Center on Capitalism and Society
Columbia University
Working Paper #103

15th Annual Conference
The Age of the Individual: 500 Years Ago Today
Session 5: Individualism in the Economy

Individualistic Welfare Analysis in the Age of Behavioral Science

Michael Woodford
Columbia University
October 31, 2017

The dominant approach within economics to the evaluation of potential interventions into market arrangements, at least since the work of Pareto, has been to ground welfare judgments on the degree to which an intervention is predicted to improve outcomes for individuals, from the standpoint of those individuals' own preferences, rather than what a social planner might think is good for people or for society collectively. It is further supposed that these individual preferences should be the ones that people express in their decisions in the marketplace, and that can be inferred from their observed behavior (the theory of "revealed preference").

However, developments in the scientific study of human behavior, especially since the middle of the 20th century, raise a challenge to the view of the individual that is presumed in welfare analysis of this kind. The theory supposes that people have well-defined and stable preferences, of which they are fully aware, and that they can be counted upon to consistently choose the outcome that they rank highest under those preferences, on every occasion when they are given a choice. Instead, many experiments suggest that people often make choices in different circumstances that cannot be rationalized in terms of *any* consistent preference ordering.

A well-known example is provided by the "decoy effects" sometimes employed by marketers: consumers' willingness to buy one good rather than another can be increased by adding a third good — the so-called "decoy" — to the set that are available for purchase, as documented for example by Huber, Payne and Puto. This is inconsistent with the notion that people choose the good that they rank highest, according to some preference ordering that is independent of the set of options on offer. The work of Lichtenstein and Slovic on "preference reversals" — demonstrating that people do not appear to assign the same relative valuation to two options (say, two different lottery tickets) when asked which of them they would choose as when asked to assign a dollar value to each of the tickets individually — provides another example.

Examples of this kind challenge the assumption that people have stable preferences, and know and consistently act upon them. Of course, to the extent that such context or framing effects are predictable, they pose no problem for the use of empirically grounded models of behavior to predict the effects of public policies. Indeed, an influential recent literature proposes to use documented behavioral anomalies to more effectively achieve goals such as increasing the degree to which people save for their retirement. But such results do raise a question about whether policy interventions can be justified on individualistic grounds.

Proponents of “libertarian paternalism” such as Cass Sunstein and Richard Thaler argue that there should be no objection to “nudges” designed to further some social objective, as long as there is no change in the set of options that continue to be available to individuals. I admit that this is a clever rhetorical defense of the policies — a proponent of classic rational choice theory cannot object to their proposed policies, because his intellectual presuppositions imply that the policies can have no effect. However, it remains the case that if the policies have any effect — and the evidence indicates that they do — then the question must arise as to who decides in which direction people’s decisions should be modified by the “nudges,” and on what grounds. Welfare judgments of *some sort* must be made. The question that I’d like to pose is, must we completely abandon individualism as a basis for such judgments?

The answer that I would like to propose is that it remains possible to preserve the idea that interventions should be justified by the degree to which individuals’ situations are improved, from the standpoint of individual values that are imputed to them. But rather than supposing that these individual values correspond to whatever people are observed to choose under any circumstances, we must define them as what one predicts that people would choose under an “ideal choice situation.” By this I mean a hypothetical situation in which the absence of information processing limitations would make it possible for a consistent and well-considered preference ordering to be expressed.

Such conditions are evidently often not satisfied in actual choices — this is why “nudges” are observed to be effective. Nonetheless we might be able to reach conclusions, based on how people actually behave, about how they would behave under such an ideal situation. The idea of basing normative analysis on such a counterfactual is in the spirit of John Rawls’ proposal to ground a normative theory of desirable social arrangements not on the contracts that people are willing to enter into with one another under their actual circumstances, but rather on what the analyst concludes that they would freely choose if choosing “behind a veil of ignorance.”

One might try to infer people's "true" preferences, while recognizing that observed choices are not always consistent with them, in either of two ways. One approach would be to compare the choices that people make under alternative circumstances, identifying one of these as a context in which preferences are expected to be expressed in an undistorted form. (Jacob Goldin calls contexts of this ideal type "revelatory frames.") One could then use the choices observed in the ideal circumstances to identify the preferences that would be used for purposes of welfare analysis, even in settings where different behavior is observed.

A recent paper by Hunt Allcott and Dmitry Taubinsky in the AER provides a good example of this method. Allcott and Taubinsky seek to assess the welfare gains from subsidizing the adoption of compact fluorescent light bulbs as an alternative to incandescent bulbs. They conduct an experiment in which a representative sample of US consumers make decisions about which kind of light bulbs to buy at various prices, with and without an "information treatment" that draws their attention to publicly available facts about the energy costs of using the two types of bulbs. The authors interpret the demand curve that they estimate under the "information treatment" as reflecting people's true preferences, when they are fully informed about the consequences of their choice, while the different demand curve that is estimated under the "no-information treatment" reflects *inattention*, rather than expressing consumers' true preferences. They find substantially greater willingness to purchase the CFL bulbs under the "information treatment," which they take to show that inattention to available information about energy costs seriously distorts the choices that people make in the marketplace.

