Beyond Revealed Preference: Choice Theoretic Foundations for Behavioral Welfare Economics[*]

B. Douglas Bernheim
Princeton University
and
NBER

Antonio Rangel
California Institute of Technology
and
NBER

October 2007

Abstract

This paper proposes a universal choice-theoretic framework for evaluating economic welfare with the following features. (1) In principle, it encompasses all behavioral models; it is applicable irrespective of the processes generating behavior, or of the positive model used to describe behavior. (2) It subsumes standard welfare economics both as a special case (when standard choice axioms are satisfied) and as a limiting case (when behavioral anomalies are small). (3) Like standard welfare economics, it requires only data on choices. (4) It is easily applied in the context of specific behavioral theories, such as the β, δ model of time inconsistency, for which it has novel normative implications. (5) It generates natural counterparts for the standard tools of applied welfare analysis, including compensating and equivalent variation, consumer surplus, Pareto optimality, and the contract curve, and permits a broad generalization of the of the first welfare theorem. (6) Though not universally discerning, it lends itself to principled refinements.

[*] We would like to thank Colin Camerer, Andrew Caplin, Vincent Crawford, Robert Hall, Peter Hammond, Botond Koszegi, Preston McAfee, Paul Milgrom, and seminar participants at Stanford University, U. C. Berkeley, Princeton University, the 2006 NYU Methodologies Conference, the summer 2006 Econometric Society meetings, the summer 2006 ASHE meetings, the winter 2007 ASSA meetings, the 2007 Conference on Frontiers in Environmental Economics sponsored by Resources for the Future, the spring 2007 SWET meetings, and the summer 2007 PET meetings, for useful comments. B. Douglas Bernheim gratefully acknowledges financial support from the NSF (SES-0452300). Antonio Rangel gratefully acknowledges financial support from the NSF (SES-0134618) and the Moore Foundation.

Interest in behavioral economics has grown in recent years, stimulated largely by accumulating evidence that the standard model of consumer decision-making may provide an inadequate positive description of human behavior. Behavioral models are increasingly finding their way into policy evaluation, which inevitably involves welfare analysis. Yet behavioral economics fundamentally challenges our ability to formulate appropriate normative criteria. If an individual's choices do not reflect a single coherent preference relation, how can an economist hope to justify a coherent non-paternalistic welfare standard?

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 then treat the utility index as welfare-relevant. Unfortunately, such an approach is often problemmatic 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 importantly, in many cases the nature and significance of the condition under which the choice is made changes when the choice is transferred to a social planner. Consider the example of time inconsistency. Suppose an individual chooses alternative x over alternative y at time t, and y over x at time $t-1$. One could account for the pursuit of different objectives at time t and $t-1$ by inserting the time of the individual's decision into the utility function (quasihyperbolic discounting is an example, e.g., Laibson [1997]). But that rationalization does not tell us how the decision's timing should enter the utility function once it has been delegated to a social planner. Thus, one could argue that the planner should choose x at time t and y at time $t-1$, the same alternatives that the individual would select. But one could also argue that, in either case, the planner should choose y at time $t-1$, on the grounds that the planner's decision, like the individual's decision at time $t-1$, is always at "arm-length" from the experience. Much of the literature on self-control takes this second view. However, neither answer is plainly superior.

The obvious problems with the normative methodology described in the last two paragraphs have led many behavioral economists to distinguish between "decision utility," which provides an as-if representation of choices, and "true" or "experienced" utility, 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] Many economists are deeply troubled by this wholesale abandonment of the revealed preference paradigm.

In seeking appropriate principles for behavioral welfare analysis, it is important to recall that standard 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 himself. 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 he would choose for himself rather than something that someone else would choose for him.

The central premise of this paper is that the same normative principle remains applicable even when individuals make "anomalous" choices of the various types commonly identified in behavioral research. We submit that confusion about normative criteria arises in the context of behavioral models only when we ignore this guiding principle, and proceed as if welfare analysis must respect a *rationalization* of choice (that is, utility or preferences) rather than choice itself. As we argue, welfare analysis requires no rationalization of behavior. When choice data lacks a consistent rationalization, the normative guidance it provides may be ambiguous in some circumstances, but is typically unambiguous in others. As we show, this partially ambiguous guidance always provides a foundation for rigorous welfare analysis.

This paper represents our effort to develop a universal choice-theoretic framework for

[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]).

evaluating economic welfare. As we explain, our framework has the following attractive features. (1) In principle, it encompasses all behavioral models; it is applicable irrespective of the processes generating behavior, or of the positive model used to describe behavior. (2) It subsumes standard welfare economics both as a special case (when standard choice axioms are satisfied) and as a limiting case (when behavioral anomalies are small). (3) Like standard welfare economics, it requires only data on choices. (4) It is easily applied in the context of specific behavioral theories. It leads to novel normative implications for the familiar β, δ model of time inconsistency. For a model of coherent arbitariness, it provides a choice-theoretic (non-pscyhological) justifications for multi-self Pareto optimality. (5) It generates natural counterparts for the standard tools of applied welfare analysis, including compensating and equivalent variation, consumer surplus, Pareto optimality, and the contract curve, and permits a broad generalization of the of the first welfare theorem. (6) Though not universally discerning, it lends itself to principled refinements.

The paper is organized as follows. Section 1 reviews the foundations of standard welfare economics. Section 2 presents a general framework for describing choices and behavioral anomalies. Section 3 sets forth choice-theoretic principles for evaluating individual welfare in the presence of choice anomalies. It also explores the implications of those principles in the context of quasihyperbolic discounting and coherent arbitrariness. Section 4 describes the generalizations of compensating variation and consumer surplus in this setting. Section 5 generalizes the notion of Pareto optimality and examines competitive market efficiency as an application. Section 6 sets forth an agenda for refining our welfare criterion and identifies a potential (narrowly limited) role for non-choice evidence. Section 7 offers some concluding remarks. Simple proofs appears in the text, but we defer longer and more technical proofs to the Appendix.

1 Standard welfare economics: a brief review

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 anomolous choices identified in the behavioral literature. As we discuss later, 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 objective here is to review the standard perspective on individual welfare.

We will use $\mathbb{X}$ 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.[2] A *standard choice situation* (SCS) consists of a constraint set $X \subseteq \mathbb{X}$. 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 situtations. We will use $\mathcal{X}$ to denote the domain of standard choice situations with which welfare economics is concerned. Usually, the standard framework restricts $\mathcal{X}$ to include only compact sets (but we will not impose that restriction in subsequent sections). Throughout this paper we will make the following weak assumption:

Assumption 1: $\mathcal{X}$ includes all non-empty finite subsets of $\mathbb{X}$ (and possibly other subsets).

[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 Pesendorder [2006].

An individual's choices are described by a correspondence $C : \mathcal{X} \rightrightarrows \mathbb{X}$, 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 his 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 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 strong consistency property. One version of this property, *weak congruence*, 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 (weak preference), P (strict preference), and I (indifference) defined over the choice objects in $\mathbb{X}$, which are derived from the choice correspondence in the following way:

$$xRy \text{ iff } x \in C(\{x, y\}) \tag{1}$$

$$xPy \text{ iff } xRy \text{ and } \sim yRx \tag{2}$$

$$xIy \text{ iff } xRy \text{ and } yRx \tag{3}$$

Under the weak congruence axiom, the relation R is an ordering, commonly interpreted as *revealed preference*.[3] 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 necessarily embrace that suggestion; instead, R is simply a summary of what the individual chooses in a wide range of situations.

[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 xPy, where P is defined as in the text (Sen [1971]).

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. Though we often define them using language that invokes notions of well-being, we can dispense with such language entirely. 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. Likewise, in settings with many individuals, an alternative x is Pareto efficient if there is no other alternative that everyone would voluntarily choose over x.

2 A general framework for describing choices

In behavioral economics as in standard economics, we are concerned with choices among sets of objects drawn from some broader set $\mathbb{X}$. 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 as relevant to a social planner's choice once the decision has been delegated. 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 could be 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.

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 [2007] have independently formulated a similar framework for describing the impact of choice procedures on decisions; they refer to ancillary conditions as "frames."

An individual's choices are described by a correspondence $C : \mathcal{G} \rightrightarrows \mathbb{X}$, 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, faced with any set of alterantives, the individual is always willing to make some choice:

Assumption 2: $C(G)$ is non-empty for all $G \in \mathcal{G}$.

2.1 What are ancillary conditions?

As a general matter, it is difficult to draw a bright line between the characteristics of the objects in $\mathbb{X}$ 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 changes 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. As we explained in the introduction, the time at which a choice is made does not necessarily hold the same significance for the individual's welfare when the 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 chosing 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 chosing 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 that affect choice. 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 characterstic 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.

When judgment is involved, different analysts may wish to draw different lines between the characteristics of choice objects and ancillary conditions. It is therefore important to emphasize that the tools we develop here provide a coherent method for conducting choice-based welfare analysis no matter how one draws that line. For example, it allows economists to perform welfare analysis without abandoning the standard notion of a consumption good.

2.2 Scope of the framework

Our framework can incorporate non-standard 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 often 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. Indeed, throughout most of this paper, we allow for *all* non-empty choice correspondences (Assumption 2). Hence the framework accomodates choice reversals based on "irrelevant alternatives," as well as 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 an individual must choose a dollar prize from the interval $[0, 100)$. That set does not lie in the domain of a standard choice correspondence. And yet, one can easily imagine someone making a choice from it; they 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 $x \in \mathbb{X}$ 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].

2.3 Positive versus normative analysis

Before proceeding, it is important to draw a clear distinction between positive and normative analysis. That distinction will allow us to clarify our tasks, which are confined to normative analysis.

In standard economics, we generally assume that such data are available for elements of some restricted set of SCSs, $\mathcal{X}^D \subset \mathcal{X}$. The objective of standard 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 $\mathbb{X}$, 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 $\mathcal{X} \backslash \mathcal{X}^D$ (by maximizing that function for each $X \in \mathcal{X} \backslash \mathcal{X}^D$).

