Journal of Public Economics 92 (2008) 1821–1832

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/econbase

Choices, situations, and happiness

Botond Kőszegi, Matthew Rabin *

Department of Economics, University of California, Berkeley, United States

ARTICLE INFO

Article history:
Received 6 November 2007
Received in revised form 17 March 2008
Accepted 27 March 2008
Available online 7 April 2008

Keywords:
Happiness
Revealed preference

ABSTRACT

This article explores some conceptual issues in the study of well-being using the traditional economic approach of inferring preferences solely from choice behavior. We argue that choice behavior alone can never reveal which situations make people better off, even with unlimited data and under the maintained hypothesis of 100% rational choice. Ancillary assumptions or additional forms of data such as happiness measures are always needed. With such ancillary assumptions and additional data, however, the use of revealed preference to study well-being can be significantly improved, so that the choices people make can jointly identify preferences, mistakes, and well-being.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Happiness is a very good thing. And it is a very good topic of social–scientific investigation. Indeed, beyond the study of behavior, a central focus of Economics has always been to understand how economic behavior and institutions affect well-being.

The most traditional technique within Economics for assessing the well-being in different situations is what might (aggressively but accurately) be called "unquestioned-rationality revealed preference": observed behavior is assumed to reflect fully rational maximization of utility, and this leads economists to infer that welfare is higher in one situation than another if it lets a person attain the outcome she seems most inclined to choose. Yet other approaches to studying well-being have been used outside and (especially recently) inside economics. In this article we clarify some assumptions underlying the use of the traditional choice-based approach to the study of well-being, and make the case for supplementing and combining it with some of these other approaches. We do so in light of three strands of research: 1) theoretical models of how utility may depend not just on outcomes of choice, but on the context in which those outcomes were chosen; 2) psychological research showing ways in which people are less than 100% rational; and 3) the growing literature that investigates the determinants of well-being by measuring happiness in different situations.

When doing so with sensible ancillary assumptions, inferring people's well-being based on the presumption that observed choices are rational is in our view the best scientific program for studying well-being yet formulated. Yet in Section 2 we clarify just how crucial these ancillary assumptions are to rational-choice welfare analysis. Whether unnoticed or merely 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 below, is trivial: as is clear from psychology, and has been recently been better appreciated and elegantly modeled within economics, well-being may depend not just on the outcome resulting from choice, but on the choice set itself. Yet the effect of different choice sets on well-being is not observable by the choices taken within each choice set. If, for instance, a person derives satisfaction from having resisted a temptation imposed on her but not one which she imposes on herself, we cannot know that from the fact that she never imposes a temptation on herself; and if a person dislikes being allocated less money than somebody else but dislikes even more intensely hurting others, we will never observe in her behavior that she dislikes coming out behind. We also formalize a straightforward logic about why there is unlikely to be clever elicitation techniques involving choices over choice

* Corresponding author.
E-mail address: rabin@econ.berkeley.edu (M. Rabin).

0047-2727/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpubeco.2008.03.010

sets, etc., to identify preferences like these: if somebody dislikes hurting others to avoid coming out behind, asking her whether she wants to let others to impose the harm makes her do the harm herself. While the ancillary assumptions needed beyond observing behavior and assuming rationality are often innocuous, we provide in Section 2 examples where choice-set dependence is so fundamental a component of preferences and the ancillary assumptions sufficiently non-obvious as to make new methods to measure well-being crucial.

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 clarifying the need for ancillary assumptions to do revealed-preference welfare analysis even with the assumption of 100% rationality, in Section 3 we briefly review some of the arguments in Kszegi and Rabin (2008) about how, *with* such ancillary assumptions, revealed preference can be used to simultaneously infer what people's preferences are *and* the ways that they sometimes fail to maximize those preferences. These combined inferences can, in turn, be used to assess welfare across different situations. If somebody reveals by her behavior that she makes errors in statistical reasoning, then we can infer her preferences based on her apparent beliefs, rather than try to construct the potentially incoherent preferences that we'd infer if we attributed Bayesian reasoning to her. And then we can judge her well-being in different environments by how, according to her revealed preferences, she assesses the outcomes her behavior induces given her combination of preferences and errors.

Hence, by liberating the insight that choices often reveal much about people's preferences from the debilitating tautology that everything people do maximizes their utility, economists can extend the power of revealed preference to the many cases where unquestioned-rationality seems a problematic maintained hypothesis. In this light, if Section 2 suggests that economics has (implicitly but ubiquitously) exaggerated the degree to which our welfare conclusions are neutral reflections of people's own choices, the arguments in Section 3 suggest that economists have also been significantly underestimating how powerful a tool revealed preference can be to reach welfare conclusions outside the extreme rational-choice framework — when combined with the same type of ancillary assumptions that must be invoked to reach welfare conclusions within the rational-choice framework.

While the formalizations of Sections 2 and 3 delineate the meaning of and implications of rationality within our framework, and as a matter of psychological interpretation and practical productivity it is certainly useful to distinguish between partially mistaken vs. fully rational behavior, Section 4 fleshes out both the conceptual precariousness and limited urgency in categorizing choices as mistaken or not. Namely, if one is interested solely in mapping situations to choices and welfare, then in principle it may not matter whether something is a mistake or not. If a person always chooses x from the choice set $\{x, y\}$, and yet we observe that she is happier with the forced choice $\{y\}$, then we have observed a primary object of concern — which choice set makes her happier. If we never observe her choosing y from $\{x, y\}$, without ancillary assumptions it may be hard to tell whether she "would have" been happier had she done so, or instead that she is fully rational and has choice-set-dependent preferences.

More broadly, the arguments of Sections 2 and 4 clarify the role that non-choice measures of well-being should play in light of accumulating evidence and theory about choice is influenced by choice sets, framing, and other aspects of how choices are elicited that have traditionally been downplayed in economics. Tversky and Kahneman (1981) and many subsequent researchers have found that the framing of questions affects choices; many experiments have found "preference reversals" showing that people seem to rank lotteries differentially depending on whether they are asked to choose or if they are to state prices for the lotteries (see, e.g., Tversky and Thaler (1990), and — more directly related to choice-set effects — a massive literature in the fields of psychology and marketing show various ways that the set of options available to people influence their preference ordering among those options (see, e.g., Huber et al., 1982 and Simonson and Tversky, 1992). Once we recognize such context effects, the joint hypothesis of full rationality and context-free preferences must be rejected, and it is clear that revealed preference as classically understood in economics cannot be sufficient to identify preferences.

We are unfamiliar with any formal treatment or explicit arguments along the lines we make in this article, and are unfamiliar with many papers in economics that makes salient how the welfare results reached in the paper rely on more than revealed preference and rationality. Yet many of the arguments in this paper are likely to be pretty obvious once we relax the assumptions of rationality *or* of the context-independence of preferences. As such, we see this article as organizing some of the principles many researchers are beginning to recognize and as aiding those researchers less familiar with these new lines of reasoning. In the process, we also believe a simple clarification of the nature of welfare economics that inheres in our approach is itself useful. Namely, we take the central question of welfare economics to be what might be called "situational comparative statics" for welfare: how do different situations or economic environments affect people's well-being? While this is not the only question one could ask, we do see it as an abstraction of the essential question that economists are often interested in.

