PART III

NEW DIRECTIONS FOR NORMATIVE ECONOMICS
Interest in behavioral economics has grown in recent years, stimulated largely by accumulating evidence that the standard model of consumer decision making provides an inadequate positive description of human behavior. Behavioral models are increasingly finding their way into policy evaluation, which inevitably involves welfare analysis. No consensus concerning the appropriate standards and criteria for behavioral welfare analysis has yet emerged.

One common strategy in behavioral economics is to add arguments to the utility function (including all of the conditions upon which choice seems to depend) in order to rationalize choices, and to treat the new arguments as welfare relevant. Unfortunately, such an approach is often problematic as a guide for normative analysis, and in some instances simply untenable. For example, if an individual’s decision depends on whether he has first viewed the last two digits of his social security number (as the literature on anchoring suggests, e.g., Tversky and Kahneman [1974]), should a social planner attempt to determine whether the individual has recently seen those digits before making a choice on his behalf? Perhaps more important,
in many cases the nature and significance of the condition under which the choice is made change when the choice is transferred to a social planner. Consider the example of time inconsistency: the individual chooses alternative $x$ over alternative $y$ at time $t$, and $y$ over $x$ at time $t - 1$. One could rationalize this behavior by inserting the time of the decision into the utility function, so that the individual pursues different objectives at $t - 1$ and $t$ (quasi-hyperbolic discounting is an example, e.g., Laibson [1997]). But if a social planner must choose for the individual from the set \{ $x$, $y$ \} at time $t$, which perspective should it adopt? One could argue that the planner should choose $x$, the same alternative that the individual would pick at time $t$. But one could also argue that the planner should mimic the choice of $y$ at time $t - 1$, on the grounds that the planner’s decision, like the individual’s decision at time $t - 1$, is at “arms length” from the experience. Much of the literature on self-control takes this second view. However, neither answer is obviously superior.

The obvious problems with the normative methodology described in the preceding paragraph have led many behavioral economists to distinguish between “decision utility,” which provides an “as if” representation of choices, possibly by invoking unconventional assumptions concerning preferences, and “true” or “experienced” utility, which is viewed as the proper measure of well-being. This approach forces one to take a stand on the nature of true utility. But the objective basis for making any assumptions about true utility is, at best, obscure.¹

This chapter is a partial summary of Bernheim and Rangel [2007], which reflects our effort to develop a unified framework for behavioral welfare economics—one that can be viewed as a natural extension of standard welfare economics. (It is also related to work by Rubinstein and Salant [2007]; we explain the relations below.) The standard approach to welfare analysis is based on choice, not on utility, preferences, or other ethical criteria. In its simplest form, it instructs the social planner to respect the choices an individual would make for herself. The guiding normative principle is an extension of the libertarian deference to freedom of choice, which takes the view that it is better to give a person the thing she would choose for herself rather than something that someone else would choose for her. Thus, with respect to public policy decisions, the standard approach instructs the planner to mimic individual choices.

We show that it is possible to extend standard choice-theoretic welfare analysis to situations in which individuals make “anomalous” choices of the various types commonly identified in behavioral research. Indeed, standard welfare economics is a special case of the framework proposed here; specifically, it is a limiting case in the sense that our framework converges to the standard framework as behavioral anomalies become small. We also show that it is possible to generalize standard tools such as compensating and equivalent variations, consumer surplus, and Pareto optimality.

The chapter is organized as follows. First we review the perspective of standard welfare economics, and then present a general framework for describing choices.
Next we set forth choice-theoretic principles for evaluating individual welfare in the presence of choice anomalies. Then, after describing the generalizations of compensating variation and consumer surplus in this setting, we explain how we generalize the notion of Pareto optimality, and examine market efficiency as an application. Last, we set forth an agenda for refining our welfare criterion and explain the role of nonchoice evidence in behavioral welfare economics, followed by our conclusion.

A Review of the Foundations for Standard Welfare Economics

Standard welfare economics consists of two separate tasks. The first task involves an assessment of each individual’s welfare; the second involves aggregation across individuals. Our object here is to develop a general framework for executing the first task—one that encompasses the various types of anomalous choices identified in the behavioral literature. As we discuss further below, aggregation can then proceed much as it does in standard welfare economics, at least with respect to common concepts such as Pareto efficiency. Consequently, our main objective here is to review the standard perspective on individual welfare.

Choices and Welfare

To summarize the standard perspective formally, we must introduce some notation. We will use $\mathcal{X}$ to denote the set of all possible choice objects. The standard framework allows for the possibility that choice objects are lotteries, and/or that they describe state-contingent outcomes with welfare-relevant states. A standard choice situation (SCS) consists of a constraint set $X \subseteq \mathcal{X}$. When we say that the standard choice situation is $X$, we mean that, according to the objective information available to the individual, the alternatives are the elements of $X$. The choice situation thus depends implicitly both on the objects among which the individual is actually choosing, and on the information available to him concerning those objects.

The objective of standard welfare economics is to provide coherent criteria for making welfare judgments concerning possible selections from standard choice situations. We will use $\mathcal{X}'$ to denote the domain of standard choice situations with which standard welfare economics is concerned. Usually, the standard framework takes $\mathcal{X}'$ to include some reasonably exhaustive collection of compact sets. To minimize technical details, throughout this chapter assume that $\mathcal{X}'$ includes all nonempty, finite subsets of $\mathcal{X}$; naturally, it may also include other subsets.
The choices that an individual would make are described by a correspondence \( C : \mathcal{X} \rightarrow \mathcal{X} \), with the property that \( C(X) \subseteq X \) for all \( X \in \mathcal{X} \). We interpret \( x \in C(X) \) as an action that the individual is willing to choose when her choice set is \( X \). Though we often speak as if choices are derived from preferences, the opposite is actually the case. Standard economics makes no assumption about how choices are actually made; preferences are merely constructs that summarize choices. Accordingly, meaningful assumptions pertain to choices, not to preferences.

The standard framework assumes that the choice correspondence satisfies a consistency property known as **weak congruence**, which generalizes the weak axiom of revealed preference (see Sen [1971]). According to the weak congruence axiom, if there exists some \( X \) containing \( x \) and \( y \) for which \( x \in C(X) \), then \( y \in C(X') \) implies \( x \in C(X') \) for all \( X' \) containing \( x \) and \( y \). In other words, if there is some set for which the individual is willing to choose \( x \) when \( y \) is present, then the individual is never willing to choose \( y \) but not \( x \) when both are present.

In the standard framework, welfare judgments are based on binary relationships \( R \), \( P \), and \( I \) defined over the choice objects in \( \mathcal{X} \), which are derived from the choice correspondence in the following way:

\[
\begin{align*}
    xRy & \iff x \in C(\{x, y\}) \\
xPy & \iff xRy \text{ and } \sim yRx \\
xIy & \iff xRy \text{ and } yRx
\end{align*}
\]

(7.1)  
(7.2)  
(7.3)

Under the weak congruence axiom, the relation \( R \) is an ordering, commonly interpreted as "revealed preference." Though this terminology suggests a model of decision making in which preferences drive choices, it is important to remember that the standard framework does not embrace that suggestion; instead, \( R \) is simply a summary of what the individual chooses in a wide range of situations. Further technical assumptions allow us to represent \( R \) with a continuous utility function.

When we use the orderings \( R \), \( P \), and \( I \) to conduct welfare analysis, we are simply asking what an individual would choose. For example, for any set \( X \), we can define an **individual welfare optimum** as the set of maximal elements in \( X \) according to the relation \( R \)—that is, \( \{ x \in X \mid xRy \text{ for all } y \in X \} \). Under the weak congruence axiom, this set coincides exactly with \( C(X) \), the set of objects the individual is willing to select from \( X \).

All of the tools of applied welfare economics are built from this choice-theoretic foundation. For example, the compensating variation associated with some change in the economic environment equals the smallest payment that would induce the individual to choose the change. A similar observation holds for the equivalent variation. The notion of consumer surplus is also entirely choice-theoretic because it measures the compensating variation of a price change. In settings with
many individuals, Pareto efficiency is a choice-theoretic concept: an alternative $x$ is Pareto efficient if there is no other alternative that everyone would voluntarily choose over $x$.

**Positive Versus Normative Analysis**

Usually, choice data are not available for all elements of $\mathcal{X}$, but rather for elements of some restricted set $\mathcal{X}^D \subset \mathcal{X}$. The objective of positive economic analysis is to extend the choice correspondence $C$ from observations on $\mathcal{X}^D$ to the entire set $\mathcal{X}$. This task is usually accomplished by defining a parametrized set of utility functions (preferences) defined over $\mathcal{X}$, estimating the utility parameters with choice data for the opportunity sets in $\mathcal{X}^D$, and using these estimated utility function to infer choices for opportunity sets in $\mathcal{X} \setminus \mathcal{X}^D$ (by maximizing that function for each $X \in \mathcal{X} \setminus \mathcal{X}^D$).

The objective of normative economic analysis is to identify desirable outcomes. In conducting standard choice-based welfare analysis, we take the product of positive analysis—the individual’s extended choice correspondence, $C$, defined on $\mathcal{X}$ rather than $\mathcal{X}^D$—as an input, and then proceed as described in the preceding section.

Preferences and utility functions, which are constructs used to extend $C$ from $\mathcal{X}^D$ to $\mathcal{X}$, are therefore positive tools, not normative tools. They simply reiterate the information contained in the extended choice function $C$. Beyond that reiteration, they add no new information that might pertain to welfare analysis.

**A General Framework for Describing Choices**

In behavioral economics as in standard economics, we are concerned with choices among objects drawn from some set $\mathcal{X}$. To accommodate certain types of behavioral anomalies, we introduce the notion of an *ancillary condition*, denoted $d$. An ancillary condition is an *observable* feature of the choice environment that may affect behavior, but that is not taken to be a welfare-relevant characteristic of the chosen object. Typical examples of ancillary conditions include the manner in which information is presented at the time of choice, or the presentation of a particular option as the “status quo.” With respect to intertemporal choice, the ancillary condition is the particular decision tree used to choose from a fixed opportunity set (which includes the points in time at which the component choices are made, and the set of alternatives available at each decision node); hence this framework can accommodate dynamically inconsistent choices.
We define a *generalized choice situation* (GCS), $G$, as a standard choice situation, $X$, paired with an ancillary condition, $d$. Thus, $G = (X, d)$. We will use $\mathcal{G}$ to denote the set of generalized choice situations of potential interest. When $\mathcal{X}$ is the set of SCSs, for each $X \in \mathcal{X}$ there is at least one ancillary condition $d$ such that $(X, d) \in \mathcal{G}$. Rubinstein and Salant (chapter 5) have independently formulated a similar framework for describing the impact of choice procedures on decisions; they refer to ancillary conditions as “frames.”