Likewise, in behavioral economics, we assume that choice data are available for some subset of the environments of interest, $\mathcal{G}^D \subset \mathcal{G}$. The objective of positive behavioral analysis is to extend the choice correspondence C from observations on $\mathcal{G}^D$ to the entire set $\mathcal{G}$. As in standard economics, this may be accomplished by defining preferences over some appropriately defined set of objects, estimating preference parameters using choice data drawn from sets in $\mathcal{G}^D$, and then using those estimated preferences to infer choices for

GCSs in $\mathcal{G} \backslash \mathcal{G}^D$. However, a behavioral economist might also use other positive tools, such as models of choice algorithms, neural processes, or rules of thumb.

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 Section 2. Likewise, in conducting choice-based behavioral welfare analysis, we take as given the individual's choice correspondence, C, defined on $\mathcal{G}$ rather than $\mathcal{G}^D$. 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.[4]

Thus, preferences and utility functions, which are constructs used to extend C from $\mathcal{X}^D$ to $\mathcal{X}$ in the standard framework, and which may (or may not) be used to extend C from $\mathcal{G}^D$ to $\mathcal{G}$ in our framework, are positive tools, not normative tools. They simply reiterate the information contained in the extended choice function C (both the observed choices and the inferred choices). Beyond that reiteration, they add no new information that might pertain to welfare analysis. In a behavioral setting, these constructs cannot meaningfully reconcile choice inconsistencies; they can only reiterate those inconsistencies. Thus, one cannot resolve normative puzzles by identifying classes of preferences that rationalize apparently inconsistent choices.[5]

[4] Thus, our concerns are largely orthogonal to issues examined in the literature that attempts to identify representations of non-standard choice correspondences, either by imposing conditions on choice correspondences and deriving properties of the associated representations, or by adopting particular representations (e.g., preference relations that satisfy weak assumptions) and deriving properties of the associated choice correspondences. Recent contributions in this area include Kalai, Rubinstein, and Spiegler [2002], Bossert, Sprumont, and Suzumura [2005], and Ehlers and Sprumont [2006].

[5] For a related point, see Koszegi and Rabin [2007], who argue that, as a general matter, utility is fundamentally unidentified in the absence of assumptions unsupported by choice data.

3 Individual welfare

In this section, we propose a general approach for extending standard choice-theoretic welfare analysis to situations in which individuals make anomolous 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 indivdual's welfare.

3.1 Individual welfare relations

Sometimes, welfare analysis involves the identification of an individual's "best" alternative (for example, 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 (discussed in Section 4) and to welfare analysis in settings with many people (discussed in Section 5). It is also equivalent to the construction of a binary relation, call it Q, where xQy 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 (1), that definition implictly 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:

$$xRy \text{ iff, for all } X \in \mathcal{X} \text{ with } x, y \in X,\ y \in C(X) \text{ implies } x \in C(X) \tag{4}$$

Similarly, we can define P, the asymmetric component of R, as follows:[6]

$$xPy \text{ iff, for all } X \in \mathcal{X} \text{ with } x, y \in X, \text{ we have } y \notin C(X) \tag{5}$$

[6] Note, however, that this does not correspond to the definition proposed by Arrow [1959], which requires only that there is some $X \in \mathcal{X}$ for which $x \in C(X)$ and $y \notin C(X)$.

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 (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 $x, y \in \mathbb{X}$, we have that $xR'y$ if, whenever x and y are available, y is never chosen unless x is as well. When $xR'y$, we will say that x is *weakly unambiguously chosen over* y.

As usual, we can define the symmetric and asymmetric components of R'. We say that $xP'y$ if $xR'y$ and $\sim yR'x$. 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 immediate (and ultimately more useful) generalization of (5):

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

In other words, for any $x, y \in \mathbb{X}$, we have xP^*y iff, whenever x and y are available, y is never chosen. When xP^*y, we will say that x is *strictly unambiguously chosen over* y. For the sake of brevity, we will sometimes drop the modifier "strictly."

Corresponding to P^*, there is an alternative generalization 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 $x, y \in \mathbb{X}$, 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 $\sim yR^*x$ implies 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 [2007] have separately proposed a binary relation that is related to P' and P^*.[7]

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^*.

As we've seen, $R^{'}$ is more faithful to the standard notion of weak revealed preference, while P^* is 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 of the weak relations, R^* and R'.[8] The relation R^* is obviously complete: for any $x, y \in \mathbb{X}$, 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 1.

Example 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.

Wihout further structure, there is no guarantee that any of the relations defined here will be transitive. Example 2 makes this point with respect to P^*; it is also easy to construct counterexamples for the other relations. □

Example 2: Suppose that $\mathcal{G} = \{X_1, X_2, X_3, X_4\}$ with $X_1 = \{a, b\}$, $X_2 = \{b, c\}$, $X_3 =$

[7] The following is a description Rubinstein and Salant's [2007] 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 satsifying a condition related to weak congruence, and thus – in contrast to P' or P^* – depends only on binary comparisons. Rubinstein and Salant [2006] considered a special case of the relation $\succ$ for decision problems involving choices from lists, without reference to welfare. Mandler [2006] proposed a welfare relation that is essentially equivalent to Salant and Rubinstein's $\succ$ for the limited context of status quo bias.

[8] As in the standard framework, one would never expect either the symmetric or asymmetric components of these relations to be complete.

$\{a, c\}$, 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 that case, we have aP^*b and bP^*c; in contrast, we can only say that aI^*c. □

Fortunately, to conduct useful welfare analysis, 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 comparisons involves P^*, the most natural generalization of strict revealed preference.

Theorem 1: *Consider any x_1,...,x_N such that $x_i R' x_{i+1}$ for $i = 1, ..., N-1$, with $x_k P^* x_{k+1}$ for some k. Then $\sim x_N R' x_1$.*

Proof: Suppose on the contrary that $x_N R' x_1$. Without loss of generality, we can renumber the alternatives so that $k = 1$. Let $X^0 = \{x_1, ..., x_N\}$. Since $x_1 P^* x_2$ and $x_1 \in X^0$, we know that $x_2 \notin C(X^0, d)$ for all d such that $(X^0, d) \in G$. Now suppose that, for some $i \in \{2, ..., N\}$, we have $x_i \notin C(X^0, d)$ for all d such that $(X^0, d) \in G$. We argue that $x_{i+1(\text{mod}\, N)} \notin C(X^0, d)$ for all d such that $(X^0, d) \in G$. This follows from the following facts: $x_i R' x_{i+1}$, $x_i \in X^0$, and $x_i \notin C(X^0, d)$ for all d such that $(X^0, d) \in G$. By induction, this means $C(X^0, d)$ is empty, contradicting Assumption 2. Q.E.D.

Theorem 1 assures us that a planner who evaluates alternatives based on R' (to express "no worse than") and P^* (to express "better than") cannot be turned into a "money pump." In the context of standard consumer theory, Suzumura's [1976] analogous *consistency* property plays a similar role.[9] The theorem has an immediate and important corollary:

Corollary 1: *P^* is acyclic. That is, for any x_1,...,x_N such that $x_i P^* x_{i+1}$ for $i = 1, ..., N-1$, we have $\sim x_N P^* x_1$.*

[9] A preference relation R is *consistent* in Suzumura's sense if $x_1 R x_2 ... R x_N$ with $x_i P x_{i+1}$ for some i implies $\sim x_N R x_1$.

Acyclicity is weaker than transitivity, but in most contexts it suffices to guarantee the existence of maximal elements, and it allows us to identify and measure unambiguous improvements. The power of Corollary 1 is that it delivers an acyclic welfare criterion without imposing *any* assumption on the choice correspondence, other than non-emptiness. Regardless of how poorly behaved the choice correspondence C may be, P^* is nevertheless acyclic.

As our next example demonstrates, the relation P' does not share this desirable property. On the contrary, P' may be cyclic.

Example 3: Suppose that $\mathcal{G} = \{X_1, X_2, X_3, X_4\}$ 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 $aP'bP'cP'a$. □

3.2 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; in other words, if 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 $yP'x$; in other words, if 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 (which follow immediately from definitions) address this question.

Observation 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. Our next observation characterizes the set of individual welfare optima more precisely.

Observation 2: 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.

For an illustration of Observation 2, let's revisit Example 2. 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^*).

Notice that Observation 1 guarantees the existence of weak welfare optima (but not of strict welfare optima). The fact that we have established existence without making any additional assumptions, e.g., related to continuity and compactness, may at first seem surprising, 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 $\mathcal{G}$; this is treated as data.

Standard existence issues arise when the choice function is built up from other components. The following example clarifies these issues.

Example 4: Consider the same choice data as in Example 2, but suppose we limit attention to $\mathcal{G}' = \{X_1, X_2, X_3\}$. In this case we have that $aP^*bP^*cP^*a$. Here, the intransitivity is apparent; P^* is cyclic because Assumption 1 is violated ($\mathcal{G}'$ does not contain all finite sets).[10] Naturally, if we are interested in creating a preference or utility representation based on the data contained in $\mathcal{G}'$ in order to project what the individual would choose from the set 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 select from 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.

□

According to Observation 2, some alternative x may be an individual welfare optimum for the set X even though there is no ancillary condition d under which $x \in C(X, d)$. (The fact that a is an individual welfare optimum in X_3 in Example 2 illustrates this possibility.) However, that property is still consistent with the spirit of the libertarian principle: the individual welfare optimum x is chosen despite the availability of each $y \in X$ in *some* circumstances, though not necessarily ones involving choices from X. In contrast, an alternative x that is *never* chosen when some alternative $y \in X$ is available cannot be an individual welfare optimum in X.

[10] 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. This is no accident: Observation 1 does not depend on Assumption 1.

The following example, based on an experiment reported by Iyengar and Lepper [2000], illustrates why it may be unreasonable to exclude the type of individual welfare optima described in the preceding paragraph. (For another more formal argument, see Section 3.3, below.) 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 thirty flavors (including strawberry) are available. Since the individual might not want the planner to act overwhelmed when choosing on his behalf, it is important to allow for the possibility that the planner should pick strawberry jam on his behalf even when thirty alternatives are available. Similar concerns would arise whenever the act of thinking about X causes the individual to experience feelings (e.g., temptation) that affect his 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 remain neutral by adopting a notion of individual welfare optima that can accommodate such possibilities.