This more direct focus on the choice set (or, more broadly, the choice context) as the primitive object of welfare economics would seem to help focus the task of welfare economics, and makes our observations about the limits of revealed preference all the more obvious. Indeed, our emphasis on the relationship between choice sets, choices from those choice sets, and implied well-being also has implications for the existing research program searching for direct evidence of the determinants of happiness. In Section 5, in fact, we turn to a brief treatment of how the framework we propose suggests that more systematic care should be taken in happiness research to recognize the role of both choice behavior and the possible influence of choice sets.

This article does not substantially address any of three central questions. First, how does one measure happiness or well-being?[1] Second, how does one make judgments about social outcomes when interpersonal comparisons of well-being are involved? Third,

[1] In fact, this article concentrates solely on the outlined conceptual issues associated with the study of happiness rather than substantial findings of the literature or the many different ways to measure well being. For some good introductions to this literature, see Kahneman et al. (1999), Frey and Stutzer (2002), and Layard (2005).

should economists be concerned with well-being? In Section 6 we briefly discuss these questions, as well as some other issues that economists are likely to confront in making progress in this tradition of studying well-being in the face of increasing awareness that our current methods are not as sufficient as we thought and that alternative approaches can provide insight.

Despite conceptual problems with the use of choice behavior as the sole criterion to reach judgments of well-being, in many cases it seems clear that both rationality and choice-set independence of preferences are good enough approximations that in fact familiar approaches are quite sufficient. Indeed, many of our examples below are clearly 'boundary cases' which clarify that (even when rationality goes unquestioned) all existing welfare conclusions in economics rely on choice–unobservable assumptions about what makes people happy, rather than raise substantive challenges to particular conclusions economists have reached. In our view, revealed preference is too powerful a tool for studying well-being, and the ancillary assumptions needed to render the tool effective are often too minimal and reasonable, to fret much about the conclusions economists are reaching except in cases where there are specific reasons to doubt these assumptions. We do, however, think the fact that unlimited choice data and unquestioned-rationality are not sufficient to reach any conclusions about well-being is an important thing for economists to recognize explicitly. Besides the many cases where the insufficiency is in fact real, we believe the open recognition that economics universally uses choice–unobservable psychological assumptions about well-being to reach all of our welfare conclusions can be an antidote to methodologically motivated aversion to greater scientific focus on these psychological assumptions and alternative ways of measuring well-being. This, in turn, will facilitate our discipline moving forward towards improved research — inside and outside the rational-choice framework — that combines choice data with other observable data pertaining to well-being.

2. Rationality, choice sets, and happiness

To see the problem with using choice behavior alone to determine well-being, we begin with an extreme example, and observe: 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 also being able to avoid death and (say) eat cake instead. Suppose that for any decision problem the person is facing — including arbitrarily complicated, dynamic decision problems possibly involving decisions to restrict her future options in various ways — her preferences are over the set of final outcomes available and the outcome ultimately chosen. If her utility satisfies $u(\text{death}|\{\text{death}\}) > u(\text{cake}|\{\text{death, cake}\}) > u(\text{death}|\{\text{cake, death}\})$, 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 over choice sets rather than final outcomes, and observing that she chooses {cake, death} over {death} in this meta-choice, does not mean she would not like to have 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 assumption of economics that well-being is independent of choice sets. We could simultaneously assume that people maximize utility and assume that utility does not depend on choice sets. Under these assumptions, we know which situations are better, equal, or worse than other situations by the traditional revealed-preferences approach. With this pair of (choice–unobservable) psychological assumptions, economists have for a very long time reached strong welfare conclusions based on the rational-choice model. In cases like this — and in many far more interesting and important examples economists have studied over the years — the discipline's approach has been more than adequate.

Research has shown, however, that in many situations this approach is problematic. In light of the ample evidence that behavior is importantly choice-set dependent, choice-set dependence of preferences is necessary to maintain the rational-choice approach. Choices over outcomes in many contexts depend on choice sets and other features of situations. Indeed, earlier arguments and examples by Sen (1993) and recent axiomatic research by Gul and Pesendorfer (2001) and Dekel et al. (2007) have emphasized that some observed behavior that violates 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 $u(\text{apple}|\{\text{apple}\}) > u(\text{candy}|\{\text{candy, apple}\}) > u(\text{apple}|\{\text{candy, apple}\})$. This has the natural interpretation that having the option of candy creates an unpleasant sensation of temptation. It seems in conflict with traditional conceptions of consistent choice, since somebody forced to choose from the choice set {candy, apple} is making a different choice than somebody involved in a meta-choice of committing herself to either {candy} or {apple}.

Unlike Sen's (1993) more general framework, recent literature building from Gul and Pesendorfer (2001) preserves the primacy of traditional decision-theory axiomatic restrictions when studying temptation by imposing those restrictions on the higher-order domain of choices over choice sets. In doing so, they develop a utility representation of such temptation preferences that embeds the traditional choice–unobservable welfare assumptions needed to reach conclusions. In their framework, if somebody chooses the choice set {candy, apple} over either {candy} or {apple}, then the person is represented as having higher utility with the choice set than with only the option apple. This assumption rules out preferences such as an aversion to committing oneself because it makes one feel like a weakling; with such preferences, a person may be happier not having access to candy yet never costlessly eliminate it from a larger choice set. 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 with the more straightforward (and more familiar to economists) preference for candy over apples without temptation disutility.

Whether or not they are prevalent, such preferences have a simple psychological intuition: a person might feel humiliated and weak if he ever eliminates an option in his choice set solely to prevent himself from choosing it, and will never do so directly or indirectly — even if (and even if he is aware) having such a restriction imposed exogenously would make him happier. It could be that a person knows that life on a fat farm or a "non-smoking" farm will be pleasant, healthy, temptation-free and a good place to

live — but feel miserable about having made the decision to move to the farm. With or without such psychological interpretations, however, there is clearly a fully coherent and fully rational set of preferences that is choice-set independent but whose welfare is choice-set dependent. Less plausibly, but also logically possible, it could be that a person employs certain types of commitment devices whenever they become available, but would be happier not to have the opportunity to do so. So whether or not people choose commitment devices when given the opportunity to do so does not, from rationality alone, tell us whether they are happy having the commitment devices themselves.

A natural reaction to the above possibility is to take one more step back, and consider choice sets over choice sets as the primitives, imposing choice-set independence on these primitives. Indeed, one might be tempted to argue that once we allow for the seemingly innocuous assumption that choice-set dependence does not feature infinite regress, we can overcome the inability of choices to reveal welfare. But this argument would muddle the point. Our cake/death examples shows exactly that some conceptually very simple preferences *do* induce choice-set dependence at every level, even when a person chooses over arbitrary choice sets over choice sets, etc. Indeed, we argue below that the ancillary psychological assumption needed to find a set of choice sets that can distinguish among preferences imposes are in fact likely to be quite stringent and unnatural.