The choices that an individual would make in each GCS are described by a correspondence $C : \mathcal{G} \Rightarrow \mathcal{X}$, with the property that $C(X, d) \subseteq X$ for all $(X, d) \in \mathcal{G}$. We interpret $x \in C(G)$ as an action that the individual is willing to choose when the generalized choice situation is $G$. We will assume throughout that $C(G)$ is nonempty for all $G \in \mathcal{G}$; in other words, faced with any set of alternatives, the individual can always make a choice.

### What Are Ancillary Conditions?

As a general matter, it is difficult to draw a bright line between the characteristics of the objects in $\mathcal{X}$ and the ancillary conditions $d$. The difficulty, as described below, is that one could view virtually any ancillary condition as a characteristic of objects in the choice set. How, then, do we decide whether a feature of the choice environment is an ancillary condition?

In some cases, the nature and significance of a condition under which a choice is made change when the choice is delegated to a planner. It is then *inappropriate* to treat the condition as a characteristic of the objects among which the planner is choosing. Instead, it necessarily becomes an ancillary condition.

Consider the example of time inconsistency discussed in the introduction: the time at which a choice is made does not necessarily hold the same significance for the individual’s welfare when a decision is delegated to a planner, as when the individual makes the decision himself. We can, of course, include the time of choice as a characteristic of the chosen object: when choosing between $x$ and $y$ at time $t$, the individual actually chooses between “$x$ chosen by the individual at time $t$” and “$y$ chosen by the individual at time $t$”; likewise, when choosing between $x$ and $y$ at time $t - 1$, the individual actually chooses between “$x$ chosen by the individual at time $t - 1$” and “$y$ chosen by the individual at time $t - 1$.” With that formulation, we can then attribute the individual’s apparently different choices at $t$ and $t - 1$ to the fact that he is actually choosing from different sets of objects. But in that case, when the decision is delegated, we must describe the objects available to the planner at time $t$ as follows: “$x$ chosen by the planner at time $t$” and “$y$ chosen by the planner at time $t$.” Since this third set of options is entirely new, a strict interpretation of libertarianism implies that neither the individual’s choices at time $t$ nor his choice at time $t - 1$ provides us with any useful guidance. If we wish to construct a theory of welfare based on choice data alone, our only viable alternative
is to treat $x$ and $y$ as the choice objects and to acknowledge that the individual’s conflicting choices at $t$ and $t - 1$ provide the planner with conflicting guidance. That is precisely what we accomplish by treating the time of the individual’s choice as an ancillary condition.

The same reasoning applies to a wide range of conditions. Although we can in principle describe any condition that pertains to the individual as a characteristic of the available objects, we would typically have to describe that characteristic differently once the decision is delegated to the planner. So, for example, “$x$ chosen by the individual after the individual sees the number 47” is different from “$x$ chosen by the planner after the individual sees the number 47,” as well as from “$x$ chosen by the planner after the planner sees the number 47.” Thus, we would necessarily treat “seeing the number 47” as an ancillary condition.

In some cases, the analyst may also wish to exercise judgment in distinguishing between ancillary conditions and objects’ characteristics. Such judgments may be controversial in some situations but relatively uncontroversial in others. For example, there is arguably no plausible connection between certain types of conditions, such as seeing the number 47 immediately prior to choosing, and well-being. According to that judgment, seeing the number 47 is properly classified as an ancillary condition. Conceivably, in some cases the analyst’s judgment could be informed by evidence from psychology or neuroscience, but the foundations for drawing pertinent inferences from such evidence remain unclear.

Within our framework, the exercise of judgment in drawing the line between ancillary conditions and objects’ characteristics is analogous to the problem of identifying the arguments of an “experienced utility” function in the more standard approach to behavioral welfare analysis. Despite that similarity, there are some important differences between our framework and the experienced utility approach. First, within our framework, choice remains the preeminent guide to welfare; one is not free to invent an experienced utility function that is at odds with behavior. Second, our framework allows for ambiguous welfare comparisons where choice data conflict; in contrast, an experienced utility function admits no ambiguity. It is important to emphasize that the tools we develop here provide a coherent method for conducting choice-based welfare analysis no matter how one draws the line between ancillary conditions and objects’ characteristics.

Sometimes, the appropriate definition of the ancillary conditions for a given choice problem can be rather subtle. Consider, for example, the hypothesis that hunger affects choices pertaining to future food consumption, even when the individual knows that her hunger won’t persist (e.g., Read and van Leeuwen [1998]). To simplify this discussion, suppose that hunger is determined both by recent consumption and by randomly occurring, transient, and privately observed emotional states. In that case, each $x \in X$ would specify consumption for each emotional state and each point in time.
To capture the effect of hunger on future consumption choices, it may be tempting to think of either hunger, or the emotional states that drive it, as ancillary conditions. It is important to remember, however, that ancillary conditions must be observable; otherwise, we would not be aware that a choice anomaly (i.e., the dependence of choice on the ancillary condition) exists in the first place. With current technology, we can observe manifestations of hunger and emotions, but we cannot observe them directly; when we say that someone is hungry, we are merely interpreting those manifestations.

To specify the ancillary conditions properly, it helps to think in terms of the experimental conditions that give rise to the choice anomaly. Read and van Leeuwen [1998], for example, presented subjects with different decision trees. For some, the decision tree required them to choose future consumption immediately after lunch; others made this choice in the late afternoon. Because the individual is more likely to be hungry in the late afternoon than immediately after lunch, they interpret their experiment as indicating that temporary hunger affects choices of future consumption. However, for our purposes, that interpretation is immaterial. The important point is that the individual can be induced to choose different objects from the same set by modifying the decision tree used to make the selection (in particular, by delaying one component decision, e.g., from 1 P.M. to 4 P.M.). Thus, the decision tree is the ancillary condition.

Scope of the Framework

Our framework can incorporate nonstandard behavioral patterns in four separate ways. First, as discussed above, it allows for the influence of ancillary conditions on choice. Standard economics proceeds from the assumption that choice is invariant with respect to ancillary conditions. Positive behavioral economics challenges this basic premise. Documentation of a behavioral anomaly usually involves identifying some SCS, X, along with two ancillary conditions, d′ and d″, for which there is evidence that \( C(X, d') \neq C(X, d'') \). This is sometimes called a preference reversal, but in the interests of greater precision we will call it a choice reversal.

Second, our framework does not impose any choice axiom analogous to weak congruence. Hence, it allows for choice reversals based on “irrelevant alternatives,” as well as for intransitivities. For example, even when ancillary conditions are irrelevant, we might still have \( C(\{x, y\}) = \{x\} \), \( C(\{y, z\}) = \{y\} \), and \( C(\{x, z\}) = \{z\} \).

Third, our framework subsumes the possibility that people can make choices from opportunity sets that are not compact. For example, suppose we ask an individual to choose a dollar prize from the interval [0, 100). This set does not lie in the domain of a standard choice correspondence. And yet, one can easily imagine
someone making a choice from this set; he might be willing to choose any element of \([99.99, 100]\), on the grounds that any such payoff is good enough. In that case, we would have \(C((0, 100)) = [99.99, 100]\).

Fourth, we can interpret a choice object \(x \in X\) more broadly than in the standard framework. For example, if \(x\) is a lottery, we might want to allow for the possibility that anticipation is welfare relevant. In that case, the description of \(x\) would include information concerning the point in time at which uncertainty is resolved, as in Caplin and Leahy [2001].

**More on Positive Versus Normative Analysis**

In this chapter, we are concerned with normative analysis. We draw the same distinction between positive and normative analysis discussed as above, in the context of the standard model. In particular, we assume that choice data are available for some subset of the environments of interest, \(G^D \subset G\). The objective of positive economic analysis is to extend the choice correspondence \(C\) from observations on \(G^D\) to the entire set \(G\). As in standard economics, this may be accomplished by defining preferences over some appropriately defined set of objects, estimating these preferences using choice data drawn from sets in \(G^D\), and then using those estimated preferences to infer choices for GCS in \(G \setminus G^D\). However, a behavioral economist might also use other positive tools, such as models of choice algorithms, neural processes, or rules of thumb.

In conducting choice-based normative analysis, we take as given the individual’s choice correspondence, \(C\), defined on \(G\) rather than \(G^D\). The particular model used to extend \(C\)—whether it involves utility maximization or a decision algorithm—is irrelevant; for choice-based normative analysis, only \(C\) matters. Preferences and utility functions, which may (or may not) be used to extend the choice correspondence, are thus positive tools, not normative tools, just as in the standard framework. These constructs cannot meaningfully reconcile choice inconsistencies; they can only reiterate the information contained in the extended choice correspondence \(C\) (both the observed choices and the inferred choices). Thus, one cannot resolve normative puzzles by identifying classes of preferences that rationalize apparently inconsistent choices.4

**Individual Welfare**

In this section, we propose a general approach for extending standard choice-theoretic welfare analysis to situations in which individuals make “anomalous” choices of the various types commonly identified in behavioral research. We begin
by introducing two closely related binary relations, which will provide the basis for evaluating an individual’s welfare.

Individual Welfare Relations

Sometimes, welfare analysis involves the identification of an individual’s “best” alternative (e.g., when solving an optimal tax problem with a representative consumer). More often, however, it requires us to judge whether one alternative represents an improvement over another, even when the new alternative is not necessarily the best one. Identifying improvements is central both to the measurement of changes in individual welfare and to welfare analysis in settings with many people (both discussed above). It is also equivalent to the construction of a binary relation, call it $R$, where $xRy$ means that $x$ improves upon $y$. Accordingly, behavioral welfare analysis requires a binary relation analogous to revealed preference.