3.3 Why this approach?

It is natural to wonder whether there is some other, potentially more attractive 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 than P^* to the libertarian principle.

Consider a choice correspondence C defined on $\mathcal{X}$ and an asymmetric binary relation Q defined on $\mathbb{X}$. For any $X \in \mathcal{X}$, let $m_Q(X)$ be the maximal elements in X for the relation Q, i.e.,

$$m_Q(X) = \{x \in X \mid \nexists y \in X \text{ with } yQx\}$$

We will also define, for $X \in \mathcal{X}$, the set

$$D(X) = \{d \mid (X, d) \in \mathcal{G}\}$$

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:

Definition: Q is an *inclusive libertarian relation* for C if, for all $X \in \mathcal{X}$, we have $\cup_{d \in D(X)} C(X) \subseteq m_Q(X)$.

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:

Definition: Q is an *exclusive libertarian relation* for C if, for all $X \in \mathcal{X}$, we have $m_Q(X)$ non-empty, and $m_Q(X) \subseteq \cup_{d \in D(X)} C(X)$.

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.[11]

Definition: Q is a *libertarian relation* for C if, for all $X \in \mathcal{X}$, Q is both inclusive and exclusive.

[11] 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]).

We have already demonstrated that P^* is always an inclusive libertarian relation (Observation 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 example 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.[12]

Example 5: Consider a choice correspondence C with the following properties:
(i) $x \notin C(\{x, y, z\}, d)$ for all ancillary conditions $d \in D(\{x, y, z\})$,
(ii) $C(\{x, y\}, d) = \{x\}$ for all ancillary conditions $d \in D(\{x, y\})$, and
(iii) $C(\{x, z\}, d) = \{x\}$ for all ancillary conditions $d \in D(\{x, z\})$.

(Note that this example resembles the strawberry jam experiment described above. Here, the individual chooses x in all pairwise comparisons, but does not choose x when overwhelmed with alternatives.)

We claim that there is no exclusive libertarian relation (and hence no libertarian relation) for C. Assume, contrary to the claim, that Q is an exclusive libertarian relation for C. Then, from (i), we know that $x \notin m_Q(\{x, y, z\})$, from which it follows that either yQx or zQx. From (ii), we know that $m_Q(\{x, y\}) = \{x\}$, from which it follows that xQy. From (iii), we know that $m_Q(\{x, z\}) = \{x\}$, from which it follows that xQz. But these conclusions contradict the requirement that Q is asymmetric. □

The preceding observation implies that there exists no general procedure for finding either libertarian welfare relations or exclusive libertarian welfare relations for all non-empty choice correspondences. Consideration of inclusive libertarian relations is therefore unavoidable. There are, of course, inclusive libertarian relations other than P^*. For example, the null relation, R^{Null} ($\sim xR^{Null}y$ for all $x, y \in \mathbb{X}$), falls into this category; for any set X, the maximal

[12] One naturally wonders about the properties that a generalized choice correspondence must have to guarantee the existence of a liberatarian relation. See Rubinstein and Salant [2007] for an analysis of that issue.

elements under R^{Null} consist of X, which of course includes all of the chosen elements. Yet R^{Null} is far less discerning, and further from the libertarian principle, than P^*. In fact, we have the following result:

Theorem 2: *Consider any choice correspondence C, and any inclusive libertarian relation $Q \neq P^*$. Then P^* is finer than Q. Furthermore, for all $X \in \mathcal{X}$, the set of maximal elements in X for the relation P^* is contained in the set of maximal elements in X for the relation Q (that is, $m_{P^*}(X) \subseteq m_Q(X)$).*

Proof: Suppose on the contrary that P^* is not finer than Q. Then for some x and y, we have xQy but $\sim xP^*y$. Because $\sim xP^*y$, we know that there exists some X containing x and y, as well as some ancillary condition d, for which $y \in C(X, d)$. Since Q is an inclusive libertarian relation, we must then have $y \in m_Q(X)$. But since $x \in X$, that can only be the case if $\sim xQy$, a contradiction. The statement that $m_{P^*}(X) \subseteq m_Q(X)$ for all $X \in \mathcal{X}$ follows trivially. Q.E.D.

Thus, for all choice correspondences C and all choice sets X, P^* is *always* the most discriminating inclusive libertarian relation. Theorem 2 also implies that no libertarian relation exists unless P^* is libertarian.

3.4 Relation to multi-self Pareto optima

Under certain restrictive conditions, our notion of an individual welfare optimum coincides with the idea of a multi-self Pareto optimum, which is sometimes used as a behavioral welfare criterion (see, e.g., Laibson et. al. [1998], or 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 (that is, $\mathcal{G} = \mathcal{X} \times D$, where $d \in D$); in that case, we say that $\mathcal{G}$ is *rectangular*. Imagine also that, for each $d \in D$, choices obey standard axioms; they correspond to the maximal elements of a preference ranking R_d, and hence to the alternatives that maximize a utility function u_d.[13]

If one imagines that each ancillary condition activates a different "self," then one can apply the Pareto criterion across selves. We will say that y *weakly multi-self Pareto dominates* x, abbreviated yMx, iff $u_d(y) \geq u_d(x)$ for all $d \in D$, with strict inequality for some d. We will say that y *strictly multi-self Pareto dominates* x, abbreviated yM^*x, iff $u_d(y) > u_d(x)$ for all $d \in D$. Moreover, $x \in X \subset \mathbb{X}$ is a *weak multi-self Pareto optimum* in X if there is no $y \in X$ such that yM^*x, and x is a *strict multi-self Pareto optimum* in X if there is no $y \in X$ such that yMx.

Theorem 3: *Suppose that $\mathcal{G}$ is rectangular, and that choices for each $d \in D$ maximize a utility function u_d. Then $M^* = P^*$ and $M = P'$. Therefore, $x \in X$ is a weak (strict) multi-self Pareto optimum in X iff it is a weak (strict) individual welfare optimum.*

Proof: First we verify that $M^* = P^*$. Assume yM^*x. By definition, $u_d(y) > u_d(x)$ for all $d \in D$. It follows that for any $G = (X, d)$ with $x, y \in X$, the individual will not select x. Therefore, yP^*x. Now assume yP^*x. By definition, the individual will not be willing to select x given any generalized choice situation of the form $G = (\{x, y\}, d)$. That implies $u_d(y) > u_d(x)$ for all $d \in D$. Therefore, yM^*x.

Next we verify that $M = P'$. Assume yMx. By definition, $u_d(y) \geq u_d(x)$ for all $d \in D$, with strict inequality for some d'. It follows that for any $G = (X, d)$ with $x, y \in X$, the individual will never be willing to choose x but not y. Moreover, for d' he is only willing to choose y from $(\{x, y\}, d)$. Therefore, $yP'x$. Now assume $yP'x$. By definition, if the individual is willing to select x given any generalized choice situation of the form $G = (\{x, y\}, d)$, then he is also willing to choose y, and there is some GCS, $G' = (X', d')$ with $\{x, y\} \subseteq X'$ for which he is willing to choose y but not x. That implies $u_d(y) \geq u_d(x)$ for all $d \in D$, and $u_{d'}(y) > u_{d'}(x)$. Therefore, yMx.

[13] To guarantee that best choices are well-defined, we would ordinarily restrict $\mathcal{X}$ to compact sets and assume that u_d is at least upper-semicontinuous, but these assumptions play no role in what follows.

The final statement concerning optima follows immediately from the equivalence of the binary relations. Q.E.D

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). However, our framework does *not* justify the multi-self Pareto criterion for quasi-hyperbolic consumers because $\mathcal{G}$ is not rectangular; see Section 3.5.2, below. It *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]; see Section 3.5.1. Ironically, the multi-self Pareto criterion has not to our knowledge been proposed as a welfare standard in such settings.

For the narrow settings that are consistent with the assumptions stated in Theorem 3, one can view our approach as a justification for the multi-self Pareto criterion that does not rely on untested and questionable psychological assumptions. Critically, the justification is choice-theoretic, not psychological. Our approach is also more general in that it does not require the GCS to be rectangular, or the choice correspondence to be well-behaved conditional on each ancillary condition.

3.5 Some applications

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

3.5.1 Coherent arbitrariness

Behavior is coherently arbitrary when some psychological anchor (for example, calling attention to one's social security number) affects behavior, but the individual nevertheless conforms to standard choice theory for any fixed anchor. This phenomenon was documented by Ariely, Loewenstein, and Prelec [2003], and led them to question the legitimacy of welfare judgments based on revealed preference.

To illustrate, let's consider a case in which an individual consumes two goods, y and z. Suppose that positive analysis delivers the following decision-utility representation:

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

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 anchor, simply shifts the weight on decision utility from z to y. Given any particular anchor, the individual behaves coherently, but his behavior is arbitrary in the sense that it depends on the signal.

Our normative framework easily accommodates this positive model. In fact, since $\mathcal{G}$ is rectangular, and since choices maximize $U(y, z \mid d)$ for each d, Theorem 3 implies that our welfare criterion is equivalent to the multi-self Pareto criterion, where each d indexes a different self.

For this positive model, it is easy to check that

$$(y', z')R'(y'', z'') \text{ iff } u(y') + dv(z') \geq u(y'') + dv(z'') \text{ for } d = d_L, d_H \tag{6}$$

Replacing the weak inequality with a strict inequality, we obtain a similar equivalence for P^*.

For a graphical illustration, see Figure 1(a). We have drawn two decision-indifference curves (that is, indifference curves derived from decision utility) through the bundle (y', z'), one for d_L (labelled I_L) and one for d_H (labelled I_H). For all bundles (y'', z'') lying below both decision-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 decision-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 decision-indifference curves, we have *neither* $(y', z')R'(y'', z'')$ nor $(y'', z'')R'(y', z')$; 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 individual's choice from this set clearly depends on the ancillary condition d. As shown in Figure 1(b), he 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 southeast 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').

Notice that, as the gap between d_L and d_H shrinks, the set $(y'', z'')P^*(y', z')$ converges to a standard upper contour set, and the set of individual welfare optima converges to a single utility maximizing choice. Thus, our welfare criterion converges to a standard criterion as the behavioral anomaly becomes small. We will return to this theme in Section 3.6.