Before further consideration of these and other (more realistic) examples of difficult-to-identify preferences, we set up a formal framework. Consider a given person choosing from some set of options, X. When trying to relate this to empirical work below, we'll consider a distribution of different choices and measures of happiness across a population of people all choosing from a choice set; but here we present analysis in terms of a single person. And although of course both choices and happiness might be influenced by environmental contextual features or (as has been repeatedly demonstrated) even by how logically identical choices are framed, here we abstract from those; we assume that happiness depends solely on the choice set and outcome. We conjecture that the main points we are making extend straightforwardly to such enriching of the framework. Indeed, some of the main points — about the difficulty of using choice alone to infer welfare — are more obvious and classically recognized when one considers such factors.

Let the real number $H(X)$ be the happiness or well-being when the person faces choice set X. We take the maximization of this to be a primary goal of policy analysis, and hence inferring these values to be the task of welfare economics. Let $c(X) \in X$ be the person's observed choice from X, which we will assume for ease of presentation to be unique and deterministic.[2] Let $h(x|X)$ be how happy the person is when she makes the choice $x \in X$. By construction, $H(X) \equiv h(c(X)|X)$.

With this framework, it is clear how to formalize the rationality hypothesis that is traditionally made in economics. Again noting that we are being loose here and throughout the article by ignoring the possibility of indifference among two or more choices, the first assumption that underlies the unquestioned-rationality revealed-preference approach is:

$$\text{Rationality: } c(X) = \arg\max_{x \in X} h(c(X)|X).$$

This hypothesis says that if we observe somebody choosing candy from the set of {candy, apple}, then she is happier with candy from that set than she is with an apple from that set. That is, $c(\{\text{candy, apple}\}) = \text{candy}$ implies that $h(\text{candy}|\{\text{candy, apple}\}) > h(\text{apple}|\{\text{candy, apple}\})$.

But economics has traditionally made a further assumption in addition to rationality: that a person's satisfaction with an outcome is independent of the choice set from which she chooses that outcome. That is, it is typically assumed that the happiness with an outcome doesn't depend on the choice set from which it is chosen:

$$\text{Menu Independence of Welfare}: h(x|X) = h(x|Y) \text{ for all } X, Y, x \in X \cap Y.$$

Under the maintained hypothesis that people choose rationally, menu independence of welfare in turn implies the classical "independence from irrelevant alternatives" axiom of choice, which we state in a particular form as "menu independence of choice":

$$\text{Menu Independence of Choice}: \{c(X), c(Y)\} \subset X \cap Y \text{ implies } c(X) = c(Y).$$

The connection between the two menu-independence assumptions is clear: if 1) people maximize their happiness and 2) their happiness with an outcome doesn't depend on their choice set, then the set of options in a choice set cannot change which of two particular options is chosen. These menu-independence assumptions are essentially what allow the core structure of analysis of positive and normative economic theory. "How does changing a person's constraints influence her choices and her well-being?" is almost always asked with the point of view that utility is independent of these constraints. This constraint-independence is inherent in the traditional revealed-preference approach to welfare.

Yet new approaches and insights have led to challenges to the focus on menu independence. There is convincing evidence from psychology on both the influence of choice sets on preference orderings, and — more fundamentally challenging — the influence of the framing or presentation of the same choice set. Much of this literature is agnostic about whether "preference reversals" necessarily involve mistakes, and is essentially consistent with the rational-utility-maximization perspective; in other contexts, there is a clear presumption that mistakes are involved. Nevertheless, assuming that the happiness with different alternatives is choice-set dependent is the only way to reconcile the evidence with rationality.

We now turn to clarifying the central non-identifiability principle: if we allow for preferences to be choice-set-dependent, arbitrarily rich choice data interpreted under the maintained hypothesis of rationality tells us nothing about what situations

[2] Of course X itself can contain stochastic lotteries, or randomization over some more primitive set, in which case $c(X)$ will be a lottery.

are better for people. To formally state the result, let Ξ be some proposed set of possible choice sets X a person might face, and $c(\Xi)=\{c(X)\}_{X\in\Xi}$ observed (or observable) choices from all of these choice sets. We put no restrictions on Ξ; it can be anything, and can include all feasible meta-choices, etc.

Choices alone cannot reveal preferences: For all Ξ, for all $c(\Xi)$, for all $Z\in\Xi$, there exists h consistent with $c(X)=\arg\max_{x\in X} h(x|X)$ for all $X\in\Xi$ such that $H(Z)>H(Y)$ for all $Y\in\Xi\backslash Z$.

The theorem says that any observed pattern of behavior is consistent with any situation being the one that makes a person happiest. Once stated, it is trivial. Nothing we can observe about behavior within the choice sets Y and Z (or any other choice sets) tell us anything about which choice set is better, because it only tells us about the relative happiness of different choices within a choice set. Note again that we have not put any restrictions on Ξ, so that each $X\in\Xi$ could be a set of choice sets, a set of choice sets over choice sets, etc. Hence, to say these more complicated objects as the domain of choice does not solve the fundamental impossibility of eliciting well-being from choice.

In fact, to formalize and generalize our point above that no choice procedure can elicit welfare in our cake–death and commitment-aversion examples, consider a particular class of preferences. Suppose that there are some set of outcomes (best conceived of as "physical" outcomes, such as what you end up eating), Q, over which a person has choice-set dependent preferences. But suppose further that she is implementing one of these outcomes in some more complicated decision-making procedure, where Ξ is the set of possible procedures. As such, for any such set there will a "physical-outcome mapping", $q(\cdot)$, that, for all $X\in\Xi$ and outcomes $x\in X$ assigns an element from a set of physical outcomes Q. That is, let q define $q(x,X)\in Q$ for all $X\in\Xi$, $x\in X$, and let the set of physical outcomes associated with a particular choice set a person faces be $Q(X)=\cup_{x\in X} q(x,X)$. Say that Ξ implements Q if for all subsets $Z\subseteq Q$ there is $X\in\Xi$ where $Q(X)=Z$. Now consider preferences that depend solely on $q(x)$ and $Q(X)$:

Procedure-independent preferences: In an environment Ξ that implements Q, the happiness function h is a procedure-independent function of Q if for all X, $Y\in\Xi$ and outcomes $x\in X$, $y\in Y$ such that $q(x)=q(y)$ and $Q(X)=Q(Y)$, $h(x|X)=h(y|Y)$.

Procedure-independence means here that a person's well-being depends solely on what options in Q she could implement and which one she does implement. It is 'procedure-independent' in the sense that it relies on her de facto constraints and choices not how she implements. Such preferences rule out, for instance, a person feeling differently about first choosing the choice set {apple} over {candy, apple} and then "choosing" apple from {apple}, versus just choosing apple from {apple, candy} directly.