What is the appropriate generalization of the standard welfare relation, $R$? While there is a tendency in standard economics to define $R$ according to expression 7.1, that definition implicitly invokes the axiom of weak congruence, which assures that choices are consistent across different sets. To make the implications of that axiom explicit, it is useful to restate the standard definition of $R$ as follows:

$$xRy \text{ iff, for all } X \in \mathcal{X} \text{ with } x, y \in X, y \in C(X) \text{ implies } x \in C(X) \quad (7.4)$$

Similarly, we can define $P$, the asymmetric component of $R$, as follows:

$$xPy \text{ iff, for all } X \in \mathcal{X} \text{ with } x, y \in X, \text{ we have } y \notin C(X) \quad (7.5)$$

These alternative definitions of weak and strict revealed preference immediately suggest two natural generalizations. The first involves a straightforward generalization of weak revealed preference, as defined in relation (7.4):

$$xR'y \text{ iff, for all } (X, d) \in \mathcal{G} \text{ such that } x, y \in X, y \in C(X, d) \text{ implies } x \in C(X, d)$$

In other words, for any $x, y \in \mathcal{X}$, we say that $xR'y$ if, whenever $x$ and $y$ are available, $y$ is never chosen unless $x$ is as well.

As usual, we can define the symmetric and asymmetric components of $R'$. We say that $xP'y$ if $xR'y$ and ~ $yR'x$. The statement “$xP'y$” means that, whenever $x$ and $y$ are available, sometimes $x$ is chosen but not $y$, and otherwise either both or neither are chosen. Likewise, we can define $xI'y$ as $xR'y$ and $yR'x$. The statement “$xI'y$” means that, whenever $x$ is chosen, so is $y$, and vice versa.

The relation $P'$ generalizes the usual notion of strict revealed preference. However, within our framework, there is a more natural (and ultimately more useful)
generalization of relation 7.5: Specifically:

\[ xP^* y \text{ iff, for all } (X, d) \in \mathcal{G} \text{ such that } x, y \in X, \text{ we have } y \notin C(X, d) \]

In other words, for any \( x, y \in X \), we say that \( xP^* y \) iff, whenever \( x \) and \( y \) are available, \( y \) is never chosen. Corresponding to \( P^* \), there is an alternative notion of weak revealed preference:

\[ xR^* y \text{ iff, for some } (X, d) \in \mathcal{G} \text{ such that } x, y \in X, \text{ we have } x \in C(X, d) \]

The statement “\( xR^* y \)” means that, for any \( x, y \in X \), there is some GCS for which \( x \) and \( y \) are available and \( x \) is chosen. It is easy to check that \( P^* \) is the asymmetric component of \( R^* \); that is, \( xR^* y \) and \( yR^* x \) imply \( xP^* y \). Similarly, we can define the symmetric component of \( R^* \) as follows: \( xI^* y \) iff \( xR^* y \) and \( yR^* x \). The statement “\( xI^* y \)” means that there is at least one GCS for which \( x \) and \( y \) are available for which \( x \) is chosen, and at least one such GCS for which \( y \) is chosen. We note that Rubinstein and Salant (chapter 5) have separately proposed a binary relation that is related to \( P' \) and \( P^* \).

How are \( R' \), \( P' \), and \( I' \) related to \( R^* \), \( P^* \), and \( I^* \)? We say that a binary relation \( A \) is weakly coarser than another relation \( B \) if \( xAy \) implies \( xBy \). When \( A \) is weakly coarser than \( B \), we say that \( B \) is weakly finer than \( A \). It is easy to check that \( P^* \) is weakly coarser than \( P' \), that \( R' \) is weakly coarser than \( R^* \), and that \( I' \) is weakly coarser than \( I^* \).

\( R' \), \( P' \), and \( I' \) are more faithful to the standard notion of weak revealed preference, while \( R^* \), \( P^* \), and \( I^* \) are more faithful to the standard notion of strict revealed preference. Which of these two generalizations is most useful? Intuitively, since we are ultimately interested in identifying improvements, faithfulness to strict revealed preference may prove more important. However, the choice between these orderings should ultimately rest on their formal properties.

We begin with completeness. The relation \( R^* \) is obviously complete: for any \( x, y \in X \), we know that \( \{x, y\} \in \mathcal{X} \), and the individual must choose either \( x \) or \( y \) from any \( G = \{(x, y), d\} \). In contrast, \( R' \) need not be complete, as illustrated by Example 7.1.

**Example 7.1.**

If \( C(\{x, y\}, d') = \{x\} \) and \( C(\{x, y\}, d'') = \{y\} \), then we have neither \( xR'y \) nor \( yR'x \), so \( R' \) is incomplete.

Without further structure, \( R' \), \( P' \), \( R^* \), and \( P^* \) need not be transitive; example 7.2a shows shows that both \( P' \) and \( R' \) need not be transitive; example 7.2b makes the same point with respect to \( R^* \). Example 7.3, presented later for another purpose, shows that transitivity can also fail for \( P^* \).
Example 7.2a.
Suppose that $G = \{X_1, X_2, X_3, X_4\}$ with $X_1 = \{a, b\}$, $X_2 = \{b, c\}$, $X_3 = \{a, c\}$, and $X_4 = \{a, b, c\}$ (there are no ancillary conditions). Suppose also that $C(\{a, b\}) = \{a\}$, $C(\{b, c\}) = \{b\}$, $C(\{a, c\}) = \{c\}$, and $C(\{a, b, c\}) = \{a, b, c\}$. Then $aP'bP'cP'a$. Thus, $aP'b$ and $bP'c$, but $\not\sim aR'c$.

Example 7.2b.
Suppose that $G = \{X_1, X_2, X_3, X_4\}$ with $X_1 = \{a, b\}$, $X_2 = \{b, c\}$, $X_3 = \{a, c\}$, and $X_4 = \{a, b, c\}$ (there are no ancillary conditions). Suppose also that $C(\{a, b\}) = \{a\}$, $C(\{b, c\}) = \{b\}$, $C(\{a, c\}) = \{c\}$, and $C(\{a, b, c\}) = \{c\}$. Then $aR^*b$ and $bR^*c$, but $\not\sim aR^*c$.

Fortunately, to conduct useful welfare analysis—in particular, to identify maximal elements of arbitrary sets, and to measure improvements—one does not necessarily require transitivity. Our first main result establishes that there cannot be a cycle involving $R'$, the most natural generalization of weak revealed preferences, if even one of the links involves $P^*$, the most natural generalization of strict revealed preference. In other words, it generalizes the standard property that, if $x_1Rx_2 \ldots Rx_N$ with $x_iP_{i+1}$ for some $i$, then it is not the case that $x_NRx_1$.

Theorem 7.1.
Consider any $x_1, \ldots, x_N$ such that $x_iR'x_{i+1}$ for $i = 1, \ldots, N - 1$, with $x_kP^*x_{k+1}$ for some $k$. Then $\not\sim x_NR'x_1$.

This result has an immediate and important corollary:

Corollary 7.1.

$P^*$ is acyclic. That is, for any $x_1, \ldots, x_N$ such that $x_iP^*x_{i+1}$ for $i = 1, \ldots, N - 1$, we have $\not\sim x_NP^*x_1$.

Acyclicity is weaker than transitivity, but in most contexts it suffices to guarantee the existence of maximal elements.

It is worth emphasizing that theorem 7.1 holds under extremely weak assumptions; we require only that choice is well defined for all finite sets of alternatives. Regardless of how poorly behaved the choice correspondence $C$ might be, the binary relations $R'$ and $P^*$ are nevertheless well behaved in the sense of theorem 7.1.

Individual Welfare Optima
Both $P'$ and $P^*$ capture the notion of a welfare improvement, but $P^*$ leads to a more demanding notion than $P'$. Accordingly, we will say that is possible to strictly
improve upon a choice \( x \in X \) if there exists \( y \in X \) such that \( yP^x x \); in other words, there is an alternative that is unambiguously chosen over \( x \). We will say that it is possible to weakly improve upon a choice \( x \in X \) if there exists \( y \in X \) such that \( yP^x x \); in other words, there is an alternative that is sometimes chosen over \( x \), and that \( x \) is never chosen over (except in the sense that both could be chosen).

Our two different notions of welfare improvements lead to two separate concepts of individual welfare optima. When a strict improvement is impossible, we say that \( x \) is a weak individual welfare optimum. In contrast, when a weak improvement is impossible, we say that \( x \) is a strict individual welfare optimum.

When is \( x \in X \) an individual welfare optimum? The following simple observations address this question:

**Observation 7.1.**

If \( x \in C(X,d) \) for some \((X,d) \in \mathcal{G}\), then \( x \) is a weak individual welfare optimum in \( X \). If \( x \) is the unique element of \( C(X,d) \), then \( x \) is a strict welfare optimum in \( X \).

This first observation assures us that our notions of individual welfare optima respect the most obvious implication of libertarian deference to voluntary choice: any action voluntarily chosen from a set \( X \) under some ancillary condition is a weak individual welfare optimum within \( X \). Moreover, any action that the individual uniquely chooses from \( X \) under some condition is a strict individual welfare optimum within \( X \).

As a general matter, alternatives chosen from \( X \) need not be the only individual welfare optima within \( X \). Observation 7.2 characterizes the set of individual welfare optima more precisely:

**Observation 7.2.**

Suppose that, for all \( a, b \in X \), we have \( \{a, b\} \in \mathcal{X} \). Then \( x \) is a weak individual welfare optimum in \( X \) if and only if for each \( y \in X \) (other than \( x \)), there is some GCS for which \( x \) is chosen with \( y \) available (\( y \) may be chosen as well). Moreover, \( x \) is a strict individual welfare optimum in \( X \) if and only if for each \( y \in X \) (other than \( x \)), either \( x \) is chosen and \( y \) is not for some GCS with \( y \) available, or there is no GCS for which \( y \) is chosen and \( x \) is not with \( x \) available.

According to observation 7.2, some alternative \( x \) may be an individual welfare optimum for the set \( X \) even though there is no ancillary condition \( d \) under which \( x \in C(X,d) \) (see example 7.5 below). Note, however, that this is still consistent with the spirit of the libertarian principle: \( x \) is chosen over every \( y \in X \) in some circumstances, though not necessarily ones involving choices from \( X \). In contrast,
an alternative $x$ that is never chosen over some alternative $y$ in the set $X$ cannot be an individual welfare optimum in $X$.

Example 7.3 below illustrates why it may be unreasonable to exclude the type of individual welfare optima described in the preceding paragraph. (See further below for another more formal argument.) Suppose a subject chooses a free sample of strawberry jam when only two other flavors are available, but feels overwhelmed and elects not to receive a free sample when 30 flavors (including strawberry) are available. Since the individual might not want the planner to act overwhelmed when choosing on her behalf, it is important to allow for the possibility that the planner should pick strawberry jam on her behalf even when 30 alternatives are available. Similar concerns would arise whenever thinking about $X$ causes the individual to experience feelings (e.g., temptation) that affect her choice from $X$, and that vanish when the decision is delegated to a planner. Since we are confining ourselves at this juncture to choice evidence, we do not take a position as to whether these considerations are present; rather, we avoid adopting a notion of individual welfare optima that assumes away such possibilities.