3.5.2 Dynamic inconsistency

In this section, we examine the well-known β, δ model of hyperbolic discounting popularized by Laibson [1997] and O'Donoghue and Rabin [1999]. Economists who use this positive model for policy analysis tend to employ one of two welfare criteria: either the multi-self Pareto criterion, which associates each moment in time with a different self, or the "long-run criterion," which treats high short-term discounting as unrepresentative of true preferences. As we'll see in the section, our framework leads to an entirely different criterion. We will have more to say concerning possible justifications for the multi-self Pareto criterion and the long-run criterion in Section 6.3.

In the finite-horizon version of the β, δ model, the consumer's task is to choose a consumption vector, $C_1 = (c_1, ..., c_T)$, where c_t denotes the level of consumption at time t. For $t = 1, ..., T$, we will use C_t to denote the continuation consumption vector $(c_t, ..., c_T)$. At time t, all discretion is resolved to maximize the function

$$U_t(C_t) = u(c_t) + \beta \sum_{k=t+1}^{T} \delta^{k-t} u(c_k) \ , \tag{7}$$

where $\beta, \delta \in (0, 1)$. We assume that the individual has perfect foresight concerning future decisions, so that behavior is governed by subgame perfect equilibria. We also assume that $u(0)$ is finite; for convenience, we normalize $u(0) = 0$.[14]

To conduct normative analysis, we must recognize the fact that there is actually only one decision maker, and recast this positive model as a correspondence from GCSs into lifetime consumption vectors. Here, $\mathbb{X}$ is a set of lifetime consumption bundles. An SCS consists of some $X \subset \mathbb{X}$, for example the set implied by a standard intertemporal budget constraint. A GCS involves a choice set, X, and a decision tree, R, for selecting an element of X; thus, $G = (X, R)$. The description of the tree includes the point in time at which each choice is made. For any given X, there can be multiple trees that select from X. Because decisions may depend on the points in time at which they are made, R serves as the ancillary condition. Note that $\mathcal{G}$ is not rectangular. For example, a decision tree that involves no choice in period 1 cannot be used to select from a choice set that could produce different consumption levels in period 1. Hence, Theorem 3, which identifies conditions that justify the multi-self Pareto criterion, does not apply.

To state our main result concerning the β, δ model, we require the following definition:

$$W_t(C_t) = \sum_{k=t}^{T} (\beta\delta)^{k-t} u(c_k)$$

In other words, $W_t(C_t)$ discounts future values of the index u at the rate $\beta\delta$.

Our next result completelycharacterizes the welfare relations implied by the β,δ model.

Theorem 4: (i) $C_1' R' C_1''$ *iff* $W_1(C_1') \geq U_1(C_1'')$

(ii) $C_1' P^* C_1''$ *iff* $W_1(C_1') > U_1(C_1'')$

(iii) $P' = R'$

(iv) $C_1' R^* C_1''$ *iff* $U_1(C_1') \geq W_1(C_1'')$

[14] The role of this assumption is to rule out the possibility that a voluntary decision taken in the future can cause unbounded harm to the individual in the present. Such possibilities can arise when $u(0) = -\infty$, but seem more an artifact of the formal model than a plausible aspect of time-inconsistent behavior.

(v) *R', P', and P^* are transitive.*

Proof: See the appendix.

Parts (i) and (ii) of the theorem tell us that, to determine whether one lifetime consumption vector, C_1', is (weakly or strictly) unambiguously chosen over another, C_1'', we compare the first period decision utility obtained from C_1'' (that is, $U_1(C_1'')$) with the first period utility obtained from C_1' discounting at the rate $\beta\delta$. Given our normalization ($u(0) = 0$), we necessarily have $U_1(C_1') \geq W_1(C_1')$. Thus, $U_1(C_1') > U_1(C_1'')$ is a necessary (but not sufficient) condition for C_1' to be unambiguously chosen over C_1''.[15] That observation explains the transitivity of the preference relation (part (v)).[16] It also implies that the welfare relation never contradicts decision utility at $t = 1$, the first moment in time. For completeness, parts (iii) and (iv) of the Theorem characterize P' and R^*.[17]

Using this result, we can easily characterize the set of individual welfare optima within any choice set X.

Corollary: *For any consumption set X, C_1 is a weak welfare optimum in X iff*

$$U_1(C_1) \geq \max_{C_1' \in X} W_1(C_1')$$

Moreover, if

$$U_1(C_1) > \max_{C_1' \in X} W_1(C_1')$$

then C_1 is a strict welfare optimum in X.[18]

In other words, C_1 is a weak welfare optimum if and only if the decision utility that C_1 provides at $t = 1$ is at least as large as the highest available discounted utility, using

[15] Also, $U_1(C_1') \geq U_1(C_1'')$ is a necessary (but not sufficient) condition for C_1' to be weakly unambiguously chosen over C_1''.

[16] For similar reasons, it is also trivial to show that $C_1^1 R' C_1^2 P^* C_1^3$ implies $C_1^1 P^* C_1^3$.

[17] It follows from part (iii) that $\sim C_1' I' C_1''$ for all C_1' and C_1''. It follows from part (iv) that $C_1' I^* C_1''$ iff $U_1(C_1') \geq W_1(C_1'')$ and $U_1(C_1'') \geq W_1(C_1')$.

[18] C_1 may also be a strict welfare optimum in X even though $U_1(C_1) = \max_{C_1' \in X} W_1(C_1')$ provided that C_1 is also the unique maximizer of W_1 (which can only be the case if C_1 involves no consumption after the first period).

$\beta\delta$ as a time-consistent discount factor. Given that $W_1(c) \leq U_1(c)$ for all c, we know that $\max_{C_1' \in X} W_1(C_1') \leq \max_{c \in X} U_1(c)$, which confirms that the set of weak individual welfare optima is non-empty.

Notice that, for all C_1, we have $\lim_{\beta \to 1}[W_1(C_1) - U_1(C_1)] = 0$. 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 β, 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 $\beta = 1$.

3.6 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 plainly holds in the applications considered above; here, we establish the point with generality.

For the purpose of this analysis, we add the following mild technical assumption concerning the choice domain (the role of which is primarily to simplify the statement of our results):

Assumption 3: $\mathbb{X}$ (the set of potential choice objects) is bounded, and for all $X \in \mathcal{X}$, we have $\text{clos}(X) \in \mathcal{X}^c$ (the compact elements of $\mathcal{X}$).

Now consider a sequence of choice functions C^n, $n = 1, 2, ...$, defined on $\mathcal{G}$. Also consider a choice function $\widehat{C}$ defined on $\mathcal{X}^c$ that reflects maximization of a continuous utility function, u. We will say that C^n weakly converges to $\widehat{C}$ if and only if the following condition is satisfied: for all $\varepsilon > 0$, there exists N such that for all $n > N$ and $(X, d) \in \mathcal{G}$, each point in $C^n(X, d)$

is within ε of some point in $\widehat{C}(\text{clos}(X))$.[19] In other words, any sequence of alternatives the individual chooses from (X, d) with C^n converges to an alternative the individual chooses from the closure of X with $\widehat{C}$, and convergence is uniform across $\mathcal{G}$.

Note that we allow for the possibility that the set X is not compact. In that case, our definition of convergence implies that choices must approach the choice made from the closure of X. So, for example, if the opportunity set is $X = [0, 1)$, where the chosen action x entails a dollar payoff of x, we might have $C^n(X) = [1 - \frac{1}{n}, 1)$, whereas $\widehat{C}(X) = \{1\}$. Notice that $C^n(X)$ weakly converges to $\widehat{C}(X)$ in this example. This convergence is intuitive: the individual is satisficing, but as n increases, he demands something that leaves less and less room for improvement.

To state our next result, we require some additional definitions. For the limiting (conventional) choice correspondence $\widehat{C}$ and any $X \in \mathcal{X}^C$, we define $\widehat{U}^*(u) \equiv \{y \in X \mid u(y) \geq u\}$ and $\widehat{L}^*(u) \equiv \{y \in X \mid u(y) \leq u\}$. In words, $\widehat{U}^*(u)$ and $\widehat{L}^*(u)$ are, respectively, the standard weak upper and lower contour sets relative to a particular level of utility u for the utility representation of $\widehat{C}$. Similarly, for each choice correspondence C^n and $X \in \mathcal{X}$, we define $U^n(x) \equiv \{y \in X \mid yP^{n*}x\}$ and $L^n(x) \equiv \{y \in X \mid xP^{n*}y\}$. In words, $U^n(x)$ and $L^n(x)$ are, respectively, the strict upper and lower contour sets relative to the alternative x, defined according to the welfare relation P^{n*} derived from C^n.

We now establish that the strict upper and lower contour sets for C^n, defined according to the relations P^{n*}, converge to the conventional weak upper and lower contour sets for $\widehat{C}$.

Theorem 5: *Suppose that the sequence of choice correspondences C^n weakly converges to $\widehat{C}$, where $\widehat{C}$ is defined on $\mathcal{X}^c$, and reflects maximization of a continuous utility function, u. Consider any x^0. For all $\varepsilon > 0$, there exists N such that for all $n > N$, we have $\widehat{U}^*(u(x^0) + \varepsilon) \subseteq U^n(x^0)$ and $\widehat{L}^*(u(x^0) - \varepsilon) \subseteq L^n(x^0)$.*

Proof: See the Appendix.

[19] Technically, this involves convergence in the upper Hausdorff hemimetric; see the Appendix for details.

Because $U^n(x^0)$ and $L^n(x^0)$ cannot overlap, and because the boundaries of $\widehat{U}^*(u(x^0)+\varepsilon)$ and $\widehat{L}^*(u(x^0)-\varepsilon)$ converge to each other as ε shrinks to zero, it follows immediately (given the boundedness of $\mathbb{X}$) that $U^n(x^0)$ converges to $\widehat{U}^*(u(x^0))$ and $L^n(x^0)$ converges to $\widehat{L}^*(u(x^0))$.