Conceiving of preferences over the primitive physical outcomes and de facto choice sets is especially important in understanding the nature of existing axiomatic choice-set dependent models such as Gul and Pesendorfer (2001) that specify axioms on preferences over choice sets. These models assume that preferences over the choice sets do not themselves depend on the menu of choice sets they are offered. That is, they propose a choice domain that itself excludes choice-set dependence. While the literature often does not emphasize the second stage of choices from choice sets, by being explicit about this second stage, we can observe that the menu independence on these meta-choices is in fact demanding that preferences over outcomes depend on procedures. The prototype of temptation preferences, for instance, is that person chooses the choice set {apple} over the choice set {candy, apple}, yet if she instead were given the choice set {candy, apple} imposed on her in the second stage, she would choose candy. But this says that when she is choosing between apple and candy one way (by first choosing among fruity choice sets and then choosing fruit) she chooses apple, but if she chooses another way (by choosing directly between fruit) she chooses candy. While assuming a person feels differently about foregoing candy indirectly versus directly is compelling insofar as these models are meant to capture intertemporal preferences — embedding the intuitions of Strotz (1956) and Laibson (1997), which treat time rather than choice sets as the essential determinant of preference reversals — they implicitly rule out procedure-independent preferences. The fault lies not with the particular structure of domains of choice studied, and would not be solved by higher-order or different elicitation techniques. Rather, the problem is that simple procedure-independent functions are not observable:

Procedure-independence implies 'meta menu dependence': Suppose for some set of physical outcomes Q, h is a menu-dependent happiness function in the environment consisting of subsets of Q: $h(q|A)\neq h(q|B)$ for some $A\subseteq Q$, $B\subseteq Q$, and $q\in A\cap B$. Then in any environment Ξ that implements Q, the procedure-independent happiness function h' in Ξ that corresponds to h is menu-dependent in Ξ: there exists X, $Y\in\Xi$ and $x\in X\cap Y$ such that $h'(x|X)\neq h'(x|Y)$.

This result says that one tempting, innocuous-seeming restriction on well-being that researchers might intuit as a way to fully identify preferences — that we find *some* domain of choice where preferences are choice-set independent — excludes very simple forms of menu-dependent preferences. Whether talking about apples, painful death, aversion to earning less than others (discussed below) or anything else, the simple examples we discuss all fit into this class.

A second, even more straightforward feature of procedure-independent preferences is that their non-identifiability in simple settings cannot be solved in more complicated settings, even allowing for the necessary menu-dependence in these broader settings. Formally:

Procedure-independent preferences are unidentifiable: Consider any set of physical outcomes Q and any pair h and h' that are procedure-independent functions of Q such that $\arg\max_{x\in Z} h(c(Z)|Z)=\arg\max_{x\in Z} h'(c(Z)|Z)$ for all $Z\subseteq Q$. Then for all Ξ that implements Q, and for the happiness functions $\hat{h}$ and $\hat{h}'$ in Ξ that correspond to h and h', $\arg\max_{x\in X}\hat{h}(c(X)|X)=\arg\max_{x\in X}\hat{h}'(c(X)|X)$ for all $X\in\Xi$.

This result establishes that no elicitation procedure can identify procedure-independent preferences; the same ancillary assumptions to designate any restriction on well-being as a function of choice in the simple settings are needed in the more complicated settings.

While common sense can identify well-being in the cake–death example, and acknowledgment that it is the contrast between prospective and immediate choice that is the essence of self-control can be combined with the assumption that it is the prospective preferences that determine well-being to conduct compelling welfare analysis of temptation preferences, the choice–unobservability of well-being is in other domains likely to be very real — and where people's preferences over meta-choice sets may in fact not reveal their relative happiness of different outcomes imposed. In these cases, psychological research or happiness studies may identify systematic but choice–unobservable patterns of happiness.

Specifically, the connection between choice and welfare is often genuinely tenuous and hard to discern in the context of "social preferences" — how people depart from self-interest in assessing the allocation of resources between themselves and others. Suppose, for instance, we observe a person who always chooses to share with others if she can. It could be that having the option to share makes her happy. 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 enviously dislike doing worse than those around her, but never choose to rectify this because she would feel even worse about hurting others. None of these distinctions are observable in choice.

In fact, the potential wedge between "preferred" outcomes and implemented outcomes can take virtually any form. To show such a general possibility, we borrow Gul and Pesendorfer's (2001) elegant representation. Let Z be a finite, deterministic choice set of whatever structure, and let $M(Z)=\{(x_i, y_i)\}_{i=1,\ldots}$ be the possible allocations of money between the decision maker and another person that can arise from any choice in Z. Let $f(x, y)$ and $g(x, y)$ be any two functions assigning a real number to each allocation. For each $z\in Z$, let $(x(z), y(z))\in M(Z)$ be the allocation generated by Z. Then define

$$h(z|Z)\equiv f(x(z),y(z)) - \theta|\mathrm{Max}_{z'\in Z} g(x(z'),y(z')) - g(x(z),y(z))|$$

for $\theta > 0$. That is, the person gets disutility from not maximizing $g(x, y)$, but her utility is also determined by $f(x, y)$. If $\theta \to \infty$, observed behavior — again, on *any* complicated choice sets, including choices over choice sets of allocations, etc. — will be fully determined by $g(\cdot, \cdot)$, but well-being will be fully determined by $f(\cdot, \cdot)$. This shows that observable behavior is completely separate, in principle, from the situations and outcomes that people find desirable.

Note again: because these preferences are defined over what final allocations she can achieve, no choice set can be given to find out if she wants the opportunity to share without giving her the opportunity to share. The functions $h(z|Z)$, $h(c(Z)|Z)$, and $H(Z)$ — the last needed to identify what situations people are better off in — are unrecoverable from behavior alone.

This example where significant questions cannot be adequately inferred from choice is in fact quite pertinent to some burgeoning literatures in economics. An explosion in experimental economics studying interpersonal allocations has recently led to various models of social preferences. Experimental evidence is strong, for instance, that people have a taste for fair outcomes and sacrifice money to help others who are getting less than they are, and that people sacrifice money to retaliate against unfair treatment.[3] But another hypothesis that accords to the psychology of envy and social comparison receives far less behavioral support: outside of cases where these others have misbehaved, do people behave as if they are bothered by getting worse allocations than others? The evidence indicates (see, e.g., Charness and Rabin, 2002) that a majority of subjects would in fact choose (self, other) allocations such as ($11, $25) over ($10, $10) when allocating between themselves and anonymous other parties who have done them no harm — not sacrificing $1 to avoid coming out behind.[4]

Yet, per our observations above, experimental research demonstrating relatively little "behindedness aversion" as a *behaviorally* prevalent aspect of allocational preferences cannot be seen as determinative as to whether or not people have a strong distaste for doing worse than others. It could be that most people *are* less happy with allocation ($11, $25) than being *given* the allocation ($10, $10), but they don't let themselves choose ($10, $10) from the choice set or the choice set {($10, $10), ($11, $25)} because they would feel even worse choosing this.

Indeed, this very real possibility sheds light on comparing choice data to other ways of inferring preferences. Loewenstein et al. (1989), for instance, asked subjects in a survey how they would feel about various allocations of money between themselves and others following various real-world scenarios, and find that subjects report lower satisfaction with settlements of a conflict that give the other party more money — *fixing their own allocation*. Articles such as Charness and Rabin (2002) argue implicitly (and by the utility function they write down, if interpreted to be applicable to welfare comparisons across situations) that their own choice-based data are more reliable than the hypothetical satisfaction questions. But this conclusion is unwarranted, and suggestive of the possibility that the natural welfare conclusions researchers might take away from the models of preferences based on choice behavior in the laboratory might be substantially misleading. Using hypothesized satisfaction, reported satisfaction, and other measures of happiness that yield different conclusions than experimental choices may in fact be giving the right welfare conclusions.

