Mistakes in Choice-Based Welfare Analysis

By Botond Kőszegi and Matthew Rabin*

Economics has always been concerned not only with describing or predicting economic behavior but also with understanding economic well-being. The traditional, official way economists (claimed to) have assessed well-being is from "revealed preference"—observing what people choose under the maintained hypothesis of 100 percent rationality. For instance, when we teach millions of students each year the conditions under which interfering with free-market exchange will make people worse off, by interfering with satisfying their wants, we do so under the compelling assumption that people tend to make choices that rationally maximize their own well-being. While economists are humble about the fact that we are not in a privileged position to declare what goals society should pursue, welfare economics has provided guidance on the determinants of well-being according to this restrictive set of criteria. Although all of us also assess well-being in other ways, it is only recently that economists have begun to do so in a more focused way. A growing number of researchers have begun to study alternatives to the 100 percent rationality assumption, while other researchers have used sundry techniques to measure well-being directly. In this article, we explore some conceptual issues as to why and how one might use these new assumptions and approaches to supplement and modify the revealed-preference approach as it is conventionally conceived.

We take the central question of welfare economics to be: how do different situations or economic environments affect people's well-being? When doing so with sensible ancillary assumptions, inferring people's well-being from their choices, based on the presumption that they are rationally pursuing their goals, is, in our view, the best scientific approach to research on well-being yet formulated. Our first goal in this paper is to clarify just how crucial these ancillary assumptions are to rational-choice welfare analysis. Whether unnoticed or unemphasized, assumptions that are unobservable in choice behavior drive *all* welfare conclusions in economics. Any combination of observed behavior and assertions about what environments enhance well-being is consistent with utility maximization. The basic logic, formalized as a mathematically trivial theorem in Kőszegi and Rabin (2007), is simple. As is clear from psychology, and recently has been better appreciated and elegantly modeled within economics, well-being may depend not only on the outcome resulting from choice, but also on the choice set itself. The effect of different choice sets on well-being is not observable by the choices made within each choice set. Indeed, we discuss examples below where choice-set dependence is so fundamental a component of preferences and the ancillary assumptions sufficiently nonobvious as to make new methods to measure well-being crucial.

Hence, even assuming a priori that people are fully rational, the question of whether economists should consider new types of evidence on well-being is *not* the question of whether economists should make assessments that go beyond what choice behavior tells us. Rather, the question is about how we select choice-unobservable assumptions and how valid they are. Whether conclusions are reached by smart psychologists observing people's smiles, smart neuroscientists observing fMRI data, smart empiricists observing happiness survey data, or smart economic theorists introspecting about which assumptions about well-being seem attractive, these conclusions are reached based on something beyond observed choices and the rationality hypothesis.

Of course, the reasonable interpretation of much evidence is that the rationality assumption may itself be wrong enough to warrant welfare analysis that allows for the possibility that people make mistakes. More than pointing out the need for ancillary assumptions to do revealed-preference welfare analysis with the

*Kőszegi: Department of Economics, University of California, Berkeley, 549 Evans Hall #3880, Berkeley, CA 94720 (e-mail: botond@econ.berkeley.edu); Rabin: Department of Economics, University of California, Berkeley, 549 Evans Hall #3880, Berkeley, CA 94720 (e-mail: rabin@econ.berkeley.edu).

100 percent rationality assumption, our more positive observation below is that, with such ancillary assumptions, revealed preference can be used to simultaneously infer what people's preferences are *and* the ways they sometimes fail to maximize those preferences. Hence, while we believe the case is clear for moving forward with nonchoice evidence of well-being, following our fleshing out of the nonexistence of exclusively choice-based welfare analysis in Section I, we will argue that choice can be used to understand a person's mistakes as well as her preferences.