Notice that observation 7.1 guarantees the existence of weak welfare optima (but not of strict welfare optima). Thus, the existence of a solution to the planner’s policy problem is guaranteed even in situations where the individuals make conflicting choices across ancillary conditions.

The fact that we have established existence without making any additional assumptions, for example, related to continuity and compactness, may at first seem confusing, but this is simply a matter of how we have posed the question. Here, we have assumed that the choice function is well defined over the set $G$—this is treated as data. Standard existence issues arise when the choice function is built up from other components. The following two examples clarify these issues.

Example 7.3.

Suppose that $G = \{X_1, X_2, X_3, X_4\}$ with $X_1 = \{a, b\}$, $X_2 = \{b, c\}$, $X_3 = \{a, c\}$, and $X_4 = \{a, b, c\}$ (there are no ancillary conditions). Imagine that the individual chooses $a$ from $X_1$, $b$ from $X_2$, $c$ from $X_3$, and $a$ from $X_4$. In this case, we have $aP^* b$ and $bP^* c$; in contrast, we can only say that $aI^* c$. So, despite the intransitivity of choice between the sets $X_1$, $X_2$, and $X_3$, the option $a$ is nevertheless a strict welfare optimum in $X_4$, and neither $b$ nor $c$ is a weak welfare optimum. Note that $a$ is also a strict welfare optimum in $X_1$ ($b$ is not a weak optimum), $b$ is a strict welfare optimum in $X_2$ ($c$ is not a weak optimum), and both $a$ and $c$ are strict welfare optima in $X_3$ ($a$ survives because it is chosen over $c$ in $X_4$, which makes $a$ and $c$ not comparable under $P^*$).

Example 7.4.

Consider the same choice data as in example 7.3, but suppose we limit attention to $G' = \{X_1, X_2, X_3\}$. In this case we have that $aP^* bP^* cP^* a$. Here,
the intransitivity is apparent; $P^*$ is cyclic because the assumption leading to
Theorem 1 is violated ($G'$ does not contain all finite sets). Even so, individual
welfare optima exist within every set that falls within the restricted domain: $a$
is a strict welfare optimum in $X_1$, $b$ is a strict welfare optimum in $X_2$, and $c$
is a strict welfare optimum in $X_3$. Naturally, if we are interested in creating a
preference or utility representation based on the data contained in $G'$ in order
to project the individual's choice from $X_4$, the intransitivity would pose a
difficulty. And if we try to make a welfare judgement concerning $X_4$ without
knowing (either directly or through a positive model) what the individual
would choose in $X_4$, we encounter the same problem—$a$, $b$, and $c$ are all
strictly improvable, so there is no welfare optimum. But once we know what
the individual would do in $X_4$ (either directly or by extrapolating from a
reliable positive model), the existence problem for $X_4$ vanishes. It is therefore
important to emphasize again that our interest here is in forming welfare
judgements from individual choices, not in the problem of representing or
extending those choices to unobserved domains. We are in effect assuming
that an adequate positive model of behavior already exists, and we are asking
how normative analysis should proceed.

Depending on the nature and extent of choice conflicts, the welfare criteria
proposed here may not be particularly discerning. Example 7.5 provides an
illustration:

Example 7.5.

Suppose that $X = \{X_1, X_2, X_3, X_4\}$ (defined in example 7.4), and that $G = X 
\times \{d, d'\}$. Suppose that, with ancillary condition $d$, $b$ is never chosen when $a$
is available, and $c$ is never chosen. However, with ancillary condition $d'$, $b$
is never chosen with $c$ available, and $a$ is never chosen. Then there no
alternatives are comparable with $P'$ or $P^*$, and the set of individual welfare
optima (weak and strict) in $X_i$ is simply $X_i$, for $i = 1, 2, 3, 4$.

In Example 7.5, two ancillary conditions produce diametrically opposed choice
patterns. In most practical situations, the amount of choice conflict, and hence
the sets of individual welfare optima, will be smaller. Generally, with greater
choice conflict, it becomes more difficult to identify alternatives that consti-
tute unambiguous welfare improvements, so the set of individual welfare optima
expands.

Why This Approach?

It is natural to wonder whether there is some other, potentially more attrac-
tive approach to formulating a choice-theoretic foundation for behavioral welfare
analysis. In this section, we provide further formal justifications for our approach.
Consider the following natural alternative to our approach: classify $x$ as an individual welfare optimum for $X$ iff there is some ancillary condition for which the individual is willing to choose $x$ from $X$. This alternative approach would appear to adhere more closely to the libertarian principle than does our approach. However, it does not allow us to determine whether a change from one element of $X$ to another is an *improvement*, except in cases where either the initial or final element in the comparison is one that the individual would choose from $X$. As explained at the outset of this section, for that purpose we require a binary relation that identifies improvements. Accordingly, our object in this section is determine whether there exists a general method of constructing an asymmetric binary welfare relation, $Q$, that is more faithful to the libertarian principle than the relations proposed above.

We will say that $Q$ is an *inclusive libertarian relation* for a choice correspondence $C$ if, for all $X$, the maximal elements under $Q$ include all of the elements the individual would choose from $X$ for some ancillary condition. We will say that $Q$ is an *exclusive libertarian relation* for a choice correspondence $C$ if, for all $X$, the maximal elements under $Q$ are contained in the set of elements the individual would choose from $X$ for some ancillary condition. Finally, we will say that $Q$ is a *libertarian relation* for $C$ if it is both inclusive and exclusive—that is, if the maximal elements under $Q$ always coincide exactly with the set of elements the individual would choose from $X$ for some ancillary condition.6

We have already demonstrated that $P^*$ is always an inclusive libertarian relation (observation 7.1). We have also argued, by way of example, that there are good reasons to treat the “extra” maximal elements under $P^*$—the ones not chosen from the set of interest for any ancillary condition—as individual welfare optima. However, the following result shows that there is an even more compelling reason not to search for a general procedure that generates either a libertarian relation, or an exclusive libertarian relation, for all choice correspondences: none exists.

**Theorem 7.2.**

*For some choice correspondences, exclusive libertarian relations do not exist.*

Theorem 7.2 implies that, if we want to derive a completely general procedure for expressing libertarianism, we must confine attention to inclusive libertarian relations. There are, of course, inclusive libertarian relations other than $P^*$. For example, the null relation, $R_{\text{Null}}$ ($\sim x R_{\text{Null}} y$ for all $x, y \in X$), falls into this category; for any set $X$, the maximal elements under $R_{\text{Null}}$ consist of $X$, which of course includes all of the chosen elements. Yet $R_{\text{Null}}$ is far less discerning, and further from the libertarian principle, than $P^*$. In fact, one can prove the following result:
Consider any choice correspondence $C$, and any inclusive libertarian relation $Q \neq P^\ast$. Then $P^\ast$ is finer than $Q$.

It follows that for all choice correspondences $C$ and choice sets $X$, the set of maximal elements in $X$ under $P^\ast$ is (weakly) smaller than the maximal elements in $X$ under $Q$. Thus, $P^\ast$ is always the most discriminating inclusive libertarian relation.

Relation to Multi-Self Pareto Optima

Our notion of an individual welfare optimum is related to the idea of a multi-self Pareto optimum, which has been used as a welfare criterion in a number of behavioral studies [see, e.g., Laibson, Repetto, and Tobacman, 1998; Bhattacharya and Lakdawalla, 2004]. Suppose in particular that the set of GCSs is the Cartesian product of the set of SCSs and a set of ancillary conditions (i.e., $G = \mathcal{X} \times D$, where $d \in D$); in that case, we say that $G$ is rectangular. Imagine also that, for each $d \in D$, the choice correspondence is consistent with the maximization of a well-behaved preference ranking $R_d$. If one imagines that each ancillary condition activates a different “self,” then one can conduct welfare analysis by examining multi-self Pareto optima.

Observation 7.3.

Under the stated conditions, a weak multi-self Pareto optimum corresponds to a weak individual welfare optimum (as we have defined it), and a strict multi-self Pareto optimum corresponds to a strict individual welfare optimum.

The multi-self Pareto criterion has been used primarily in the literature on quasi-hyperbolic discounting, where it is applied to an individual’s many time-dated “selves” (as in the studies identified above). Ironically, our framework does not justify the multi-self Pareto criterion for quasi-hyperbolic consumers, because $G$ is not rectangular (see below).

In contrast, our framework does justify the use of the multi-self Pareto criterion for cases of “coherent arbitrariness,” such as those studied by Ariely, Loewenstein, and Prelec [2003]. In that context, $d$ is some psychological anchor. Although the anchor affects behavior, the individual conforms to standard consumer theory for any fixed anchor. To our knowledge, the multi-self Pareto criterion has not been proposed as a natural welfare standard in the presence of coherent arbitrariness.

For the narrow settings that are consistent with the assumptions described at the beginning of this section, one can view our approach as a way to justify the multi-self Pareto criterion without relying on untested and questionable psychological...
assumptions. Note, however, that while our framework justifies the application of the multi-self Pareto criterion in such contexts, the justification is choice-theoretic, not psychological. Furthermore, our approach is more general in that it does not require the GCS to be rectangular, nor the choice correspondence within each ancillary condition to be well behaved.

Some Applications

In this section, we examine the implications of our framework for some particular behavioral anomalies.

Coherent Arbitrariness

Let’s consider a case in which an individual consumes two goods, $y$ and $z$. Suppose that positive analysis delivers the following utility representation:

$$U(y, z | d) = u(y) + dv(z),$$

with $u$ and $v$ strictly increasing, differentiable, and strictly concave. Notice that the ancillary condition, $d \in [d_L, d_H]$, which we interpret here as an irrelevant signal, simply shifts the weight on “utility” from $z$ to $y$. Given any particular signal, the individual behaves coherently, but his behavior is arbitrary in the sense that it depends on the signal. This type of behavior (coherent arbitrariness) has been documented by Ariely, Loewenstein, and Prelec [2003] and has led some to question the practice of basing welfare judgments on revealed preference.

Our normative framework easily accommodates this positive model of behavior. In fact, since $G$ is rectangular, our welfare criterion is equivalent to the multi-self Pareto criterion, where each $d$ indexes a different “self.”