Envy-based unhappiness is, in fact, a primary theme that comes out of the expanding literature studying the determinants of happiness based on society-wide surveys: many researchers (see, e.g., Luttmer, 2005) have gathered evidence suggesting that, fixing one's own material positions, people seem to be less satisfied when their neighbors are better off. Finding ways (even in the voting booth) of identifying whether people would act on their desire to have their neighbors be poorer at no benefit to themselves

[3] For a good recent paper that reviews much of the range of the social-preferences literature in the process of reporting new evidence, see Falk et al. (2008).

[4] In fact, Charness and Rabin (2002) and other papers find that a substantial number of subjects are willing to sacrifice small amounts of money (e.g., choose ($9.50, $25) over ($10, $10)) to help the anonymous others by a lot and making the comparison far less favorable, and a majority of subjects would probably to help others if no sacrifice is involved, even if they come out further behind (choosing, e.g., ($10, $25) over ($10, $10)).

may be extremely difficult and hence one of the reasons for using happiness data. Beyond the practical data difficulties, however, the arguments here suggest that the behavioral evidence suggesting that most people do not hurt others won't answer the question of whether or not they are happier.

Similar problems arise in inferring welfare with respect to other "behavioral" social-preferences models. As seems clear, the reason so many subjects turn down allocations such as ($2, $8) in the widely studied ultimatum game (and hence generating an outcome of ($0, $0)) is that they want to punish unfair treatment by proposers. Yet this does not — and cannot — tell us what situations make people happier. It could be that a person is happier being able to take revenge on others than she would be without the opportunity. Or it could be that she is made miserable by being mistreated. Good psychology, and modes of analysis besides choice behavior alone, are needed to glean welfare from behavior. This is a practical area where a type of preferences are using choice behavior to identify welfare that may be unidentifiable from such choice behavior.

Although a bit outside our framework, it is worth noting that some very similar issues regarding the choice–unobservability of preferences arise in the domain of "belief-based preferences," where a person's well-being may be influenced for non-instrumental reasons by her beliefs. A recent spate of theoretical research (see, e.g., Caplin and Leahy, 2001; Kszegi, 2006, and Brunnermeier and Parker, 2005) have proposed utility functions capturing the notion that beliefs directly influence well-being. Happiness may, for instance, be influenced by 'ego utility' — our assessment of our own talents and status — as well as beliefs about future economic or health prospects. In this domain, identifying whether particular beliefs will make the person happier may be very difficult, especially when the utility is linear in beliefs and agents obey (as full rationality demands) the law of iterated expectations. Indeed, if we were interested in finding out whether somebody would be happier knowing that a movie star has checked into a drug clinic or less happy (presumably many people have each type of preference), it may be difficult to find that out from choice behavior. Letting $h(q)$ be a person's happiness when believing the star has a drug problem with probability q, by eliciting preferences by over information, we can presumably find out how $h(q)$ compares to $q \cdot h(1)+(1-q) \cdot h(0)$, but this or richer information-choice data cannot tell us whether $h(\cdot)$ is increasing or decreasing.

Similar types of problems to those illustrated above might arise in identifying even whether people are on average better off or worse off with information: part of the psychology of curiosity may involve an unpleasantness associated with not finding out *available* information. A person may, for instance, be happier (and know that she will be happier) not having the ability to find out when and how she will die than being able to find out, but nonetheless choose to find out if she can. Such preferences, and hence the happiness associated with making information available to a person, would be impossible to distinguish solely with the choice behavior we observe.

3. Mistaken choice and happiness

In the previous section we observed the necessity for rational-choice welfare economics of ancillary choice–unobservable assumptions. In this section we demonstrate that *with* reasonable ancillary assumptions, choice behavior can be a powerful tool in revealing preferences and well-being even without assuming extreme rationality *a priori*. In fact, in the many domains where interpreting behavior as fully implementing preferences seems to yield incoherent or ridiculous conclusions, revealed preference can only be rescued as a powerful tool if the notion that choice always implements preferences is abandoned.

While explored in more detail in Kszegi and Rabin (2008), here we outline briefly the approaches to improving the power of revealed preference better by acknowledging mistakes. The basic approach is to find a setting where the nature of some state-contingent preferences is obvious, so that choices reveal beliefs about the likelihood of those states, including any systematic mistakes. Then use the mistakes in beliefs that are revealed, rather than rational expectations, to interpret what preferences are in situations where those preferences 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_H, x_T) represent a lottery that pays $\$x_H$ if the next flip of a coin comes up heads (H) and $\$x_T$ if the next flip comes up tails (T). Suppose we observe the following choices:

1. If the person observes that the previous flips come up HHH, she chooses $(x_H, x_T)=(85, 120)$ over $(x_H, x_T)=(120, 90)$ 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.
2. If instead she observes TTT, she chooses $(x_H, x_T)=(120, 85)$ over $(x_H, x_T)=(90, 120)$ on the next flip. This amounts to a bet that H will come up next.
3. If she has observed no flips, she chooses $(x_H, x_T)=(90, 120)$ over $(x_H, x_T)=(120, 85)$ and $(x_H, x_T)=(120, 90)$ over $(x_H, x_T)=(85, 120)$ 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 we observe some fruit bets based on coins. Letting (f_H, f_T) be the bundle of fruits she gets following a H or T, suppose we observe:

1. If the person observes that the previous flips come up HHH, she chooses $(f_H, f_T)=$(4 apples, 4 oranges) over $(f_H, f_T)=$(5 oranges, 5 apples).

2. If she observes that the previous flips come up TTT, she chooses (f_H, f_T)=(4 oranges, 4 apples) over (f_H, f_T)=(5 apples, 5 oranges).
3. If she has observed no flips, she chooses (f_H, f_T) of either (5 apples, 5 oranges) or (5 oranges, 5 apples) over (f_H, f_T) of either (4 apples, 4 oranges) or (4 oranges, 4 apples).

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 non-random situation. Her preferences are revealed by behavior, even though they are not implemented by behavior.[5] And 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. Importantly, she may be better off with that choice set than being able to choose among 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.

This procedure can be used in far more practical settings. Do people have a reasonable theory of stock markets? Are they over-trading because they enjoy trading, or because they think they are outsmarting the market? If we directly elicit their beliefs about the market, we can discover whether they like the sport of it, and are happy to lose money doing it, or whether their goal is in fact higher return and lower risk for retirement savings, and their behavior reflects a mistake. In general, it is likely that with enough data it is easy to identify mistaken beliefs about the way the world works by combining the observation of all the implicit bets people are making with the minimal identifying assumption of stochastic monotonicity of monetary preferences.