I. Choice Behavior Alone Cannot Reveal Welfare

One reason choice behavior may not reveal well-being is that we do not observe enough choices. The main motivation among economists for the growing literature on happiness does not seem to be based on doubts that people are fully rational, but rather on the fact that observable choice behavior is not rich enough to use revealed preference. While such practical limits to revealed preference are surely the best reason to explore alternative ways of assessing well-being, we turn to our observation that, even if an arbitrarily large amount of data *were* available, choice evidence alone cannot provide guidance on welfare. An extreme example illustrates the point. A person's choice behavior can never reveal whether she would find it best to have painful early death as her only possible option, rather than (say) being able to avoid death and eat cake instead. Suppose that for any decision problem the person is facing—including arbitrarily complicated, dynamic decision problems—her preferences are over the set of final outcomes available and the outcome ultimately chosen. If her utility satisfies $u(\text{death}|\{\text{death}\}) > u(\text{cake}|\{\text{cake}, \text{death}\}) > u(\text{death}|\{\text{cake}, \text{death}\})$, she will choose cake whenever given the opportunity, but her utility is higher when death is unavoidable. Note that asking the person to make a decision between choice sets rather than final outcomes, and observing that she chooses {death, cake} over {death}, does not mean she would not prefer death imposed on her, since, by assumption, her preference is to have no mechanism for avoiding death.

There is, of course, a simple way around agnosticism about whether unavoidable painful death makes a person happy—use common sense and assume that it does not. Or, for this and other situations, one could revert to the older choice-unobservable psychological assumption of economics that well-being is independent of choice sets.

This latter response is, however, not generally adequate. First, in light of the ample evidence that behavior is choice-set dependent, choice-set dependence is necessary to maintain the rational-choice approach. Indeed, recent research by Faruk Gul and Wolfgang Pesendorfer (2001) has emphasized that this is one way that some observed violations of traditional choice axioms can be reconciled with rational choice. Somebody could choose candy from the choice set {candy, apple} but be better off with the choice set {apple} if she has choice-set-dependent preferences, where $u(apple|\{apple\}) > u(candy|\{candy, apple\}) > u(apple|\{candy, apple\})$. This has the natural interpretation that the option of eating candy creates an unpleasant sensation of temptation.

Gul and Pesendorfer (2001) are still able to reach strong welfare conclusions, however, based on choice-unobservable welfare assumptions on an expanded choice domain. They assume that if we observe somebody choose the choice set {*candy*, *apple*} over {*apple*} then we know that the person cannot be better off having only the option apple. Yet this rules out the possibility that a person may be happier not having access to candy yet never costlessly avoiding it, because doing so is an admission that she is too weak to resist temptation. She would give herself the option {*candy, apple*} rather than {*apple*} in any choice procedure but be better off not having the option to do so. This rational model yields the same choice behavior as somebody without temptation disutility but with very different welfare implications.

There are many other instances of choice-set–dependent preferences where the ancillary assumptions needed to infer well-being from choice alone are quite substantial and unresolved. Suppose, for instance, we observe a person who always chooses to share with others if she can. Related to psychological investigations on the nature of altruism and guilt, all consistent with rational choice, it could be that the ability

to share makes her happy relative to not having the option to give. Or it could be that she gives only because she would feel guilty otherwise, and would be happier without an option to share. Similar issues arise for other forms of social preferences. A person may dislike doing worse than those around her, but never choose to rectify this because she would feel worse about hurting others. Whether she feels envy or she is happy when others do well would not be observable in choice.

II. Revealing Preferences and Revealing Mistakes

More importantly and more constructively than observing the necessity for rational-choice welfare economics of ancillary choice-unobservable assumptions, the rest of our paper argues that *with* reasonable ancillary assumptions, choice behavior can be a powerful tool in revealing preferences, even when extreme rationality is abandoned as an a priori assumption. In fact, reasonable and useful inferences about people's preferences often can be made by, and only by, recognizing some of the mistakes people make. Preferences often can be *revealed* by behavior, even when they are not *implemented* by behavior.

While discussed in more detail in Kőszegi and Rabin (2006), we outline a general approach and an example of how preferences can better be revealed by acknowledging mistakes. The first step is to find a setting where the nature of some state-contingent preferences is obvious, so that observable choices reveal beliefs about the likelihood of those states, including any systematic mistakes. We can then use the revealed mistakes in beliefs, rather than rational expectations, to interpret what preferences are in situations where those are less obvious.

This procedure has its clearest and least controversial power in revealing errors about objective facts in the world, such as non-Bayesian statistical reasoning. Although such errors are likely to affect investment behavior and other important economic decisions, we use a simple and contrived illustration. First, assume that a person's preferences for money are independent of coin flips. Let (x, y) represent a lottery that pays $\$x$ if the next flip of a coin comes up heads (H) and $\$y$ if the next flip comes up tails (T). Suppose we observe the following choices:

- If the person observes that the previous flips come up HHH, she chooses $(85, 120)$ over $(120, 90)$ on the next flip. Notice that she chooses the pair that has lower stakes overall but pays more if the next flip is T, amounting to a bet that T will come up next.