Using Theorem 5, one can also establish the following corollary:

Corollary: *Suppose that the sequence of choice correspondences C^n weakly converges to $\widehat{C}$, where $\widehat{C}$ is defined on $\mathcal{X}^c$, and reflects maximization of a continuous utility function, u. For any $X \in \mathcal{X}$ and any sequence of alternatives x^n such that x^n is a weak individual welfare optimum for C^n, all limit points of the sequence maximize u in* $\text{clos}(X)$.

Because this corollary is actually a special case of Theorem 9 below, we omit a separate proof.

Theorem 5 is important for three 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 anomolies are small when using the standard framework; they view this as a justification for both standard positive analysis and for standard normative analysis. 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 unless we have established the correct criteria for such settings, 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 anomolies 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 convergence result implies that the debate over the significance of choice anomolies 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 in the context of the β, δ model.)

Third, our convergence result suggests that our welfare criterion will always be reasonably discerning provided behavioral anomalies are not too large. This is reassuring, in that the welfare relations may be extremely coarse, and the sets of individual welfare optima extremely large, when choice conflicts are sufficiently severe. The following example provides an illustration.

Example 6: Suppose that $\mathcal{X} = \{X_1, X_2, X_3, X_4\}$ (defined in Example 2), and that $\mathcal{G} = \mathcal{X} \times \{d, d'\}$. Suppose also 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 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 $i = 1, 2, 3, 4$.$\square$

In Example 6, 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. Theorem 5 assures us that, with less choice conflict, it becomes easier to identify alternatives that constitute unambiguous welfare improvements, so the set of individual welfare optima shrinks.

4 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 will focus on compensating variation; the treatment of equivalent variation is analogous. We will also illustrate how, under more restrictive assumptions, the generalized compensating variation of a price change corresponds to an analog of consumer surplus.

4.1 Compensating variation

Let's assume that the individual's SCS, $X(\alpha, m)$, depends on a vector of environmental parameters, α, and a monetary transfer, m. Let α_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 α_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.[20]

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 (that is, $y \in C(\{x, y\})$, 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 $(X(\alpha_0, 0), d_0)$, or in the final GCS, $(X(\alpha_1, m), d_1)$. In the standard framework, this causes no difficulty as the individual must be indifferent between all alternatives chosen from the same set. However, within our framework, these alternatives my 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

[20] 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.

the initial and final sets.

The second dimension of ambiguity arises from a potential form of non-monotonicity. 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 handle the resulting ambiguity by finding a level of compensation beyond which such reversals do no occur. We discuss an alternative in the Appendix.

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, below). Accordingly, we define two notions of compensating variation:

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

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

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

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

In other words, all levels of compensation greater than the CV-A guarantee that everything selected in the new set is unambiguously chosen over everything selected from the initial set. Similarly, all levels of compensation smaller than the CV-B guarantee that everything selected from the initial set is unambiguously chosen over everything selected from the new set.[21]

It is easy to verify that $m^A \geq m^B$. Thus, the CV-A and the CV-B provide bounds on the required level of compensation. Also, when $\alpha_1 = \alpha_0$ and $d_1 \neq d_0$ (so that only the

[21] Additional continuity assumptions are required to guarantee that the individual is adequately compensated when the level of compensation *equals* CV-A (or CV-B).

ancillary condition changes), $m^A \geq 0 \geq m^B$. In other words, the welfare effect of a change in the ancillary condition, by itself, is always ambiguous.

Example 7: Let's revisit the application discussed in Section 3.5.1 (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 α 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 2 depicts the bundles (y_0, z_0) and (y_1, z_1), as well as the the CV-A and the CV-B for this change. The CV-A is given by the horizontal distance (y_1, z_1) and point a, because $(y_1 + m^A + \varepsilon, z_1)$ is chosen over (x_0, m_0) for all ancillary conditions and $\varepsilon > 0$. The CV-B is given by the vertical distance betwene (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. □

For the reasons discussed in Section 3.6, it is important to establish that our generalized framework for welfare analysis converges to the standard framework as behavioral anomalies become small. Notably, in Example 7, both the CV-A and the CV-B converge to the standard notion of compensating variation as d_H approaches d_L . Our next result (for which we again invoke Assumption 3) establishes this convergence property under innocuous assumptions concerning $X(\alpha, m)$ and u.

Theorem 6: *Suppose that the sequence of choice correspondences C^n weakly converges to $\widehat{C}$, where $\widehat{C}$ is defined on $\mathcal{X}^c$, and reflects maximization of a continuous utility function, u. Assume that $X(\alpha, m)$ is compact for all α and m, and continuous in m.*[22] *Also assume that $\max_{x \in X(\alpha,m)} u(x)$ is strictly increasing in m for all α. Consider a change from (α_0, d_0) to (α_1, d_1). Let $\widehat{m}$ be the standard comnpensating variation derived from $\widehat{C}$. Let m_A^n be the CV-A, and m_B^n be the CV-B derived from C^n. Then $\lim_{n\to\infty} m_A^n = \lim_{n\to\infty} m_B^n = \widehat{m}$.*

[22] $X(\alpha, m)$ is continuous in m if it is both upper and lower hemicontinuous in m.

Proof: See the Appendix.

The 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 the 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.[23] Similarly, if he experiences a sequence of changes and is not adequately compensated for any of them in the sense of the 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 1.

In contrast to the standard framework, the compensating variations (either CV-As or CV-Bs) associated with each step in a sequence of changes needn't be additive.[24] However, we are not particularly troubled by non-additivity. If one wishes to determine the size of the payment that compensates for a collection of changes, it is appropriate to consider these changes together, rather than sequentially. The fact that the individual could be induced to pay (or accept) a different amount, in total, provided he is "surprised" by the sequence of changes (and treats each as if it leads to the final outcome) does not strike us as a serious conceptual difficulty.

4.2 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 will continue to study the environment introduced in Section 3.5.1 and revisited in Example 7. However, we will assume here that positive analysis delivers the following more restrictive utility representation

[23] For example, if m_1^A is the CV-A for a change from $(X(\alpha_0, 0), d_0)$ to $(X(\alpha_1, m), d_1)$, and if m_2^A is the CV-A for a change from $(X(\alpha_1, m_1^A), d_1)$ to $(X(\alpha_2, m_1^A + m), d_2)$, then nothing that the individual would choose from $(X(\alpha_0, 0), d_0)$ is unambiguously chosen over anything that he would choose from $(X(\alpha_2, m_1^A + m_2^A), d_2)$.

[24] In the standard framework, if m_1 is the CV for a change from $X(\alpha_0, 0)$ to $X(\alpha_1, m)$, and if m_2 is the CV for a change from $X(\alpha_1, m_1)$ to $X(\alpha_2, m_1 + m)$, then $m_1 + m_2$ is the CV for a change from $X(\alpha_0, 0)$ to $X(\alpha_2, m)$. The same statement does not necessarily hold within our framework.

(which involves no income effects, so that Marshallian consumer surplus would be valid in the standard framework):

$$U(y, z \mid d) = y + dv(z) \tag{8}$$

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

Let M denote the consumer's initial income. Consider a change in the price of z from p_0 to p_1, along with a change in ancillary conditions from d_0 to d_1 (potentially, either $p_0 = p_1$ or $d_0 = d_1$). Let z_0 denote the amount of z purchased with (p_0, d_0), and let z_1 denote the amount purchased with (p_1, d_1); assume that $z_0 > z_1$. Since there are no income effects, z_1 will not change as the individual is compensated (holding the ancillary condition fixed). The following result provides a simple formula for the CV-A and CV-B associated with the change from (p_0, d_0) to (p_1, d_1):

Theorem 7: *Suppose that decision utility is given by equation* (8). *The CV-A and CV-B associated with a change from* (p_0, d_0) *to* (p_1, d_1) *are:*

$$m^A = [p_1 - p_0]z_1 + \int_{z_1}^{z_0} [P(z, d_H) - p_0]dz \tag{9}$$

$$m^B = [p_1 - p_0]z_1 + \int_{z_1}^{z_0} [P(z, d_L) - p_0]dz \tag{10}$$

Proof: To calculate the CV-A, we must find the infinum of the values of m that satisfy

$$U(M - p_1 z_1 + m', z_1 \mid d) > U(M - p_0 z_0, z_0 \mid d) \text{ for all } m' \geq m \text{ and } d \in [d_L, d_H]$$

Notice that this requires

$$m \geq [p_1 z_1 - p_0 z_0] + d[v(z_0) - v(z_1)] \text{ for all } d \in [d_L, d_H]$$

Since $v(z_0) > v(z_1)$, the solution is

$$m^A = [p_1 z_1 - p_0 z_0] + d_H[v(z_0) - v(z_1)]$$

$$
\begin{aligned}
&= [p_1 z_1 - p_0 z_0] + \int_{z_1}^{z_0} d_H v'(z)\, dz \\
&= [p_1 - p_0] z_1 + p_0 z_1 - p_0 [z_0 - z_1] - p_0 z_1 + \int_{z_1}^{z_0} d_H v'(z)\, dz \\
&= [p_1 - p_0] z_1 + \int_{z_1}^{z_0} [d_H v'(z) - p_0] dz
\end{aligned}
$$

The derivation of (10) is analogous. Q.E.D.

The first term in (9) is the extra amount the consumer ends up paying for the first y_1 units. The second term is the area under the demand curve and above a horizontal line at p_0 between y_1 and y_0, when d_H is the ancillary condition. Figure 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 labelled A and B.

Notice that (10), the formula for CV-B, is the same as (9), except that we use the area under the demand curve associated with d_L, rather than the one associated with d_H. Figure 3(b) provides a graphical representation of CV-B: it is the sum of the areas labelled A and C, minus the area labeled E.

As the figure 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, in accordance with Theorem 6.

5 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.

5.1 Generalized Pareto optima

Suppose there are N individuals indexed $i = 1, ..., N$. Let $\mathbb{X}$ 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 $\mathcal{G}_i$ (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 $\mathbb{X}$.

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.[25,26]

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 $(X, d) \in \mathcal{G}$ (in which case we have $\sim[yP_i^*x$ 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 curve. Is there also usually a continuum of generalized Pareto optima spanning the gap between the extreme cases described in the

[25] Between these extremes, there are two intermediate notions of Pareto optimality. One could replace P_i^* with P_i' in the definition of a weak generalized Pareto optimum, or replace R_i' with P_i' and P_i' with P_i^* in the definition of a strict generalized Pareto optimum. One could also replace P_i^* with P_i' in the definition of a strict generalized Pareto optimum.