We know that, if $u(y') + dv(z') \geq u(y'') + dv(z'')$ for all $d \in [d_L, d_H]$, then the individual will never be willing to choose $(y'', z'')$ without also being willing to choose $(y', z')$ when both are available. We can rewrite that inequality as follows:

$$u(y') - u(y'') \geq d[v(z'') - v(z')]$$

(7.6)

Notice that, if relation 7.6 holds for $d_L$ and $d_H$, then it holds for all $d \in [d_L, d_H]$. Therefore,

$$(y', z')R'(y'', z'') \text{ iff } u(y') + dv(z') \geq u(y'') + dv(z'') \text{ for all } d \in \{d_L, d_H\}$$

(7.7)

Replacing the weak inequality with a strict inequality, we obtain the definition of both $P'$ and $P^*$. Replacing “for all” with “for any,” we obtain the definition of $R^*$. 

For a graphical illustration, see figure 7.1(a). We have drawn two “indifference curves” through the bundle \((y', z')\), one for \(d_L\) (labeled \(I_L\)) and one for \(d_H\) (labeled \(I_H\)). For all bundles \((y'', z'')\) lying below both indifference curves, we have \((y', z')P^*(y'', z'')\); this is the analog of a lower contour set. Conversely, for all bundles \((y'', z'')\) lying above both indifference curves, we have \((y'', z'')P^*(y', z')\); this is the analog of an upper contour set. For all bundles \((y'', z'')\) lying between the two indifference curves, we have neither \((y', z')P^*(y'', z'')\) nor \((y'', z'')P^*(y', z')\); however, \((y', z')I^*(y'', z'')\).

Now consider a standard budget constraint, \(X = \{(y, z) \mid y + pz \leq M\}\), where \(y\) is the numeraire, \(p\) is the price of \(z\), and \(M\) is income. The set \(X\) corresponds to the triangle in figure 7.1(b). The individual’s choice from this set clearly depends on the ancillary condition \(d\). In particular, she chooses bundle \(a\) when the ancillary condition is \(d_H\), and bundle \(b\) when the ancillary condition is \(d_L\). Each of the points on the darkened segment of the budget line between bundles \(a\) and \(b\) is uniquely chosen for some \(d \in [d_L, d_H]\), so all of these bundles are strict individual welfare optima. In this case, there are no other welfare optima, weak or strict. Consider any other bundle \((y', z')\) on or below the budget line; if it lies to the northwest of \(a\), then \(aP^*(y', z')\); if it lies to the southwest of \(b\), then \(bP^*(y', z')\); and if it lies anywhere else below the budget line, then \(xP^*(y', z')\) for some \(x\) containing more of both goods than \((y', z')\).

**Dynamic Inconsistency**

The \(\beta, \delta\) model of hyperbolic discounting popularized by Laibson [1997] and O'Donoghue and Rabin [1999] has come into widespread use among economists. When our framework is applied to this positive model, what welfare criterion emerges? Here we focus on decisions involving the allocation of consumption.
new directions for normative economics

over three periods; see Bernheim and Rangel [2007] for an analysis of the general multiperiod case.

The consumer’s task is to choose a consumption vector, $c = (c_1, c_2, c_3)$, where $c_t$ denotes the level of consumption at time $t$. The positive model functions as follows.

At $t = 1$, all discretion is resolved to maximize the function

$$U_1(c) = u(c_1) + \beta [\delta u(c_2) + \delta^2 u(c_3)],$$

assuming perfect foresight with respect to continuation decisions (if any). At $t = 2$, all remaining discretion is resolved to maximize the function

$$U_2(c) = u(c_2) + \beta \delta u(c_3),$$

again assuming perfect foresight with respect to continuation decisions (if any). Finally, at $t = 3$, all remaining discretion is resolved to maximize the function

$$U_3(c) = u(c_3).$$

We assume that $0 < \beta \leq 1$ and $0 < \delta \leq 1$. We also assume that $u(0)$ is finite [and, for convenience, we normalize $u(0) = 0$].

To conduct normative analysis, we must recast this positive model as a correspondence from GCs into consumption vectors. Here, $X$ is a set of intertemporal consumption bundles; an intertemporal budget constraint gives rise to a particular $X \subset X$. An ancillary condition describes the decision tree used to choose the consumption vector. Clearly, $c_1$ must be chosen at $t = 1$. However, a decision at $t = 1$ may determine the constraint set encountered at $t = 2$. Likewise, any remaining discretion concerning $c_2$ must be resolved at $t = 2$; in addition, a decision at $t = 2$ may determine the constraint set encountered at $t = 3$. Finally, any remaining discretion concerning $c_3$ must be resolved at $t = 3$. We allow for all conceivable decision trees. Note that, in this instance, $G$ is not rectangular.

Next we define the function

$$W(c) = u(c_1) + \beta \delta u(c_2) + (\beta \delta)^2 u(c_3).$$

This represents lifetime discounted utility using $\beta \delta$ as the time-consistent discount factor. Given our normalization $|u(0) = 0|$, it follows immediately that $W(c) < U_1(c)$ iff $c_3 > 0$, and $W(c) = U_1(c)$ iff $c_3 = 0$. In particular, it is never the case that $W(c) > U_1(c)$.

For this simple model, to determine whether $c$ is unambiguously chosen over $c'$ (and hence whether the welfare relation ranks $c$ above $c'$), we compare the lifetime discounted utility associated with $c$ using $\beta \delta$ as a time-consistent discount factor, with the “decision utility” associated with $c'$ at $t = 1$. Formally:
Theorem 7.4.

(i) \( cR^c c' \) iff \( W(c) \geq U_1(c') \),
(ii) \( cR^* c' \) iff \( U_1(c) \geq W(c') \),
(iii) \( cP^* c' \) iff \( W(c) > U_1(c') \),

(iiv) \( P' = R' \), and \( P', \ P^*, \ and \ P^* \) are transitive.

Using this result, we can easily characterize the set of individual welfare optima within \( X \). For any \( X \), let

\[
W^*(X) = \max_{c \in X} W(c).
\]

Then \( c \) is a weak individual welfare optimum in \( X \) iff \( U_1(c) \geq W^*(X) \)—in other words, if the “decision utility” that \( c \) provides at \( t = 1 \) is at least as large as the highest available discounted utility, using \( \beta \delta \) as a time-consistent discount factor. Given that \( W(c) \leq U_1(c) \) for all \( c \), we know that \( W^*(c) \leq \max_{c \in X} U_1(c) \), so the set of weak individual welfare optima is plainly nonempty (as we already knew it must be).

Notice that, for all \( c \), we have \( \lim_{\beta \to 1} [W(c) - U_1(c)] = 0 \). Accordingly, as the degree of dynamic inconsistency shrinks, our welfare criterion converges to the standard criterion. In contrast, the same statement does not hold for the multi-self Pareto criterion, as that criterion is usually formulated. The reason is that, regardless of \( \beta \), each “self” is assumed to care only about current and future consumption. Thus, consuming everything in the final period is always a multi-self Pareto optimum, even when \( \beta = 1 \).

The Standard Framework as a Limiting Case

Clearly, our framework for welfare analysis subsumes the standard framework; when the choice correspondence satisfies standard axioms, the generalized individual welfare relations coincide with revealed preference. Our framework is a natural generalization of the standard welfare framework in another important sense: when behavioral departures from the standard model are small, our welfare criterion is close to the standard criterion. This conclusion, which plainly holds in the applications considered above, is proven with generality in Bernheim and Rangel [2007].

The preceding conclusion is important for two reasons. First, it offers a formal justification for using the standard welfare framework (as an approximation) when choice anomalies are known to be small. Many economists currently adopt the premise that anomalies are small when using the standard framework; they view this as a justification both for standard positive analysis and for standard normative
new directions for normative economics

In the case of positive analysis, their justification is clear: if we compare the actual choices to predictions generated from a standard positive model and discover that they are close to each other, we can conclude that the model involves little error. However, in the case of normative analysis, their justification for the standard approach is problematic. To conclude that the standard normative criterion is roughly correct in a setting with choice anomalies, we would need to compare it to the correct criterion. But if we don’t know the correct criteria for such settings, then we have no benchmark against which to gauge the performance of the standard criterion. As a result, we cannot measure the distance between the standard normative criterion and the correct criterion, even when choice anomalies are tiny. Our framework overcomes this problem by providing welfare criteria for all situations, including those with choice anomalies. One can then ask whether the criterion changes much if one ignores the anomalies. In this way, our analysis formalizes the intuition that a little bit of positive falsification is unimportant from a normative perspective.

Second, our limiting results imply that the debate over the significance of choice anomalies need not be resolved prior to adopting a framework for welfare analysis. If our framework is adopted and the anomalies ultimately prove to be small, one will obtain virtually the same answer as with the standard framework. (For the reasons described above, the same statement does not hold for the multi-self Pareto criterion.)

Tools for Applied Welfare Analysis

The concepts of compensating variation and equivalent variation are central to applied welfare economics. In this section we show that they have natural counterparts within our framework. Here, we focus on compensating variation; the treatment of equivalent variation is analogous. We illustrate how, under more restrictive assumptions, the compensating variation of a price change corresponds to an analog of consumer surplus.

Compensating Variation

Let’s assume that the individual’s SCS, $X(\alpha, m)$, depends on a vector of environmental parameters, $\alpha$, and a monetary transfer, $m$. Let $\alpha_0$ be the initial parameter vector, $d_0$ the initial ancillary conditions, and $(X(\alpha_0, 0), d_0)$ the initial GCS. We will consider a change in parameters to $\alpha_1$, coupled with a change in ancillary conditions to $d_1$, as well as a monetary transfer $m$. We write the new GCS as $(X(\alpha_1, m), d_1)$. This setting will allow us to evaluate compensating variations for fixed changes in prices, ancillary conditions, or both.8
Within the standard economic framework, the compensating variation is the smallest value of $m$ such that for any $x \in C[X(\alpha_0, 0)]$ and $y \in C[X(\alpha_1, m)]$, the individual would be willing to choose $y$ in a binary comparison with $x$ [i.e., $y \in C(\{x, y\})$, or equivalently, $yRx$].

In extending this definition to our framework, we encounter three ambiguities. The first arises when the individual is willing to choose more than one alternative in either the initial GCS $X(\alpha_0, 0)$ or in the final GCS $X(\alpha_1, m)$. In the standard framework, this causes no difficulty because the individual must be indifferent among all alternatives chosen from the same set. However, within our framework, these alternatives may fare differently in comparison to other alternatives. Here, we handle this ambiguity by insisting that compensation is adequate for all pairs of outcomes that might be chosen voluntarily from the initial and final sets.