A more intriguing possibility is that this minimal assumption that people prefer more money to less can be used to empirically identify or experimentally test whether people mispredict their *own future behavior*. To return to our earlier example: suppose we suspect that people who are choosing [{candy, apple} → candy] are really people who *think* they are choosing [{candy, apple} → apple] but underestimate their self-control problem. To identify such a possibility, suppose we have them 'bet' on what they'll choose by attaching a penny to one option: They can choose

$$[\{\text{candy} + 1 \text{ cent}, \text{apple}\} \rightarrow \text{candy} + 1 \text{ cent}] \text{ or}$$
$$[\{\text{candy} + 1 \text{ cent}, \text{apple}\} \rightarrow \text{apple}] \text{ or}$$
$$[\{\text{candy}, \text{apple} + 1 \text{ cent}\} \rightarrow \text{candy}] \text{ or}$$
$$[\{\text{candy}, \text{apple} + 1 \text{ cent}\} \rightarrow \text{apple} + 1 \text{ cent}].$$

If we see a person choosing the third (asking for a penny with the apple, but then eating candy), we can suspect a misprediction. Knowing fully what to infer about the welfare of such person is more difficult in this case — would she be better off with a choice set containing candy or not? Yet minimal restrictions on her disutility of temptation and attitudes towards money might suggest that she would be better off with the {candy+1, apple} choice set than the {candy, apple+1}.

This type of error, misprediction of one's own future preferences, seems an important issue in identifying the connection in the real world between observed behavior and well-being. Consider the possibility that unaddicted 18-year-olds underestimate the effect of addiction on their future preferences and behavior. There is some convincing evidence that even addicts who understand and want to fight their addiction do not appreciate the strength of cravings. For instance, Giordano et al. (2004) used an incentive-compatible procedure to elicit 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. Addicts valued the dose significantly more if they were deprived, even though they knew their current satiation level had nothing to do with their future satiation level. Such evidence that addicts who are sated at the moment systematically underappreciate the strength of their future cravings suggests it is possible that 18-year-old non-addicts may similarly systematically underappreciate their future cravings if they become addicted.

4. Do we need to identify mistakes to study happiness?

Although various examples above indicate that decomposing observed choices into preferences and into mistakes is quite often both feasible and desirable in practical terms, in this section we clarify and illustrate two limitations to both the feasibility and necessity of such decompositions in principle. First, it follows from our analysis above that mistaken behavior cannot, except by psychological interpretation, be distinguished from non-mistaken behavior. But second, that for welfare analysis conceived as situational comparative statics on well-being, the very inability to distinguish mistakes from non-mistakes also indicates that it does not necessarily matter that we do so.

We first note a class of behavioral phenomena that will illustrate these themes especially well: framing and focusing effects. Different descriptions of mathematically identical decision problems induce different patterns of choices. In a famous example of such "framing effects," Tversky and Kahneman (1981) gave students the following instructions: "Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer." Note

[5] Indeed, we note that a researcher armed with general knowledge of the types of mistakes that people make would in fact be savvy enough to guess the preferences for oranges over apples based solely on this second set of choices, while a credulous believer in rational choice might be tempted to conclude that this person prefers more fruit to less fruit following initial flips of a coin, but less fruit to more fruit following long streaks of coin flips.

in particular the instructions' use of concurrent decisions and the request to first examine both decisions. The decisions were:

Decision 1: Choose between
(A) Getting $240 for sure; and
(B) Getting $1000 with probability 0.25.
Decision 2: Choose between
(C) Losing $750 for sure; and
(D) Losing $1000 with probability 0.75.

When given these decisions separately, most subjects choose A and D. Yet A and D combine to a gain of $240 with probability 0.25 and a loss of $760 with probability 0.75, while B and C combine to a gain of $250 with probability 0.25 and a loss of $750 with probability 0.75. When put in this "broad" frame, nobody would choose A and D.

Focusing effects are when focusing a person's attention on an issue or choice, or making salient certain outcomes or possibilities, can alter behavior. If we remind a person of a small risk, we might induce her to buy insurance against that risk, even though she might never have thought about doing so herself. And if we make the possibility of skin cancer salient to a patient, she may get nervous and get checked for it, even if we have given her no new information.

Framing and focusing effects can be interpreted in two ways. Under one interpretation, the frame or focus of an individual affects her preferences, and these preferences are translated into frame-sensitive choices. In the alternative view, preferences do not depend on the momentary frame or focus, but some decision situations lead people to make mistakes in implementing their stable preferences.

Yet it may be difficult to tell which of these interpretations is right. In order to tell whether a choice in a particular frame is a mistake, we would have to induce the person to make other choices. But the decision frame that induces changed behavior may be changing preferences rather than inducing or stopping mistakes. So, for instance, if making salient some particular costs associated with an outcome always makes a person avoid that outcome, it may be hard to separate out whether this is because the negative frame makes that outcome a worse experience — or just mistakenly seem that way to the decision maker.

Yet the very difficulty of identifying whether such choices are mistakes or not is related to the question of whether — from a welfare point of view — we need to do so. Whether these effects are due to true context-dependent preferences, or to mistakes, happiness data can in principle help tell us which descriptions lead to the most happiness. If we have enough data to measure happiness across all situations, the channel through which environments generate happiness are unimportant. 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.

More generally, for all X, we can only observe $c(X)$ and $h(c(X)|X)$. Whether (for non-singleton choice sets) it is a mistake is unobservable, since we never observe $h(x|X)$ for any $x \neq c(X)$. The well-known positive "folk theorem" about the unobservability of rationality is that the question whether or not $c(X) = \arg\max_{x \in X} h(x|X)$ is inherently unobservable in choice data. As such, what we are observing is not only consistent with some combination of $c(\cdot)$ and $h(\cdot|\bigcirc)$ where $c(X) = \arg\max_{x \in X} h(x|X)$ for all X, but also consistent with some combination of $c(\cdot)$ and $h(\cdot|\bigcirc)$ where $c(X) \neq \arg\max_{x \in X} h(x|X)$ for all X.

An obvious corollary to this is that, given an observed $c(\cdot)$, any normative conclusions about $H(X)$ that can be reached based on some theory of $h(\cdot|\bigcirc)$ and some theory of mistakes can also be reached by some theory of $h'(\cdot|\bigcirc)$ where $c(X) = \arg\max_{x \in X} h(x|X)$ for all X. Not only does there always exist a fully rational explanation for all observed behavior that some researchers might attribute to limited rationality — there also exist rational explanations that are consistent with the *welfare conclusions* such a researcher reaches. No *combination* of behavior and welfare conclusions from non-rational theory can possibly be inconsistent with rationality. Whatever the merits of a methodological mandate to interpret behavior as rational, it never tells us anything about the conclusions reached by alternative approaches.

One could, for instance, replicate all the above gambler's-fallacy-based predictions while maintaining the assumption of rational-utility maximization. To mimic the behavioral predictions, one can assume that the person likes betting whatever she prefers 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 assume that the person is happier betting if she has observed no flips than if she has observed HHH.