[26] If one thinks of P^* as a preference relation, then our notion of a weak generalized Pareto optimum coincides with existing notions of social efficiency when consumers have incomplete and/or intransitive preferences (see, e.g., Fon and Otani [1979], Rigotti and Shannon [2005], or Mandler [2006]). It is important to keep in mind that, in that literature, an individual is always willing to select any element of a choice set X that is maximal under under the preference relation. In contrast, in our framework, an individual is not necessarily willing to select any element of X that is maximal under the individual welfare relation P^*. However, for the limited purpose of characterizing socially efficient outcomes, choice is not involved, so that distinction is immaterial. Thus, existing results concerning the structure or characteristics of the Pareto efficient set with incomplete and/or intransitive preferences apply in our setting; we mention an example below.

previous paragraph? The following example answers this question in the context of a two-person exchange economy.

Example 8: Consider a two-person exchange economy involving two goods, a and b. Suppose the choices of consumer 1 are described by the positive model set forth in Section 3.5.1 (concerning coherent arbitrariness), while consumer 2's choices respect standard axioms. In Figure 4, the area between the curves labeled T_H (formed by the tangencies between the consumers' 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.[27] 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 cuve. Thus, the standard framework once again emerges as a limiting case of our framework, in which behavioral anomolies become vanishingly small. □

More generally, in standard settings (with continuous preferences and a compact set of social alternatives X), one can start with *any* alternative $x \in X$, and find a Pareto optimum in $\{y \mid yR_i x \text{ for all } i\}$, 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 8: *For every $x \in X$, the non-empty set $\{y \in X \mid \forall i, \ \sim xP_i^* y\}$ includes at least one weak generalized Pareto optimum in X.*

[27] Notably, in another setting with incomplete preferences, Mandler [2006] demonstrates with generality that the Pareto efficient set has full dimensionality.

Proof: Consider the following set:

$$U^*(x,X) = \{y \in X \mid \forall i, \ \sim xP_i^*y \text{ and} \nexists a_1, ..., a_N \text{ s.t. } xP_i^*a_1P_i^*a_2...a_NP_i^*y\}$$

Plainly, $U^*(x,X) \subseteq \{y \in X \mid \forall i, \ \sim xP_i^*y\}$. We will establish the theorem by showing that $U^*(x,X)$ contains a weak generalized Pareto optimum.

First we claim that, if $z \in U^*(x,X)$ and there is some $w \in X$ such that wP_i^*z for all i, then $w \in U^*(x,X)$. Suppose not. Then for some k, there exists $a_1, ..., a_N$ s.t. $xP_k^*a_1P_k^*a_2...a_NP_k^*wP_k^*z$. But that implies $z \notin U^*(x,X)$, a contradiction.

Now we prove the theorem. Take any individual i. Choose any $z \in C_i\left(U^*(x,X), d\right)$ for some d with $(U^*(x,X), d) \in \mathcal{G}$. We claim that z is a weak generalized Pareto optimum. Suppose not. Then there exists $w \in X$ such that wP_j^*z for all j. From the lemma, we know that $w \in U^*(x,X)$. But then since $w, z \in U^*(x,X)$ and wP_i^*z, we have $z \notin C_i(U^*(x,X), d)$, a contradiction. Q.E.D.

Notice that Theorem 8 does not require additional assumptions concerning compactness or continuity. Rather, existence follows from the fundamental assumption that the choice correspondence is non-empty over its domain.[28]

For the reasons discussed in Section 3.6, it is once again important to establish that our generalized framework for welfare analysis converges to the standard framework as behavioral anomalies become small. Our next result (for which we again invoke Assumption 3) establishes the generalized Pareto optima have this convergence property.[29] The statement of the theorem requires the following notation: for any set X, choice domain $\mathcal{G}$, and collection of choice correspondences (one for each individual) $C_1, ..., C_J$ defined on $\mathcal{G}$, let $W(X; C_1, ..., C_J, \mathcal{G})$ denote the set of weak generalized Pareto optima within X. (When

[28] The proof of Theorem 9 is more subtle than one might expect; in particular, there is no guarantee that any individual's welfare optimum within the set $\{y \in X \mid \forall i, \ \sim xP_i^*y\}$ is a Pareto optimum within X.

[29] It follows from Theorem 9 that, for settings in which the Pareto efficient set is "thin" (that is, of low dimensionality) under standard assumptions, the set of generalized Pareto optima is "almost thin" as long as behavioral anomalies are not too large. Thus, unlike Mandler [2006], we are not troubled by the fact that the Pareto efficient set with incomplete preferences may have high (even full) dimensionality.

the choice set is compact and the choice correspondences reflect utility maximization, we will engage in a slight abuse of notation by writing the set of weak Pareto optima as $W(X; C_1, ..., C_J, \mathcal{X}^c)$).

Theorem 9: *Consider any sequence of choice correspondence profiles,* $(C_1^n, ..., C_J^n)$*, such that* C_i^n *weakly converges to* $\widehat{C}_i$*, where* $\widehat{C}_i$ *is defined on* $\mathcal{X}^c$ *and reflects maximization of a continuous utility function,* u_i*. For any* $X \in \mathcal{X}$ *and any sequence of alternatives* $x^n \in W(X; C_1^n, ..., C_J^n, \mathcal{G})$*, all limit points of the sequence lie in* $W(\text{clos}(X), \widehat{C}_1, ..., \widehat{C}_J, \mathcal{X}^c)$.

Proof: See the Appendix.

5.2 The efficiency of competitive equilibria

The notion of a generalized Pareto optimum easily lends itself to formal analysis. To illustrate, we provide a generalization of the first welfare theorem.

Consider an economy with N consumers, F firms, and K goods. We will use x^n to denote the consumption vector of consumer n, z^n to denote the endowment vector of consumer n, $\mathbb{X}^n$ to denote consumer n's consumption set, and y^f to denote the input-output vector of firm f. Feasibility of production for firm f requires $y^f \in Y^f$, where the production sets Y^f are characterized by free disposal. We will use Y to denote the aggregate production set. We will say that an allocation $x = (x^1, ..., x^N)$ is *feasible* if $\sum_{n=1}^N (x^n - z^n) \in Y$ and $x^n \in X^n$ for all n.

The conditions of trading involve a price vector π and a vector of ancillary conditions, $d = (d^1, ..., d^N)$, where d^n indicates the ancillary conditions applicable to consumer n. The price vector π implies a budget constraint $B^n(\pi)$ for consumer n – that is, $B^n(\pi) = \{x^n \in \mathbb{X}^n \mid \pi x^n \leq \pi z^n\}$.

We assume that profit maximization governs the choices of firms. Consumer behavior is described by a choice correspondence $C^n(X^n, d^n)$ for consumer n, where X^n is a set of available consumption vectors, and d^n represents the applicable ancillary condition. Let R_n be the welfare relation on $\mathbb{X}^n$ obtained from $(\mathcal{G}^n, C^n)$ (similarly for P_n and P_n^*).

A *behavioral competitive equilibrium* involves a price vector, $\widehat{\pi}$, a consumption allocation, $\widehat{x} = (\widehat{x}^1, ..., \widehat{x}^N)$, a production allocation, $\widehat{y} = (\widehat{y}^1, ..., \widehat{y}^F)$, and a set of ancillary conditions $\widehat{d} = (\widehat{d}^1, ..., \widehat{d}^N)$, such that (i) for each n, we have $\widehat{x}^n \in C^n(B^n(\widehat{\pi}), \widehat{d}^n)$, (ii) $\sum_{n=1}^N (\widehat{x}^n - z^n) = \sum_{f=1}^F \widehat{y}^f$, and (iii) $\widehat{y}^f$ maximizes $\widehat{\pi} y^f$ for $y^f \in Y^f$.

Fon and Otani [1979] have shown that a competitive equilibrium of an exchange economy is Pareto efficient even when consumers have incomplete and/or intransitive preferences (see also Rigotti and Shannon [2005] and Mandler [2006]). One can establish the efficiency of a behavioral competitive equilibrium for an exchange economy as a corollary of their theorem.[30] We offer a direct proof below to incorporate production, underscore the simplicity of the argument, and highlight its similarility to the standard demonstration of the first welfare theorem.[31]

Theorem 10: *The allocation associated with any behavioral competitive equilibrium is a weak generalized Pareto optimum.*[32]

Proof: Suppose on the contrary that x is not a weak generalized welfare optimum. Then, by definition, there is some feasible allocation $\widehat{w}$ such that $\widehat{w}^n P_n^* \widehat{x}^n$ for all n.

The first step is to show that if $w^n P_n^* \widehat{x}^n$, then $\widehat{\pi} w^n > \widehat{\pi} \widehat{x}^n$. Take any w^n with $\widehat{\pi} w^n \leq \widehat{\pi} \widehat{x}^n$. Then $w^n \in B^n(\widehat{\pi})$. Because $\widehat{x}^n \in C^n(B^n(\widehat{\pi}), \widehat{d}^n)$, we conclude that $\sim w^n P_n^* \widehat{x}^n$.

Combining this first observation with the market clearing condition, we see that

$$\widehat{\pi} \sum_{n=1}^{N} (\widehat{w}^n - z^n) > \widehat{\pi} \sum_{n=1}^{N} (\widehat{x}^n - z^n) = \widehat{\pi} \sum_{f=1}^{F} \widehat{y}^f$$

[30] Let $m_{P_i^*}(X)$ denote the maximal elements of X under P_i^*. Consider an alternative exchange economy in which $m_{P_i^*}(X)$ is the choice correspondence for consumer i. According to Theorem 1 of Fan and Otani [1979], the competitive equilibria of that economy are Pareto efficient, when judged according to P_1^*,...,P_N^*. For any behavioral competitive equilibrium, there is necessarily an equivalent equilibrium for the alternative economy. (Note that the converse is not necessarily true.) Thus, the behavioral competitive equilibrium must be a generalized Pareto optimum.