- If instead she observes TTT, she chooses $(120, 85)$ over $(90, 120)$ on the next flip. This amounts to a bet that H will come up next.

- If she has observed no flips, she chooses $(90, 120)$ over $(120, 85)$ and $(120, 90)$ over $(85, 120)$ on the next flip.

These choices suggest a specific pattern of mistakes, the gambler's fallacy—the person believes that if the same realization of the random binary process has occurred a number of times, the other realization is "due." From recognizing a person's tendency to make this mistake, we can infer her preferences when they are less clear. Suppose, for instance, that the person prefers (4 apples, 4 oranges) to (5 oranges, 5 apples) after flips HHH; prefers (4 oranges, 4 apples) to (5 apples, 5 oranges) after TTT; and prefers 5-fruit gambles to 4-fruit gambles if she has observed no flips. Having interpreted her earlier behavior as belief in the gambler's fallacy, we can conclude from this behavior that she likes oranges more than apples, and will choose oranges in a nonrandom situation. Her preferences are revealed by behavior, even though they are not implemented by behavior.

Understanding the person's mistakes also allows us to analyze welfare. After observing flips HHH of coin A, for instance, would the person be better off with the option to choose between gambles $(120, 90)$ and $(85, 120)$ based on coin A, or the option to choose between gambles $(120, 89)$ and $(84, 120)$ on a new coin B? Because she mistakenly chooses the dominated bet on coin A but the favorable bet on coin B, she would be better off with the coin B choice set. Moreover, she may be better off with that choice set than being able to choose between the two choice sets—she may mistakenly choose to bet on coin A both because it is for more money

and because the gambler's fallacy leads her to think coin A is more predictable.

III. Replicator Dynamics

As an implication of our arguments in Section I, of course, the behavior and welfare conclusions previously shown can be replicated in a fully rational model—no combination of behavior and welfare conclusions from nonrational theory can possibly be inconsistent with rationality. While we trust this realization will eventually dampen economists' enthusiasm for finding rational-choice explanations for everything, we suspect that in the short run this principle will continue to entice economists to react to new psychologically motivated theories of mistakes by formulating new rational-choice models that replicate the behavioral predictions. And depending on which of the unlimited range of choice-unobservable psychological assumptions is chosen for the "replicator model," the welfare conclusions of the mistakes-based model can be either replicated or reversed.

It is hard to say when and how such endeavors are valuable.[1] In practical terms, a (sufficiently tractable) rational-choice model should clearly be used when it provides a psychologically more reasonable, and hence more likely to be generalizable, account of some particular phenomenon—including the many instances where economists are right to attribute rational motives to behavior others deem a mistake. When a utility-maximization reframing does damage to both psychological reality and to applicability, however, insisting on the reframing is clearly an impediment to scientific progress. If one is so inclined, one could, for instance, replicate all the gambler's-fallacy-based predictions above while maintaining the assumption of rational utility maximization. To mimic the behavioral predictions, one can assume that the person likes betting on H after TTT, on T after HHH, and has no preference among her bets if she has observed no flips. To mimic the welfare conclusion, one can make the even odder assumption that it is better for the person to bet if she has observed no flips than if she has observed HHH. Because the mistakes-based theory provides general and ex ante (rather than ex post) guidance on these questions, it is better economics.

In other cases, whether behavior reflects mistakes versus rational choice is less clear. Besides the fully rational explanation given above for why somebody might be better off with smaller choice sets despite never choosing to restrict herself (because she finds such self-binding unpleasant), the same behavior-welfare combination would arise if a person naively predicts she will not be bothered and will not yield to temptation. These two theories, irrational naiveté about self-control problems versus fully rational commitment aversion to controlling oneself, are not distinguishable in simple settings. Or (to take an example from a major theme in happiness research) people may be less happy when sacrificing local status by moving into wealthier neighborhoods and yet move into such neighborhoods for one of two reasons: they know it will happen but would be even more bothered by letting their behavior be guided by envy; or they may mistakenly believe they will continue to assess their status from their old neighborhood rather than from the new one.