The second dimension of ambiguity arises from the possibility that additional compensation could in principle reduce an individual’s welfare. In the standard framework, if the payment $m$ is adequate to compensate an individual for some change, then any $m' > m$ is also adequate. Without further assumptions, that property need not hold in our framework. Here, we will simplify matters by assuming that it does hold; see Bernheim and Rangel [2007] for an analysis of the more general case.

The third dimension of ambiguity concerns the standard of compensation: do we consider compensation sufficient when the new situation (with the compensation) is unambiguously chosen over the old one, or when the old situation is not unambiguously chosen over the new one? This ambiguity is an essential feature of welfare evaluations with inconsistent choice (see example 7.6, below). Accordingly, we define two notions of compensating variation:

**Definition 7.1.**

CV-A is the level of compensation $m^A$ that solves

$$\inf \{ m \mid yP^*x \text{ for all } x \in C[X(\alpha_0, 0), d_0] \text{ and } y \in C[X(\alpha_1, m), d_1] \}$$

**Definition 7.2.**

CV-B is the level of compensation $m^B$ that solves

$$\sup \{ m \mid xP^*y \text{ for all } x \in C[X(\alpha_0, 0), d_0] \text{ and } y \in C[X(\alpha_1, m), d_1] \}$$

In other words, a level of compensation slightly greater than CV-A guarantees that any alternative selected in the new choice situation (with the compensation) is unambiguously chosen over any alternative selected in the initial situation. Similarly, a level of compensation slightly smaller than CV-B guarantees that any alternative
selected in the initial situation is unambiguously chosen over any alternative selected in the new situation (with the compensation).

CV-A and CV-B are well-behaved measures of compensating variation in the following sense: If the individual experiences a sequence of changes, and is adequately compensated for each of these changes in the sense of CV-A, no alternative that he would select from the initial set is unambiguously chosen over any alternative that he would select from the final set. Similarly, if he experiences a sequence of changes and is not adequately compensated for any of them in the sense of CV-B, no alternative that he would select from the final set is unambiguously chosen over any alternative that he would select from the initial set. Both of these conclusions are corollaries of theorem 7.1. However, in contrast to the standard framework, the compensating variations (either CV-As or CV-Bs) associated with each step in a sequence of changes need not be additive. This is not necessarily a serious problem; see Bernheim and Rangel [2007] for further discussion.

Example 7.6.

Let’s revisit the application discussed above regarding coherent arbitrariness. Suppose the individual is offered the following degenerate opportunity sets: \( X(0, 0) = \{(y_0, z_0)\} \), and \( X(1, m) = \{(y_1 + m, z_1)\} \). In other words, changing the environmental parameter \( \alpha \) from 0 to 1 shifts the individual from \((y_0, z_0)\) to \((y_1, z_1)\), and compensation is paid in the form of the good \( y \). Figure 7.2 depicts the bundles \((y_0, z_0)\) and \((y_1, z_1)\), as well as the CV-A and CV-B for this change. CV-A is given by the horizontal distance \((y_1, z_1)\) and point \( a \), because

![Figure 7.2. CV-A and CV-B for example 7.6](image-url)
(y_1 + m^A + \varepsilon, z_1) is chosen over (x_0, m_0) for all ancillary conditions and \varepsilon > 0. CV-B is given by the horizontal distance between (y_1, z_1) and point b, because (y_0, z_0) is chosen over (y_1 + m_B - \varepsilon, z_1) for all ancillary conditions and \varepsilon > 0. Note, however, that for intermediate levels of compensation, (y_1 + m, z_1) is chosen under some ancillary conditions, and (y_0, z_0) is chosen under others. Finally, note that as \delta_H converges to \delta_L, both notions of compensating variation converge to the standard notion.

**Consumer Surplus**

Next we illustrate how, under more restrictive assumptions, the compensating variation of a price change corresponds to an analog of consumer surplus. We continue to study the environment introduced above and revisited in example 7.6. However, we will assume here that positive analysis delivers the following more restrictive utility representation (which involves no income effects, so that Marshallian consumer surplus would be valid in the standard framework):

\[ U(y, z \mid d) = y + dv(z) \]

Thus, for any given \( d \), the inverse demand curve for \( z \) is given by \( p = dv'(z) \), where \( p \) is the relative price of \( z \).

Let \( M \) denote the consumer’s initial income. Consider an increase in the price of \( z \) from \( p_0 \) to \( p_1 \) while holding the ancillary conditions constant, \( d_0 = d_1 \). Let \( z_0 \) denote the amount of \( z \) purchased at price \( p_0 \), and let \( z_1 \) denote the amount purchased at price \( z_1 \); given our assumptions, \( z_0 > z_1 \). Since there are no income effects, \( z_1 \) will not change as the individual is compensated (holding the ancillary condition fixed).

Through straightforward calculations, one can obtain the following expression for CV-A:

\[ m^A = [p_1 - p_0]z_1 + \int_{z_1}^{z_0} [d_Hv'(z) - p_0]dz \]

The first term measures the extra amount the consumer ends up paying for the first \( z_1 \) units. The second term is the area under the demand curve and above a horizontal line at \( p_0 \), when \( d_H \) is the ancillary condition. Figure 7.3(a) provides a graphical representation of CV-A, analogous to the one found in most microeconomics textbooks: it is the sum of the areas labeled A, B, and C.

Similarly, it is easily shown that CV-B is given by

\[ m^B = [p_1 - p_0]z_1 + \int_{z_1}^{z_0} [d_Lv'(z) - p_0]dz. \]
Notice that this is the same as the expression for CV-A, except that we use the area under the demand curve associated with \( d_L \), rather than the one associated with \( d_H \). Figure 7.3(b) provides a graphical representation of CV-B: it is the sum of the areas labeled A and D, minus the areas labeled E.

As figure 7.3 illustrates, CV-A and CV-B bracket the conventional measure of consumer surplus that one would obtain using the demand curve associated with the ancillary condition \( d_0 \). In addition, as the range of possible ancillary conditions narrows, CV-A and CV-B both converge to standard consumer surplus. This underscores the fact that the standard framework is a special case of the framework considered here. Moreover, it also implies that, when inconsistencies are minor (i.e., \( d_H - d_L \) is small), the ambiguity in welfare, as measured by the difference between CV-A and CV-B, is small.

The compensating variation associated with a change in ancillary conditions, from \( d_0 \) to \( d_1 \neq d_0 \), with fixed prices, is calculated in a similar way. For the purpose
of illustration, assume that \( z_1 > z_0 \). Figure 7.4 shows CV-A and CV-B graphically. CV-A, the light gray area between the demand curve for the ancillary condition \( d_H \) and a horizontal line drawn at the price of \( z \), is strictly positive. In contrast, CV-B, the dark gray area between a horizontal line drawn at the price of \( z \) and the demand curve for the ancillary condition \( d_L \), is strictly negative. Once more, the ambiguity in welfare is minor when the behavioral inconsistencies between \( d_H \) and \( d_L \) are small.

**Welfare Analysis Involving More Than One Individual**

In settings with more than one individual, welfare analysis often focuses on the concept of Pareto optimality. In the standard framework, we say that a social alternative \( x \in X \) is a Pareto optimum in \( X \) if there is no other alternative that all individuals would choose over \( x \). In this section we describe a natural generalization of this concept to settings with behavioral anomalies, and we illustrate its use in establishing the efficiency of competitive market equilibria. For a more comprehensive discussion of these concepts, see Bernheim and Rangel [2007].

**Generalized Pareto Optima**

Suppose there are \( N \) individuals indexed \( i = 1, \ldots, N \). Let \( \mathcal{X} \) denote the set of all conceivable social choice objects, and let \( X \) denote the set of feasible objects. Let \( C_i \) be the choice function for individual \( i \), defined over \( G_i \) (where the subscript reflects the possibility that the set of ancillary conditions may differ from individual to individual). These choice functions induce the relations \( R'_i \) and \( P^*_i \) over \( \mathcal{X} \).

We say that \( x \) is a *weak generalized Pareto optimum* in \( X \) if there exists no \( y \in X \) with \( yP^*_i x \) for all \( i \). We say that \( x \) is a *strict generalized Pareto optimum* in \( X \) if there exists no \( y \in X \) with \( yR'_i x \) for all \( i \), and \( yP^*_i x \) for some \( i \).\(^{11}\)

Since strict individual welfare optima do not always exist, we cannot guarantee the existence of strict generalized Pareto optima with a high degree of generality. However, we can trivially guarantee the existence of a weak generalized Pareto optimum for any set \( X \): simply choose \( x \in C_i(X, d) \) for some \( i \) and \( (X, d) \in G \) (in which case we have \( \sim[yP^*_i x \text{ for all } y \in X] \)).

In the standard framework, there is typically a continuum of Pareto optima that spans the gap between the extreme cases in which the chosen alternative is optimal for some individual. We often represent this continuum by drawing a utility possibility frontier or, in the case of a two-person exchange economy, a contract
Figure 7.5. The generalized contract curve

curve. Is there also usually a continuum of generalized Pareto optima spanning the gap between the extreme cases described in the preceding paragraph? Example 7.7 answers this question in the context of a two-person exchange economy.

Example 7.7.

Consider a two-person exchange economy involving two goods, \( y \) and \( z \). Suppose the choices of consumer 1 are described by the positive model set forth in the discussion above concerning coherent arbitrariness, while consumer 2’s choices respect standard axioms. In figure 7.5, the area between the curves labeled \( T_H \) (formed by the tangencies between the consumer’s indifference curves when consumer 1 faces ancillary condition \( d_H \)) and \( T_L \) (formed by the tangencies when consumer 1 faces ancillary condition \( d_L \)) is the analog of the standard contract curve; it contains all of the weak generalized Pareto optimal allocations. The ambiguities in consumer 1’s choices expand the set of Pareto optima, which is why the generalized contract curve is thick. Like a standard contract curve, the generalized contract curve runs between the southwest and northeast corners of the Edgeworth box, so there are many intermediate Pareto optima. If the behavioral effects of the ancillary conditions were smaller, the generalized contract curve would be thinner; in the limit, it would converge to a standard contract curve. Thus, the standard framework once again emerges as a limiting case of our framework, in which behavioral anomalies become vanishingly small.

More generally, in standard settings (with continuous preferences and a compact set of social alternatives \( X \)), one can start with \textit{any} alternative \( x \in X \) and find a
Pareto optimum in \( \{ y \mid y \not\leq x \text{ for all } i \} \)—for example, by identifying some individual’s most preferred alternative within that set. Indeed, by doing so for all \( x \in X \), one generates the contract curve. Our next theorem establishes an analogous result for weak generalized Pareto optima.

