A. (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226

35. McClure, S.M. et al. (2004) Neural correlates of behavioral preference for culturally familiar drinks. *Neuron*. 44, 379–387
36. Berridge, K.C. and Robinson, T.E. (2003) Parsing reward. *Trends Neurosci*. 26, 507–513
37. Hampton, A.N. and O’Doherty, J.P. (2007) Decoding the neural substrates of reward-related decision making with functional MRI. *Proc. Natl. Acad. Sci. U. S. A.* 104, 1377–1382
38. Knutson, B. et al. (2001) Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* 12, 3683–3687
39. O’Doherty, J. et al. (2003) Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J. Neurosci*. 23, 7931–7939