Because the mistakes-based theory of the gambler's fallacy provides general and ex-ante (rather than ex-post) guidance to predicting both behavior and welfare, it is better economics than the contrived rational-choice theory. 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 naivete about self-control problems vs. fully rational commitment-aversion to controlling oneself — are in simple settings not distinguishable. Or (to take an example from a major theme in happiness research discussed earlier) people may be less happy when sacrificing local status by moving into wealthier neighborhoods, and yet move into such neighborhoods for either of two reasons. They might 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 the new one. Without non-choice measures of happiness we could not be sure whether a person is better off with the

opportunity to move into wealthier neighborhoods, irrespective of what we assume about whether people who move are rational. *With* such non-choice measures, we can make a well-being assessment irrespective of what we assume about rationality.

5. Situations, outcomes, and measured happiness

Despite our view that choice behavior, used with care, should remain a primary way of inferring well-being (even when people might make mistakes), we believe the case is strong for economists to start to integrate alternative measures of well-being. In this section, we briefly consider how some of the issues raised above might shed light on empirical approaches to studying happiness. Again, our emphasis is not on which measures of happiness are used nor on the current state of the reliability of these measures; it is on the conceptual issues of what one does with the observed measures that one believes to be informative.

If one takes the view that the primary point of welfare economics is to compare people's well-being in different situations, the most fruitful approach to happiness research is to study the choice *environment* as the determinant of happiness — rather than the choices within that environment. This suggests, for instance, for both broad and specific policy questions researchers should study whether people will on average be happier in environments where marriage is easier or harder rather than whether married people are happier than unmarried people in any given environment.

Although sometimes seemingly framed in terms of comparing outcomes rather than comparing choice sets, much happiness research is in fact highly consistent with this perspective. One reason — perhaps the main reason — that researchers have studied happiness of outcomes directly rather than used revealed preference is in fact that in many cases people do not make, or we cannot observe, the relevant choices. In situations where the only observed data are the relative happiness of people with the choice set $\{x\}$ vs. those with the choice set $\{y\}$, and the primary policy-relevant question is which of these two situations will make people better off on average, then there is little alternative to (and even less reason to refrain from) trying to measure people's well-being in these situations directly. More generally, while we can get some traction on the effects on well-being of noise pollution, crime, unemployment, social comparison, and health through voting behavior or housing prices, in practical terms getting a full picture is often difficult. Research on the effects of these factors can help inform policy about how much weight to put on affecting these outcomes. And in all such cases, none of the issues of endogeneity of choice, the influence of choice sets, etc., mitigate the insights gleaned from the research.

As in general empirical work, of course, in all such studies due caution must be taken in identifying correlates of those facing one situation rather than another, and researchers must worry about endogeneity problems. And one endogenous factor to consider is happiness itself: unhappy people lose their jobs, watch T.V., and get divorced more than happy people. For all these cases, one must use some of the classical techniques of econometric research of identifying natural experiments or instrumental variables to find exogenous variation in these environmental features in order to know the direction of causality.

But it is worth noting that in situations involving choice the endogeneity is more than a problem of econometric techniques, and reflects a more fundamental wedge between research that studies happiness as a function of outcomes rather than welfare economics conceived of as situational comparative statics. While a finding that married people are happier than unmarried people, for instance, may be interesting evidence for many questions, its relevance for situational comparative statics is limited on a couple of grounds. Even controlling for opportunities, people who get married are different than those who don't. This isn't merely a run-of-the-mill econometric selection problem, nor a case where there *might* be some unobserved differences. Seen through the lens of the rational-choice approach to economics, it is the mother of all selection problems: the core assumption of this approach is that different choices from a choice set *must* come from different preferences,. As an illustration, suppose that we observe two people with the choice set $\{x, y\}$ and find that $c_1(\{x, y\})=x$, $c_2(\{x, y\})=y$, and $h_1(x|\{x, y\}) > h_2(y|\{x, y\})$. For example, one person chooses an apple, and the other an orange, and the apple person is happier. What assumptions would allow us to conclude that $H_2(\{x\}) > H_2(\{x, y\})$ — that we should only make apples available — or the weaker conclusion that $H_2(\{x\}) > H_2(\{y\})$? Certainly under the maintained hypothesis that people are rational, and even in most plausible theories of errors, different choices within a choice set constitutes *a priori* evidence that the person is different. Seen in this light, comparing the happiness of those who choose differently in the same circumstances seems per se inappropriate.

One could imagine assumptions that would make such outcome-based happiness data of interest in comparing situations. It could simply be that, whatever heterogeneity there is in underlying happiness levels or preferences, choices themselves are uncorrelated with that heterogeneity. If that were so, we could use evidence of how outcomes influence happiness to determine what options are best for people. While such an assumption could come from assuming extreme departures from rationality, there is sometimes a plausible reason for supposing that choice reveals relatively little about a person's preference. Namely, in some settings there may in fact be very little heterogeneity in underlying preferences, and variation in behavior in such domains would come from differences in beliefs across people as to what will deliver them well-being rather than actual differences. So, for instance, it could be that in some investment contexts, variation in plausible preferences over portfolios may be quite small compared to variation in beliefs people have as to how the world works. Even if researchers are agnostic about what those underlying preferences are, studying the happiness of people employing different strategies, even if endogenously chosen, may tell us which environment might make people happiest.

Given how strong and often unrealistic are the assumptions needed for inferring situational comparative statics from outcomes alone, it would seem a more useful general technique would be to compare the happiness levels of all people given different choice sets, not the levels of those making different choices in a given context. For instance, rather than compare

$$h(\text{smoke}|\{\text{smoke, nonsmoke}\})<h(\text{nonsmoke}|(\text{smoke, nonsmoke}))$$

to infer whether a ban on smoking might be useful, it would seem more useful to compare entire populations in different environments. If we had data on groups of people who are free to smoke vs. those that are not, for instance, we could do something like comparing the average

$$p_{\text{smoke}} \cdot h(\text{smoke}|\{\text{smoke}, \text{nonsmoke}\}) + (1 - p_{\text{smoke}})h(\text{nonsmoke}|\{\text{smoke}, \text{nonsmoke}\}) \text{ vs. } h(\textit{nonsmoke}|\{\textit{nonsmoke}\}),$$

where p_{smoke} is the proportion of people who smoke when given the opportunity to do so.

Gruber and Mullainathan (2002), in fact, undertake an analogue of this exercise, showing that the average happiness of a group of people statistically identified as would-be smokers are happier in locations where tobacco taxes are higher than where they are lower. More generally, data comparing h(nonsmoke|{nonsmoke}) vs. h(smoke|{smoke}) would be a safe way to infer well-being under the maintained hypothesis that well-being is not choice-set dependent. And Oreopoulos (2007) performs a similar exercise, finding that English teenagers who can be statistically identified as having stayed an extra year in school solely because of a change in school-leaving laws were substantially better off in a variety of economic and cultural dimensions, and in reported happiness, many years later. While one cannot be certain that the initial pain caused by forced schooling didn't outweigh the long-term benefits, the results are suggestive that the lifetime well-being of people forced to stay in school until 16 years old is greater than those (with the larger choice set) of being able to drop out at 15.