Theorem 7.5.

For every \( x \in X \), the nonempty set \( \{ y \in X \mid \forall i, \sim xP^*_iy \} \) includes at least one weak generalized Pareto optimum in \( X \).

Notice that theorem 7.5 does not require additional assumptions concerning compactness or continuity. Rather, existence follows from the fundamental assumption that the choice correspondence is nonempty over its domain.\textsuperscript{12}

The Efficiency of Competitive Equilibria

To illustrate the usefulness of these concepts, in Bernheim and Rangel [2007] we provide a generalization of the First Welfare Theorem. Specifically, we consider a production economy consisting of \( N \) consumers, \( F \) firms, and \( K \) goods. The economy is standard in all respects, except that consumer \( i \)’s behavior is governed by a general choice correspondence mapping budget sets and ancillary conditions into sets of consumption vectors. We make one simple assumption (akin to nonsatiation) with respect to consumer behavior: if \( x^n > w^n \) (where \( > \) indicates a strict inequality for every good), then consumer \( n \) never chooses \( w^n \) when \( x^n \) is available.

A behavioral competitive equilibrium involves a price vector, \( \pi = (\pi^1, \ldots, \pi^K) \), along with a vector of ancillary conditions \( \dd = (\dd^1, \ldots, \dd^N) \), that clear all markets. Although behavioral competitive equilibria may not exist, those that do exist are necessarily efficient:

Theorem 7.6.

The allocation in any behavioral competitive equilibrium is a strict generalized Pareto optimum.

It is worth emphasizing that a perfectly competitive equilibrium may be inefficient when judged by a refined welfare relation, after officiating choice conflicts, as described in the next section. This observation alerts us to the fact that, in behavioral economics, choice inconsistencies lead to a new class of potential market failures. It is also possible to demonstrate the inefficiency of market equilibria according to the welfare relations we have proposed in the presence of sufficiently severe
but otherwise standard market failures, such as externalities [see Bernheim and Rangel, 2007].

**Refining the Welfare Criterion Using Nonchoice Data**

We have shown that the individual welfare orderings \( R' \), \( P' \), \( R^* \), and \( P^* \) may not be very discerning in the sense that many alternatives may not be comparable and the set of individual welfare optima may be large. This tends to occur when there are significant conflicts between the choices made under different ancillary conditions.

In this section we consider the possibility that one might refine these relations by altering the data used to construct them, either by adding new choice data, or by deleting data. We also discuss the types of evidence that could be useful in making this type of refinements.

### Refinement Strategies

The following simple observation indicates how the addition or deletion of data affects the coarseness of the welfare relation.

**Observation 7.4.**

Fix \( \mathcal{X} \). Consider two generalized choice domains \( \mathcal{G}_1 \) and \( \mathcal{G}_2 \) with \( \mathcal{G}_1 \subset \mathcal{G}_2 \), where \( \mathcal{X}_1 \) includes all pairs \( \{a, b\} \) with \( a, b \in \mathcal{X} \). Also consider two associated choice functions \( C_1 \) defined on \( \mathcal{G}_1 \), and \( C_2 \) defined on \( \mathcal{G}_2 \), with \( C_1(G) = C_2(G) \) for all \( G \in \mathcal{G}_1 \). The welfare relations \( R'_1 \) and \( P'_1 \) obtained from \( (\mathcal{G}_1, C_1) \) are weakly coarser than the welfare relations \( R'_2 \) and \( P'_2 \) obtained from \( (\mathcal{G}_2, C_2) \).

It follows that the addition of data (i.e., the expansion of \( \mathcal{G} \)) makes \( R' \) and \( P^* \) weakly coarser, while the elimination of data (i.e., the reduction of \( \mathcal{G} \)) makes \( R' \) and \( P^* \) weakly finer. Intuitively, if choices between two alternatives, \( x \) and \( y \), are unambiguous over some domain, they are also unambiguous over a smaller domain. Notice, however, the same principle does not hold for \( P' \) or \( R^* \).

The following simple observation indicates how the addition or deletion of data affects the set of individual welfare optima.

**Observation 7.5.**

Fix \( \mathcal{X} \). Consider two generalized choice domains \( \mathcal{G}_1 \) and \( \mathcal{G}_2 \) with \( \mathcal{G}_1 \subset \mathcal{G}_2 \), where \( \mathcal{X}_1 \) includes all pairs \( \{a, b\} \) with \( a, b \in \mathcal{X} \). Also consider two associated
choice functions $C_1$ defined on $G_1$, and $C_2$ defined on $G_2$, with $C_1(G) = C_2(G)$ for all $G \in G_1$. Consider any $X \in X_1$.

(a) If $x \in X$ is a weak welfare optimum for $X$ based on $(G_1, C_1)$, it is also a weak welfare optimum for $X$ based on $(G_2, C_2)$.

(b) Suppose that $x \in X$ is a strict welfare optimum for $X$ based on $(G_1, C_1)$, and that there is no $y \in X$ such that $x I_1 y$. Then $x$ is also a strict welfare optimum for $X$ based on $(G_2, C_2)$.

It follows that the addition of data cannot shrink the set of weak individual welfare optima and can only shrink the set of strict individual welfare optima in very special cases.

Observations 7.4 and 7.5 motivate an agenda involving refinements of the welfare relations considered in this chapter. Our goal is to make our proposed welfare relations more discerning while maintaining libertarian deference to individual choice by officiating between apparent choice conflicts. In other words, if there are some GCSs in which $x$ is chosen over $y$, and some other GCSs in which $y$ is chosen over $x$, we can look for objective criteria that might allow us to disregard some of these GCSs, and thereby refine the initial welfare relations. We can then construct new welfare relations based on the pruned data, which will be weakly finer than the initial ones, and which may contain fewer welfare optima.

Notably, observations 7.4 and 7.5 rule out the possibility of self-officiating—that is, discriminating between apparently conflicting behaviors through “meta-choices.” As an illustration, assume there are two GCSs, $G_1, G_2 \in \mathcal{G}$ with $G_1 = (X, d_1)$ and $G_2 = (X, d_2)$, such that the individual chooses $x$ from $G_1$ and $y$ from $G_2$. Our object is to determine which behavior the planner should mimic when choosing from $X$. Instead of letting the planner resolve this based on external criteria, why not let the individual herself resolve it? Suppose we know that the individual, if given a choice between the two choice situations $G_1$ and $G_2$, would choose $G_1$. Doesn’t this mean that $G_1$ provides a better guide for the planner (in which case the planner should select $x$)? Not necessarily. The choice between $G_1$ and $G_2$ is simply another GSC, call it $G_3 = (X, d_3)$, where $d_3$ indicates that component choices are made in a particular sequence, and under particular conditions. If the individual selects $x$ in $G_3$, all we have learned is that there is one more ancillary condition, $d_3$, in which he would choose $x$. Since choices between generalized choice situations simply create new generalized choice situations, and since the addition of data on decisions in new generalized choice situations does not usefully refine the primary welfare relation, $P^*$, or the sets of welfare optima, it does not help us resolve the normative ambiguity associated with choice conflicts.
Refinements Based on Imperfect Information Processing

When we say that an individual’s standard choice situation is $X$, we mean that, based on all of the objective information that is available to him, he is actually choosing among elements of $X$. In standard economics, we use this objective information to reconstruct $X$, and then infer that he prefers his chosen element to all the unchosen elements of $X$. But what if he fails to use all of the information available to him, or uses it incorrectly? What if the objective information available to him implies that he is actually choosing from the set $X$, while in fact he believes he is choosing from some other set, $Y$? In that case, should a planner nevertheless mimic his choice when evaluating objects from $X$? Not in our view.

Why would the individual believe himself to be choosing from some set, $Y$, when in fact, according to the available objective information, he is choosing from the set $X$? There are many possible reasons. His attention may focus on some small subset of $X$. His memory may fail to call up facts that relate choices to consequences. He may forecast the consequences of his choices incorrectly. He may have learned from his past experiences more slowly than the objective information would permit.

In principle, if we understand the individual’s cognitive processes sufficiently well, we may be able identify his perceived choice set $Y$ and reinterpret the choice as pertaining to the set $Y$ rather than to the set $X$. We refer to this process as “deconstructing choices.” While it may be possible to accomplish this in some instances (see, e.g., Köszegi and Rabin, chapter 8), we suspect that, in most cases, this is beyond the current capabilities of economics, neuroscience, and psychology.

We nevertheless submit that there are circumstances in which nonchoice evidence can reliably establish the existence of a significant discrepancy between the actual choice set, $X$, and the perceived choice set, $Y$. This occurs, for example, in circumstances where it is known that attention wanders, memory fails, forecasting is naive, and/or learning is inexplicably slow. In these instances, we say that the GCS is suspect.

We propose using nonchoice evidence to officiate between conflicting choice data by deleting suspect GCSs. Thus, for example, if someone chooses $x$ from $X$ under condition $d'$ where she is likely to be distracted, and chooses $y$ from $X$ under condition $d''$ where she is likely to be focused, we would delete the data associated with $(X, d')$ before constructing the welfare relations. In effect, we take the position that $(X, d'')$ is a better guide for the planner than $(X, d')$. Even with the deletion of choice data, these welfare relations may remain ambiguous in many cases due to other unresolved choice conflicts, but $R'$ and $P^*$ nevertheless become (weakly) finer, and the sets of weak individual welfare optima grow (weakly) smaller.

Note that this refinement agenda entails only a mild modification of the core libertarian principles that underlie the standard choice-theoretic approach to welfare economics. Significantly, we do not propose substituting nonchoice data, or any
external judgment, for choice data. Rather, we adhere to the principle that the social planner’s objective should be to select an alternative that the individual would select for herself in some generalized choice situation.

The Role of Neuroscience, Psychology, and Neuroeconomics

What types of nonchoice evidence might one use to determine the circumstances in which internal information processing systems work well, and the circumstances in which they work poorly? Evidence from neuroscience, psychology, and neuroeconomics concerning the functioning of various cognitive processes can potentially shed light on the operation of processes governing attention, memory, forecasting, and learning. This evidence can provide an objective basis for determining whether a particular choice situation is suspect. For example, if memory is shown to function poorly under certain environmental conditions, GSCs that are associated with those conditions, and that require factual recall, are suspect. Our work on addiction [Bernheim and Rangel, 2004] provides an illustration. Citing evidence from neuroscience, we argue that the repeated use of addictive substances causes specific information processing systems to malfunction under identifiable ancillary conditions. The choices made in these circumstances are therefore suspect, and welfare evaluations should be guided by choices made in other ancillary conditions.