[31] Presumably, one could also address the existence of behavioral competitive equilibria by adapting the approach developed in Mas-Colell [1974], Gale and Mas-Colell [1975], and Shafer and Sonnenschein [1975].

[32] One can also show that a behavioral competitive equilibrium is a strict generalized Pareto optimum under the following additional assumption (which is akin to non-satiation): if $x^n, w^n \in X^n$ and $x^n > w^n$ (where $>$ indicates a strict inequality for every component), then $w^n \notin C^n(X^n, d^n)$ for any d^n with $(X^n, d^n) \in \mathcal{G}^n$. In that case, $w^n R_n \widehat{x}^n$ implies $\widehat{\pi} w^n \geq \widehat{\pi} \widehat{x}^n$; otherwise, the proof is unchanged.

Moreover, since $\widehat{w}$ is feasible, we know that $\sum_{n=1}^{N}(\widehat{w}^n - z^n) \in Y$, or equivalently that there exists $v = (v^1, ..., v^F)$ with $v^f \in Y^f$ for each f such that $\sum_{n=1}^{N}(\widehat{w}^n - z^n) = \sum_{f=1}^{F} v^f$, from which it follows that

$$\widehat{\pi}\sum_{n=1}^{N}(\widehat{w}^n - z^n) = \widehat{\pi}\sum_{f=1}^{F} v^f$$

Combining the previous two equations yields

$$\widehat{\pi}\sum_{f=1}^{F} v^f > \widehat{\pi}\sum_{f=1}^{F}\widehat{y}^f$$

But this can only hold if $\widehat{\pi}v^f > \widehat{\pi}\widehat{y}^f$ for some f. Since $v^f \in Y^f$, this contradicts the assumption that $\widehat{y}^f$ maximizes firm f's profits given $\widehat{\pi}$. Q.E.D.

The generality of Theorem 10 is worth emphasizing: it establishes the efficiency of competitive equilibria within a framework that imposes almost no restrictions on consumer behavior, thereby allowing (as argued in Section 2.2) for virtually any conceivable choice pattern, including all anomalies documented in the behavioral literature. Note, however, that we have not relaxed the assumption of profit maximization by firms; moreover, the theorem plainly need not hold if firms pursue other objectives. Thus, we see that the first welfare theorem is driven by assumptions concerning the behavior of firms, not consumers.

Example 9: Figure 5 illustrates this result for the simple two-person two-good exchange economy considered in example 8. Initial endowments correspond to point z. In equilibrium, ancillary condition d' prevails, and prices permit the consumers to trade along the straight line through z. Both consumers trade to point w. In consumer 1's case, this point corresponds to the tangency between the budget line and an "indifference curve" for ancillary condition d', labeled $I_{1d'}$. The area labeled A, which lies above the indifference curves labeled I_{1L} (corresponding to ancillary condition d_L) and I_{1H} (corresponding to ancillary condition d_H) contains all allocations v for which vR_1w. The area labeled B, which lies below the indifference curve labeled I_2, contains all allocations v for which vR_2w. Since these area do not overlap (except at w), w is a strict generalized Pareto optimum. □

5.3 Market failures

Just as it is possible to establish the efficiency of behavioral competitive equilibrium with perfectly competitive markets, it is also possible to demonstrate the inefficiency of equilibria in the presence of sufficiently severe but otherwise standard market failures. Here, we show by way of example that a sufficiently large externality leads to inefficiency.

Example 10: Consider the same two-person two-good exchange economy as in examples 8 and 9. Suppose that positive analysis delivers the following utility representation of consumer 1's behavior (where the relevant choice experiments would necessarily include allocations for consumer 2):

$$U(a_1, b_1 \mid \lambda, d) = a_1 + dv(b_1) - ae(a_2)$$

where λ is a positive constant and $e(a_2)$ is a strictly increasing convex function. Otherwise, all of our assumptions are unchanged. In this setting, consumer 1's enjoyment of good a inflicts a negative externality on consumer 1, the size of which depends on the parameter λ.

Since the effect of a_2 on consumer 1's utility representation is separable, the presence of the term $-\lambda e(a_2)$ does not affect consumer 1's choices. Accordingly, changing λ does not alter the set of competitive equilibrium. Figure 6 reproduces the competitive equilibrum from Figure 5. Here, the "indifference curves" for consumer 1, now labeled I_{1H}^1 and I_{1L}^1, correspond to the portion of the positive model that governs choices over a_1 and b_1, holding a_2 fixed. In other words, they ignore the term $-\lambda e(a_2)$. To this figure, we've added the indifference curves labeled I_{1L}^2 and I_{1H}^2. These are based on the positive model that governs choices over a_1 and b_1, imposing the constraint that $a_1 + a_2$ equals the total endowment of a. Notice that these curves are steeper than the versions that hold a_2 fixed. This is because incremental consumption of a_1 creates an additional benefit when it forces a reduction in a_2; consequently, less a is required to compensate for a given loss of b. Now the area labeled A', which lies above the indifference curves labeled I_{1L}^2 (corresponding to ancillary condition d_L) and I_{1H}^2 (corresponding to ancillary condition d_H) contains all allocations v for which

vP_1^*w. For small values of λ, this area still does not overlap with area B, so the equilibrium remains a strict generalized Pareto optimum. However, for sufficiently large values of λ, the areas A' and B will overlap, as shown (since I_{1H}^2 will be steeper than I_2 at w), which means that the behavioral competitive equilibrium will not be a 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 economies, there is a new class of potential market failures involving choices made in the presence of problemmatic ancillary conditions. Our analysis of addiction (Bernheim and Rangel [2004]) exemplifies this possibility.

6 Refining the welfare relations

We have seen 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 problem tends to arise 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 for these types of refinements.

6.1 Refinement strategies

The following simple observation (the proof of which is trivial) indicates how the addition or deletion of data affects the coarseness of the welfare relation and the sets of weak and strict individual welfare optima.

Observation 3: Fix $\mathbb{X}$. Consider two generalized choice domains $\mathcal{G}_1$ and $\mathcal{G}_2$ with $\mathcal{G}_1 \subset \mathcal{G}_2$. 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 $G \in \mathcal{G}_1$.

(a) The welfare relations R'_2 and P^*_2 obtained from $(\mathcal{G}_2, C_2)$ are weakly coarser than the welfare relations R'_1 and P^*_1 obtained from $(\mathcal{G}_1, C_1)$.

(b) If $x \in X$ is a weak welfare optimum for X based on $(\mathcal{G}_1, C_1)$, it is also a weak welfare optimum for X based on $(\mathcal{G}_2, C_2)$.

(c) Suppose that $x \in X$ is a strict welfare optimum for X based on $(\mathcal{G}_1, C_1)$, and that there is no $y \in X$ such that xI'_1y. Then x is also a strict welfare optimum for X based on $(\mathcal{G}_2, C_2)$.

It follows that the addition of data (that is, the expansion of $\mathcal{G}$) makes R' and P^* weakly coarser, while the elimination of data (that is, 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.[33] Also, 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.

Observation 3 motivates an agenda involving refinements of the welfare relations considered in this paper. The goal of this agenda is to make the 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 3 rules 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

[33] Notice, however, the same principle does not hold for P' or R^*. Suppose, for example, that xI'_1y given $(\mathcal{G}_1, C_1)$, so that $\sim xP'_1y$. Then, with the addition of a GCS for which x is chosen but y is not with both available, we would have xP'_2y; in other words, the relation P' would become finer. Similarly, suppose that xP^*_1y given $(\mathcal{G}_1, C_1)$, so that $\sim yR^*_1x$. Then, with the addition of GCS for which y is chosen when x is available, we would have yR^*_2x; in other words, the relation R^* would become finer.

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 himself 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.

6.2 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., Koszegi and Rabin [2007]), we suspect that, in most cases, this task is beyond the current capabilities of economics, neuroscience, and psychology.

We nevertheless submit that there are circumstances in which non-choice 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 non-choice evidence to officiate between conflicting choice data by deleting suspect GCSs. Thus, for example, if someone chooses x from X under condition d' where he is likely to be distracted, and chooses y from X under condition d'' where he 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 become (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 the use of non-choice data, or any external judgment, as either a substitute for or supplement to choice data. Within this framework, all evaluations are ultimately based on the individual's actual choices, and must be consistent with all unambiguous choice patterns.

6.2.1 Forms of non-choice evidence

What forms of non-choice 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 psychology, neuroscience, 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 involving forecasting malfunctions. Citing evidence from neuroscience, we argue that the repeated use of addictive substances causes specific a neural system that measures empirical correlations between cues and potential rewards to malfunction in the presence of identifiable ancillary conditions. Whether or not that system *also* plays a role in hedonic experience, the choices made in the preence of those conditions are therefore suspect, and welfare evaluations should be guided by choices made under other 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 opthalmologist 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, non-choice evidence sheds light on the likelihood that the individual successfully processed information that was in principle available to him, thereby properly characterizing 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.

6.2.2 What is a mistake?

The concept of a *mistake* does not exist within the context of standard choice-theoretic welfare economics. Within our framework, however, one can define *mistake* as a choice made in a suspect GCS that is contradicted by choices in non-suspect GCSs. In other words, if the individual chooses $x \in X$ in one GCS where he properly understands that the choice set is X, and chooses $y \in X$ in another GCS where he 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 he believes he makes is, by definition, not a mistake given the set from which he believes he is choosing.

In Bernheim and Rangel [2004], 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."

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 non-choice evidence that the pedestrian attends to traffic patterns.

6.3 Refinements based on coherence

In some instances, it may be possible to partition behavior into coherent patterns and isolated anomalies. One might then argue that, for the purpose of welfare analysis, it is apporpriate to respect the coherent aspects of choice and ignore the anomalies. This argument suggests another potential approach to refining the welfare relations:identify subsets of GCSs, corresponding to particular ancillary conditions, within which choice is coherent, in the classic sense that it reflects the maximal elements a preference relation on $\mathbb{X}$. Then construct welfare relations based on those GCSs, and ignore other choice data.

Unfortunately, the coherence criterion raises difficulties. Every choice is coherent taken by itself. Accordingly, some form of minimum domain requirement is needed, and there is no obvious way to set this requirement objectively.