Under the assumption that choices under the "information treatment" reflect true preferences, Allcott and Taubinsky are able to estimate the deadweight loss associated with different possible interventions in the lightbulb market, under the assumption that reducing people's inattention to facts about energy costs is not practical outside the experimental setting. They conclude that a modest subsidy to CFL bulbs would increase welfare, and compute the optimal subsidy. At the same time, their calculations indicate that an outright ban on using incandescent bulbs — the policy actually adopted by the US — should lower welfare. This result illustrates an important point, which is that merely establishing the existence of a behavioral bias in a certain direction is not sufficient to establish that a crudely designed restriction on behavior, such as an outright ban on the action that one thinks people mistakenly choose too often, will necessarily improve welfare. But it also shows how individualistic welfare analysis is still possible, while taking account of the behavioral bias, and demonstrates that the conclusions reached may well differ from those obtained if one were to assume that market behavior under laissez-faire always reflects people's true preferences.

Another, more ambitious approach would seek to understand the predictable ways in which different contexts are expected to elicit different choices, with the help of a theory according to which people's observed patterns of behavior are generally well-adapted to some class of decision problems that they might expect on average to face, though the heuristics that they necessarily use (owing to limited attention and information-processing capacity) will not produce an ideal decision on each occasion. Such a judgment that the heuristics used are relatively efficient (among those making no greater informational demands) will have to be relative to some identification of the person's implicit goals; hence a program of this kind, if successful, would allow the individual's goals to be identified from a study of her behavior, even though it is not assumed that every observed choice represents the highest-valued element of the choice set on that occasion, as in "revealed preference" theory. The preferences recovered in this way could then be used as a basis for a welfare analysis that could claim not to be substituting the social analyst's own views for the personal values of the individual, even if one never observes actual choices made under an "ideal choice situation."

As an example of a theory of this kind, in a recent paper, Xavier Gabaix and David Laibson propose a theory of intertemporal choice in which people's choices are biased by having to base their decisions on noisy mental simulations of the future outcomes that can be expected to result from the various options available to them. The model assumes that people's decisions are made, on the basis of their noisy mental simulations, according to a criterion that maximizes the expected value of the person's utility, conditional on the decision having to be some function of the noisy simulation, rather than being based on the true consequences of the decision. To the extent that such a model fits observed behavior, the true utility function of the decision maker can be inferred, even though observed choices will generally not be optimal from the standpoint of those preferences when the true consequences are taken into account.

In particular, Gabaix and Laibson show that their theory implies that people will discount future rewards relative to immediate rewards to a greater extent than their true intertemporal preferences would prescribe; in the Gabaix-Laibson view, people are not really so unconcerned with future rewards, they simply see them less distinctly as a result of the noise in their mental simulations. The theory therefore predicts that people will under-save, relative to the amount that would actually maximize their expected utility.

The Gabaix-Laibson theory further predicts that this problem cannot be solved simply by offering people the opportunity to commit themselves to “save more later,” as proposed by Richard Thaler and Shlomo Benartzi. The Thaler-Benartzi solution might seem sufficient, if one thought that people are in an ideal choice situation, and correctly express their true preferences, as long as they are making decisions only about actions to be taken far enough in the future for “present bias” not to distort their choices. But this is not what the Gabaix-Laibson theory implies. According to that theory, people’s intertemporal decisions are biased, but even with a correct understanding of the noise in their mental simulations, they have no motive to choose to constrain their future selves not to make the decisions that appear best at that time — instead, they should realize that their current ability to accurately assess consequences far in the future is poorer than it will be later.

Nonetheless, if validated, the Gabaix-Laibson theory would allow one to infer someone’s true rate of time preference from the choices that they make on the basis of noisy mental simulations. These inferences about true preferences could then justify “paternalistic” interventions — such as government saving on behalf of individuals — without this having to be based on a view of what is good for people that is not derived from observation of their own choices.

REFERENCES

Allcott, Hunt, and Dmitry Taubinsky, “Evaluating Behaviorally Motivated Policy: Evidence from the Lightbulb Market,” *American Economic Review* 105: 2501-2538 (2015).

Gabaix, Xavier, and David Laibson, “Myopia and Discounting,” NBER Working Paper no. 23254, March 2017.

Goldin, Jacob, “Which Way to Nudge? Uncovering Preferences in the Behavioral Age,” *Yale Law Journal* 125: 226-270 (2015).

Huber, Joel, John W. Payne, and Christopher Puto, “Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis,” *Journal of Consumer Research* 9: 90-98 (1982).

Lichtenstein, Sarah, and Paul Slovic, “Reversals of Preference Between Bids and Choices in Gambling Decisions,” *Journal of Experimental Psychology* 89: 46-55 (1971).

Rawls, John, *A Theory of Justice*, Cambridge, MA: Belknap Press, 1971.

Sunstein, Cass R., and Richard H. Thaler, “Libertarian Paternalism Is Not an Oxymoron,” *University of Chicago Law Review* 70: 1159-1202 (2003).

Thaler, Richard H., and Shlomo Benartzi, “Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving,” *Journal of Political Economy* 112: S164-S187 (2004).

Thaler, Richard H., and Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth and Happiness,* New Haven: Yale University Press, 2008.