In these examples of smoking and dropping out of school, the data can be naturally interpreted as people making mistakes. Other research similarly seems either explicitly or implicitly premised on possibilities of mistakes. For instance, if studies suggest that people who can be identified as working hard for higher income are less happy than those without the opportunity to work hard for more money, insofar as this is a choice variable it probably indicates that people work too hard compared to what would be rational. But it could also be that happiness is choice-set dependent, so that people don't like working hard, but dislike turning down jobs even more. As argued in the previous section, however, identifying what choices are mistakes is not necessary for situational comparative statics to yield useful insights; being offered lucrative jobs that diminishes time for leisure, family, and friends can be identified as either conducive to happiness or detrimental independent of whether we can identify people as mispredicting their happiness.

6. Conclusion

We are confident that in the not-very-long long term economists will combine our discipline's insights and methods with other approaches to studying happiness to modify and improve welfare analysis. In the Introduction we noted three likely questions about this research program that this article does not address.

The first was the central issue in much current research of how one measures happiness or well-being. While we share most practitioners' skepticism about whether measures yet exist to confidently measure happiness, we also share the view that some measures of well-being are sufficiently feasible and sufficiently necessary to make pursuit of them worthwhile. Seen through the arguments of this paper, indeed, the question is not in general whether the measures are sufficiently reliable to overturn revealed-preference approaches to well-being. In addition to the belief that these measures will improve over time as the product of focused research, our arguments clarify that the appropriate test of their *current* value is whether, for all their imperfections, the measures are sufficiently reliable to sometimes be useful to complement the intuitions and judgment *currently* embedded in all welfare conclusions reached in economics.

The second was how one makes interpersonal comparisons of well-being when that is called for. Although we have nothing to say on the topic beyond this sentence, we feel that the role of attempting to devise interpersonally-comparable measures of cardinal happiness is obviously a legitimate exercise that equally obviously could benefit from non-choice data.

On the third question, of whether economists should study happiness, the premise of this article is that the answer is a resounding “yes.” And notwithstanding considerable variation in emphasis on welfare among strands of economic research, we also think the question of whether economic research and teaching does *now* (via our various notions of efficiency and welfare) reach strong conclusions about well-being is also a resounding “yes.” We are hopeful and enthusiastic about the possibility that the concern for and study of well-being will expand dramatically in due time.

We have some concern in the short term, however: as researchers embrace new insights and new methods that challenge the approach of inferring well-being mechanistically from choice behavior, and abandon the presumption of superhuman rationality that has until recently dominated formal economic modeling, we fear many economists will be tempted to move in the opposite direction, and abandon the discipline's traditional interest in welfare analysis.

Whatever the merits of these supplementary approaches to studying well-being, one possible reaction is to question whether these alternative approaches are really in the domain of economics. The answer to this is, of course, “yes.” Questions of whether and how well-being is influenced by institutions that affect savings, access to credit, economic growth, inequality, addiction, market pricing, different forms of contracting, investment incentives, and retirement savings, whatever the best way to answer those questions, are the stuff of welfare economics. To outsource all these traditional topics of economics to other disciplines on methodological grounds would be an odd and unlikely maneuver. Doing so would be especially regrettable because it is clear that the powerful theoretical and empirical methods, insights, and assumptions of economics, centered around role of choice and the assumption of rationality, can be fruitfully applied to and combined with the new assumptions and methods. As inferences along the lines of Giordano et al. (2004) begin to be applied to field data on addiction, economists armed with price theory and awareness of how much we learn from choice are in a unique position to study when becoming addicted is rational and when it is a mistake.

More generally, because of the heavy emphasis within economics on the role of choice as a factor in determining the satisfaction of wants, we think the study of happiness will be improved if our discipline participates in this research agenda.

Acknowledgments

We thank Tim Besley, Stefano DellaVigna, Antonio Rangel, Robert Sugden, and two anonymous referees for comments, and the National Science Foundation (Grant SES-0648659, 2007-2010) for financial support.

References

Brunnermeier, Markus, Parker, Jonathan, 2005. Optimal expectations. American Economic Review 95 (4), 1092–1118.
Caplin, Andrew, Leahy, John, 2001. Psychological expected utility and anticipatory feelings. Quarterly Journal of Economics 116 (1), 55–79.
Charness, Gary, Rabin, Matthew, 2002. Understanding social preferences with simple tests. Quarterly Journal of Economics 177 (3), 817–869.
Dekel, Eddie, Lipman, Bart, Rustichini, Aldo (2007), Temptation-driven preferences, mimeo, July 2007.
Falk, Armin, Fehr, Ernst, Fischbacher, Urs, 2008. Testing theories of fairness — intentions matter. Games and Economic Behavior 62, 287–303.
Frey, Bruno, Stutzer, Alois, 2002. What can economists learn from happiness research? Journal of Economic Literature 40 (2), 402–435 June 2002.
Giordano, Louis A., Bickel, Warren K., Loewenstein, George, Jacobs, Eric A., Marsch, Lisa, Badger, Gary J., 2004. Mild opioid deprivation increases the degree that opioid-dependent outpatients discount delayed heroin and money. Psychopharmacology 163 (2), 174–182.
Gruber, Jonathan, Mullainathan, Sendhil, 2002. Do cigarette taxes make smokers happier? NBER Working Paper #8872.
Gul, Faruk, Pesendorfer, Wolfgang, 2001. Temptation and self-control. Econometrica 69 (6), 1403–1435 Nov., 2001.
Huber, J., Payne, J.W., Puto, C., 1982. Adding asymmetrically dominated alternatives: violations of regularity and similarity hypotheses. Journal of Consumer Research 9, 90–98.
Kahneman, Daniel, Diener, Ed, Schwarz, Norbert (Eds.), 1999. Well-being: The Foundations of Hedonic Psychology. Russell Sage Foundation, New York.
Kszegi, Botond, 2006. Emotional agency. Quarterly Journal of Economics 121 (1), 121–156.
Kszegi, Rabin, 2008. Revealed mistakes and revealed preferences, in Caplin and Schotter, eds., Methodologies of Modern Economics, Oxford University Press.
Laibson, David, 1997. Hyperbolic discounting and golden eggs. Quarterly Journal of Economics 112 (2), 443–477.
Layard, Richard, 2005. Happiness: Lessons from a New Science. Penguin, London.
Loewenstein, George, Thompson, Leigh, Bazerman, Max, 1989. Social utility and decision making in interpersonal contexts. Journal of Personality and Social Psychology 57 (3), 426–441.
Luttmer, E.F.P., 2005. Neighbors as negatives: relative earnings and well-being. Quarterly Journal of Economics 120 (3), 963–1002 August 2005.
Oreopoulos, Philip, 2007. Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. Journal of Public Economics 91, 2213–2229.
Sen, Amartya, 1993. Internal consistency of choice. Econometrica 61 (3), 495–521.
Simonson, I., Tversky, A., 1992. Choice in context: tradeoff contrast and extremeness aversion. Journal of Marketing Research 29, 281295.
Strotz, R.H., 1956. Myopia and inconsistency in utility maximization. Review of Economic Studies 23 (3), 165–180.
Tversky, Amos, Kahneman, Daniel, 1981. The framing of decisions and the psychology of choice. Science 211 (4481), 453–458.
Tversky, Amos, Thaler, Richard, 1990. Anomalies: preference reversals. Journal of Economic Perspectives 4 (2), 201–211.