IV. Moving Forward

We suspect many economists believe that some people (perhaps friends, relatives, or students) make statistical errors such as the gambler's fallacy, and most economists agree that *if* there is evidence of such mistakes in domains of importance, it should be studied by economists as mistakes. Other errors posited by psychologically oriented researchers are likely to prove more controversial. People may, for instance, systematically mispredict their own future preferences. Sixteen-year-olds not addicted to tobacco may underestimate the effect of addiction on their future preferences and behavior and hence mistakenly be prone to becoming addicted. Such an underappreciation of the power of addiction

[1] Indeed, it may be philosophically difficult to say what it means to assume that somebody is making a mistake. If we observe somebody choosing x from the choice set $\{x, y\}$ and want to assess whether she is better off with this choice set than (say) having only option $\{y\}$, then, either by assumption or with measurement, we can compare her well-being when choosing x given $\{x, y\}$ versus the choice y given $\{y\}$. But given that she *does* choose x, whether the person would be happier had she chosen y given $\{x, y\}$ than x given $\{x, y\}$ seems neither observable (with any data) nor, for the purposes of welfare economics as we see it, terribly important.

is in fact consistent with a growing amount of research. Louis A. Giordano et al. (2002), for instance, show that even experienced addicts do not fully appreciate the strength of cravings when not currently experiencing those cravings. Using real money and real drugs, they elicited monetary valuations for a dose of the heroin substitute buprenorphine at a given future state and time from both currently satiated and currently deprived heroin addicts. Although all subjects were long-time addicts choosing for the same familiar future situation independent of their current state, those who were currently satiated paid significantly less for the doses than those currently deprived. If not-yet-addicted 16-year-olds might similarly underappreciate the power of craving, it seems studying whether tobacco addiction is a mistake should warrant attention as a topic of welfare economics.

While maybe too many economists treat such systematic errors as somehow implausible, surely most economists abandon their devotion to extreme rationality assumptions when youth (and tobacco) are involved. And we doubt economists really think that youth have a monopoly on doing things that might not maximize well-being. We suspect some believe that many investors mistakenly over-trade based on misconceived theories of market patterns, and some economists may even acknowledge that the prevalence of expensive consumer debt may involve an important departure from 100 percent rationality.

Yet many economists who acknowledge the possibility of such errors seem to deem it as outside their purview as economists to investigate the welfare implications of these errors. This seems to us a mistake—based on an odd combination of arrogance and humility. A common intuition and worry is that the evolving new "mistakes-are-possible" and "nonchoice-data-are-permissable" welfare economics will imply a new arrogance by economists in telling people what makes them happy. We think, on the contrary, that the progress will lead not only to a better science of well-being, but also to a humbler and less preachy one as well. Maintaining the status quo of teaching students, citizens, and policymakers what institutions and incentive structures are efficient given only one particular notion of human nature and only one source of data, but demurring from such analysis when assuming the types of mistakes and using the types of measures that many lay people and other social scientists are concerned with, is not modesty. Nor is the heavy paternalism implicit in refusing such analyses out of fear that the expanded analysis will invite the types of government interventions we disapprove of.

By contrast, the tempting alternative of advocating that economists abandon their fundamental and pervasive concern with welfare altogether strikes us as involving a strange pessimism about the power of the discipline. It is clear that the powerful theoretical and empirical methods, insights, and assumptions of economics can be fruitfully applied to, and combined with, new assumptions and methods. For instance, because inferences along the lines of Giordano et al. (2002) and the procedure used for the gambler's fallacy could be applied to field data on addiction, economists are in a unique position to study when becoming addicted is rational and when it is a mistake. Abandoning one of the central tasks of economics for fear that our discipline is not up to the task seems unwarranted. Panicked predictions of loss of discipline and scientific decline when status quo assumptions and methods are modified and expanded have proven ill-advised in many previous cases of innovation in economics, and it seems clear to us that such reactions are a mistake in this case as well.

REFERENCES

Giordano, Louis A., Warren K. Bickel, George Loewenstein, Eric A. Jacobs, Lisa Marsch, and Gary J. Badger. 2002. "Mild Opioid Deprivation Increases the Degree That the Opioid-Dependent Outpatients Discount Delayed Heroin and Money." *Psychopharmacology*, 163(2): 174–82.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica*, 69(6): 1403–35.

Kőszegi, Botond, and Matthew Rabin. 2006. "Revealed Preferences and Revealed Mistakes." Unpublished.

Kőszegi, Botond, and Matthew Rabin. 2007. "Choice and Happiness." Unpublished.