For those who question the use of evidence from neuroscience, we offer the following motivating example. An individual is offered a choice between alternative $x$ and alternative $y$. When the alternatives are described verbally, the individual chooses $x$. When the alternatives are described partly verbally and partly in writing, the individual chooses $y$. Which choice is the best guide for public policy? Based on the information provided, the answer is unclear. But suppose we learn in addition that the information was provided in a dark room. In that case, we would be inclined to respect the choice of $x$, rather than the choice of $y$. We would reach the same conclusion if an ophthalmologist certified that the individual was blind. More interestingly, we submit that the same conclusion would follow if a brain scan revealed that the individual’s visual processing was neurologically impaired. In all of these cases, nonchoice evidence sheds light on the likelihood that the individual successfully processed information that was in principle available to him, and thus was able to properly characterize the choice set $X$.

The relevance of evidence from neuroscience and neuroeconomics may not be confined to problems with information processing. Pertinent considerations would also include impairments that prevent people from implementing desired courses of action. Furthermore, in many situations, simpler forms of evidence may suffice. If an individual characterizes a choice as a mistake on the grounds that he neglected or misunderstood information, this may provide a compelling basis for declaring
the choice suspect. Other considerations, such as the complexity of a GCS, could also come into play.

What Is a Mistake?
In our work on addiction, we have characterized certain types of choices as mistakes. This has led to considerable confusion, particular among those who regard the notion of a mistake as meaningless (or even self-contradictory) within the context of sensible choice-theoretic welfare economics. This is simply a matter of terminology. We use the word “mistake” to describe a choice made in a suspect GCS that is contradicted by choices in nonsuspect GCSs. In other words, if the individual chooses \( x \in X \) in one GCS where she properly understands that the choice set is \( X \) at the moment of choice, and chooses \( y \in X \) in another GCS where she misconstrues the choice set as \( Y \), we say that the choice of \( y \in X \) is a mistake. We recognize, of course, that the choice she believes she makes is, by definition, not a mistake given the set from which she believes she is choosing.

In Bernheim and Rangel [2007], we provide the following example of a mistake:

American visitors to the UK suffer numerous injuries and fatalities because they often look only to the left before stepping into streets, even though they know traffic approaches from the right. One cannot reasonably attribute this to the pleasure of looking left or to masochistic preferences. The pedestrian’s objectives—to cross the street safely—are clear, and the decision is plainly a mistake [32]

We know that the pedestrian in London is not attending to pertinent information and/or options, and that this leads to consequences that he would otherwise wish to avoid. Accordingly, we simply disregard this GCS on the grounds that behavior is mistaken (in the sense defined above), and instead examine choice situations for which there is nonchoice evidence that the pedestrian attends to traffic patterns.

Discussion

In this chapter, we have proposed a choice-theoretic framework for behavioral welfare economics—one that can be viewed as a natural extension of standard welfare economics. We have shown that the application of libertarian welfare principles does not require all choices to be consistent, in the classic sense. Though the guidance provided by choice data may be ambiguous in some circumstances, it may nevertheless be unambiguous in others. This partially ambiguous guidance provides sufficient information for rigorous welfare analysis. Moreover, we have shown
that one can carry out this analysis using straightforward generalizations of standard tools, such as compensating and equivalent variation, consumer surplus, and Pareto optima.

The framework for welfare analysis proposed here is a natural generalization of the standard welfare framework in two separate respects. First, it nests the standard framework as a special case. Second, when behavioral departures from the standard model are small, our welfare criterion is close to the standard criterion.

Finally, we have proposed an agenda for refining our welfare criterion. This agenda necessarily relies on the circumscribed but systematic use of nonchoice data. Significantly, we do not propose substituting nonchoice data, or any external judgment, for choice data. Rather, we adhere to the principle that the social planner’s objective should be to select an alternative that the individual would select for herself under some circumstances. Nonchoice data are potentially valuable because they may provide important information concerning which choice circumstances are most relevant for welfare and policy analysis.

The approach that we have proposed has some important additional advantages. First, it allows economists to conduct welfare analysis in environments where individuals make conflicting choices, without having to take a stand on whether individuals have “true utility functions,” or on how well-being might be measured. Second, the analyst is free to use a wide range of positive models, including those that do not entail the maximization of an underlying utility function, without sacrificing the ability to evaluate welfare.

The approach that we have proposed also has some limitations. First, in some applications, our welfare criteria may not be particularly discriminating. In such cases, the refinement agenda discussed above is particularly critical. Second, it is likely that, in some extreme cases, there will be an objective basis for classifying all or most of an individual’s potential GCSs as suspect, leaving an insufficient basis for welfare analysis. Individuals suffering from Alzheimer’s disease, other forms of dementia, or severe injuries to the brain’s decision-making circuitry might fall into this category. Decisions by children might also be regarded as inherently suspect. Thus, our framework also carves out a role for paternalism.

Where will the refinements agenda ultimately take us? The two authors of this chapter hold somewhat different views on this subject. Rangel’s view is that in the coming decades neuroeconomics will generate insights and tools that will allow us to objectively officiate between conflicting choice situations to the point where many inconsistencies can be eliminated. Bernheim sees the potential for neuroeconomics to provide a basis for officiating between conflicting choices, specifically insofar as it sheds light on information processing malfunctions or on functional impairments. However, he conjectures that the effects of many ancillary conditions will not fall unambiguously within those categories, and he anticipates that other types of evidence concerning brain processes will be susceptible to a variety of normative interpretations.
NOTES

We thank Colin Camerer, Andrew Caplin, Vincent Crawford, Robert Hall, Peter Hammond, Botond Köszegi, Preston McAfee, and seminar participants at Stanford University, the 2006 NYU Methodologies Conference, the summer 2006 Econometric Society meetings, at the winter 2007 American Economic Association meetings, and the 2007 conference on Frontiers in Environmental Economics sponsored by Resources for the Future, for useful comments. B. Douglas Bernheim gratefully acknowledges financial support from the National Science Foundation (SES-0452300). Antonio Rangel gratefully acknowledges financial support from the National Science Foundation (SES-0134618) and the Moore Foundation.

1. Evidence of incoherent choice patterns, coupled with the absence of a scientific foundation for assessing true utility, has led some to conclude that behavioral economics should embrace fundamentally different normative principles than standard economics (see, e.g., Sugden [2004]).

2. In the latter case, the states may not be observable to the planner. Instead, they may reflect the individual’s private information. With respect to privately observed states, it makes little difference whether the state reflects an event that is external to the individual, or internal (e.g., a randomly occurring mood). Thus, the standard framework subsumes cases where states are internal; see, e.g., Gul and Pesendorfer [2006].

3. According to the definition proposed by Arrow [1959], \( x \) is revealed preferred to \( y \) if there is some \( X \in \mathcal{X} \) for which \( x \in C(X) \) and \( y \notin C(X) \). Under the weak congruence axiom, that definition is equivalent to the statement that \( xP_1 y \), where \( P \) is defined as in the text [Sen, 1971].

4. One can illustrate this point with reference to “temptation preference” as formulated by Gul and Pesendorfer [2001]. Though the Gul-Pesendorfer analysis may appear to resolve normative ambiguities arising from time-inconsistent choices by virtue of establishing the uniqueness of a utility representation, the apparent resolution is illusory. For a related point, see Köszegi and Rabin (chapter 8), who argue that, as a general matter, utility is fundamentally unidentified in the absence of assumptions unsupported by choice data.

5. The following is a description Rubinstein and Salant’s (chapter 5) binary relation, using our notation. Assume that \( C \) is always single valued. Then \( x \succ y \) iff \( C([x, y], d) = x \) for all \( d \) such that \(([x, y], d) \in \mathcal{G}\). The relation \( \succ \) is defined for choice functions satisfying a condition related to weak congruence and thus—in contrast to \( P' \) or \( P^* \)—depends only on binary comparisons. Rubinstein and Salant [2006] proposed a special case of the relation \( \succ \), without reference to welfare.

6. When there are no ancillary conditions and the revealed preference relation is a libertarian relation for \( C \), then \( C \) is called a normal choice correspondence [Sen 1971].

7. This is a natural assumption if we take the time periods to be short. It is possible, e.g., to compensate an individual fully for one day of fasting, but presumably not for one year of fasting, since that would be fatal.

8. This formulation of compensating variation assumes that \( \mathcal{G} \) is rectangular. If \( \mathcal{G} \) is not rectangular, then as a general matter we would need to write the final GCS as \((X(\alpha_1, m), d_1(m))\) and specify the manner in which \( d_1 \) varies with \( m \).
9. For example, if $m_1^A$ is the CV-A for a change from $(X(a_0, 0), d_0)$ to $(X(a_1, m), d_1)$, and if $m_2^A$ is the CV-A for a change from $(X(a_1, m_1^A), d_1)$ to $(X(a_2, m_1^A + m), d_2)$, then nothing that the individual would choose from $(X(a_0, 0), d_0)$ is unambiguously chosen over anything that he would choose from $(X(a_2, m_1^A + m_1^A), d_2)$.

10. In the standard framework, if $m_1$ is the CV for a change from $(X(a_0, 0), X(a_1, m_1))$, and if $m_2$ is the CV for a change from $(X(a_1, m_1))$ to $(X(a_2, m_1 + m)$), then $m_1 + m_2$ is the CV for a change from $(X(a_0, 0))$ to $(X(a_2, m)$. The same statement does not necessarily hold within our framework.

11. In Bernheim and Rangel [2007], we consider some other possible generalizations of Pareto efficiency.

12. The proof of theorem 7.5 is more subtle than one might expect; in particular, there is no guarantee that any individual’s welfare optimum within the set $\{ y \in X \mid \forall i, xP^*_iy \}$ is a Pareto optimum within $X$.

13. Suppose, e.g., that $xP^*_iy$ given $(G_1, C_1)$, so that $xP^*_iy$. Then, with the addition of a GCS for which $x$ is chosen but $y$ is not with both available, we would have $xP^*_iy$; in other words, the relation $P^*$ would become finer. Similarly, suppose that $xP^*_iy$ given $(G_1, C_1)$, so that $yP^*_2x$. Then, with the addition of GCS for which $y$ is chosen when $x$ is available, we would have $yP^*_2x$; in other words, the relation $R^*$ would become finer.

REFERENCES


References