In some circumstances, however, the coherence criterion seems reasonably natural (based once again on non-choice considerations). Take, for example, the problem of intertemporal consumption allocation for a β, δ consumer (discussed in Section 3.5.2). For many GCSs, the allocation is determined by a sequence of choices at many different points in time. However, for each point in time t, there is a class of GCSs, call it $\mathcal{G}_t$, for which all discretion is exercised at time t, through a broad precommitment. Within each $\mathcal{G}_t$, all choices reflect maximization of the same time t utility function. Therefore, each $\mathcal{G}_t$ identifies a set of GCSs for which choices are coherent. Based on the coherence criterion, one might therefore construct our welfare relations restricting attention to $\mathcal{G}_c = \mathcal{G}_1 \cup \mathcal{G}_2 \cup ... \cup \mathcal{G}_T$. We will call these relations R_c' and P_c^*. For all $G \in \mathcal{G}_c$, the ancillary condition is completely described by the point in time at which all discretion is resolved. Thus, we can write any such G as (X, t).

Based on Theorem 3, it is natural to conjecture that R_c' and P_c^* corresponds to the weak and strict multi-self Pareto criterion. However, that theorem does not apply because $\mathcal{G}_c$ is not rectangular. Consider in particular any consumption set $X \in \mathcal{X}$ such that $(X,t) \in \mathcal{G}_t$. For any two consumption vectors, x' and x'' in X, it must be the case that $x'_k = x''_k$ for all $k < t$. This is because one can never make a choice in period t affecting past consumption.

Our next result characterizes individual welfare optima under R'_c and P^*_c for conventional intertemporal budget constraints. We will assume that initial wealth, w_1, is strictly positive. Let $\lambda \equiv \frac{1}{1+r}$, where r is the rate of interest. We will use X_1 denote the standard intertemporal budget set:

$$X_1 = \left\{ (c_1, ..., c_T) \in \mathbf{R}^T_+ \mid w_1 \geq \sum_{k=1}^{T} \lambda^{k-1} c_t \right\}$$

We will also use $X_t(c'_1, .., c'_{t-1})$ denote the continuation budget set, given theat the individual has consumed c'_1,...,c'_{t-1}:

$$X_t(c'_1, ..., c'_{t-1}) = \left\{ (c'_1, ..., c'_{t-1}, c_t, ..., c_T) \in \mathbf{R}^T_+ \mid w_1 - \sum_{k=t}^{t-1} \lambda^{k-1} c'_t \geq \sum_{k=t}^{T} \lambda^{k-1} c_t \right\}$$

At time t, all discretion is resolved to maximize the function given in (7). We define

$$V_t(C_t) = \sum_{k=t}^{T} \delta^{k-t} u(c_k)$$

In other words, $V_t(C_t)$ is conventional discounted utility. We will also assume that $u(c)$ is continuous and strictly concave.

Theorem 11: *For welfare evaluations based on R'_c and P^*_c:*

(i) *the consumption vector C^*_1 is an individual welfare optimum in X_1 (both weak and strict) iff C^*_1 maximizes $U_1(C_1)$.*

(ii) *for any feasible $(c'_1, ..., c'_{t-1})$, the consumption vector C^*_1 is an individual welfare optimum (both weak and strict) in $X_t(c'_1, ..., c'_{t-1})$ iff C^*_1 maximizes $\alpha U_t(C_t) + (1-\alpha) V_t(C_t)$ for some $\alpha \in [0,1]$.*

Proof: See the Appendix.

According to Theorem 11, individual welfare optimality within X_1 under R^c is completely governed by the perspective of the individual at the first moment in time. Thus, the special status of $t = 1$, which we noted in the context of Theorem 4, is amplified when attention is restricted to $\mathcal{G}^c$. In any period $t > 1$, there is some ambiguity concerning the tradeoff between current and future consumption, with standard discounting and β,δ discounting bracketing the range of possibilities. However, a sequence of individual welfare optima for periods $t = 1, ..., T$ is time consistent if and only if it coincides with the maximization of U_1. Assuming that the first period is short, Theorem 11 therefore provides a potential formal justification for the long-run criterion.

What accounts for the dominance of the $t = 1$ perspective, and are the implications of Theorem 11 reasonable? To anwer these questions, it is helpful to understand the relationship between P'_c, P^*_c, and the multi-self Pareto criteria. If the domain of generalized choice situations were rectangular, the relations P' and P^* would coincide with the weak and strict multi-self Pareto relations (Theorem 3). Note that we can make the domain rectangular by hypothetically extending the choice correspondence C to include choices involving past consumption. If we then delete the hypothetical choice data, the welfare relations become more discerning, and the set of individual welfare optima shrinks (Observation 3). Thus, the set of individual welfare optima under P'_c and P^*_c must be contained in the set of multi-self Pareto optima *for every conceivable set of hypothetical data on backward-looking choices.* In other words, P'_c and P^*_c identify multi-self Pareto improvements that are robust with respect to all conceivable assumptions concerning such counterfactual choices.

This discussion identifies a conceptual deficiency in conventional notions of mult-self Pareto efficiency, which assumes that the time t self does not care about the past (see, e.g., Laibson et. al. [1998], or Bhattacharya and Lakdawalla [2004]).[34] Since there can be no choice experiments involving backward-looking decision, this assumption (as well as any alternative assumption) is untestable and unwarranted. In light of our inherent ignorance

[34] Other assumptions concerning backward-looking preferences appear in the literature; see, e.g., Imrohoroglu, Imrohoroglu, and Joines [2003].

concerning the nature of backward looking preferences, it would seem more appropriate to adopt a notion of multi-self Pareto efficiency that is robust with respect to a wider range of possibilities.

Imagine then that the period t self can make decisions for past consumption as well as for future consumption; moreover, decisions of the period t self correspond to maximization of the utility function

$$\widehat{U}_t(C_t) = \Gamma_t(c_1, ..., c_{t-1}) + u(c_t) + \beta \sum_{k=t+1}^{T} \delta^{k-t} u(c_k)$$

This is the same objective function as in the β,δ setting (equation (7)), except that preferences are both backward looking and forward looking. We note that no conceivable choice experiment could possibly identify Γ. Accordingly, we will say that C_1 is a (weak or strict) *robust multi-self Pareto optimum* if it is a (weak or strict) multi-self Pareto optimum for all possible $(\Gamma_2, ..., \Gamma_T)$.[35] Arguably, we should place some minimal restrictions on the Γ_t, for example that they are continuous and increasing, but such restrictions do not affect the following result:

Theorem 12: *A consumption vector C_1 is a both a weak and a strict robust multi-self Pareto optimum in X_1 iff it maximizes $U_1(C_1)$.*

Proof: See the Appendix.

Together, Theorems 11 and 12 imply that the set of individual welfare optima under P_c' and P_c^* coincides exactly with the set of robust multi-self Pareto optima. Intuitively, the time t perspective dominates robust multi-self Pareto comparisons, and thus P_c' and P_c^*, because we lack critical information (backward-looking preferences) concerning all other perpspectives.

[35] We omit Γ_1 because there is no consumption prior to period 1.

6.4 Refinements based on preponderance

Another natural criterion for officiating between conflicting choices is preponderance. In other words, if someone ordinarily chooses x over y (that is, in almost all choice situations where both are available and one is chosen), and rarely chooses y over x, it might be appropriate to disregard the exceptions and follow the rule. It appears that this criterion is often invoked (at least implicitly) in the literature on quasi-hyperbolic (β,δ) discounting to justify welfare analysis based on long-run preferences.

Conceptually, we see two serious problems with the preponderance criterion. First, the use of this criterion presupposes that there is some natural measure on $\mathcal{G}$. The nature of this measure is far from obvious. Since it is presumably easy to proliferate variations of ancillary conditions, one cannot simply count GCSs. There may also be competing notions of preponderance. For example, in the QHD environment, there is an argument for basing preponderance on commonly encountered, and hence familiar, GCSs. If the individual makes most of his decisions “in the moment,” this notion of preponderance would favor the short-run perspective.

Second, a rare ancillary condition may be highly conducive to good decision-making. That would be the case, for example, if an individual typically misunderstands available information concerning his alternatives unless it is presented in a particular way. Likewise, in the QHD setting, one could argue that people may appreciate their needs most accurately when those needs are immediate and concrete, rather than distant and abstract.

We suspect that the economics profession’s “revealed preference” for the long-run welfare perspective emerges from the widespread belief that short-run decisions sometimes reflect lapses of self-control, rather than an inclination to credit preponderance. Implicitly or explicitly, we recognize a choice as a lapse of control based on non-choice considerations, such as introspection.

7 Discussion

In this paper, 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.

Our framework is a natural generalization of standard welfare economics 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.

In principle, our framework encompasses all behavioral models; it is applicable irrespective of the processes generating behavior, or of the positive model used to describe behavior. 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. Thus, the framework potentially opens the door to greater integration of economics, psychology, and cognitive neuroscience.

Like standard welfare economics, our framework requires only data on choices. 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.

We have also demonstrated that our framework is easily applied. It leads to novel normative implications for the familiar β, δ model of time inconsistency. For a model of coherent arbitariness, it provides a choice-theoretic (non-pscyhological) justifications for multi-self Pareto optimality. It generates natural counterparts for the standard tools of applied welfare analysis, including compensating and equivalent variation, consumer surplus, Pareto optimality, and the contract curve, and permits a broad generalization of the of the first welfare theorem.

Finally, we have also suggested (perhaps more controversially) that our framework lends itself to principled refinements, some of which may rely circumscribed but systematic use of non-choice data. Significantly, we do not propose the use of non-choice data, or any external judgment, as either a substitute for or supplement to choice data. Within this framework, all evaluations are ultimately based on the individual's actual choices, and must be consistent with all unambiguous choice patterns. Non-choice 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 also has some limitations. First, in some applications, our welfare criteria may not be particularly discriminating. In such cases, the refinement agenda discussed in Section 6 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 insufficent 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.

References

[1] Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "Coherent Arbitrariness: Stable Demand Curves without Stable Preferences." *Quarterly Journal of Economics*, 118(1):73-105.

[2] Arrow, Kenneth J. 1959. "Rational Choice Functions and Orderings." *Economics*, 26(102): 121-127.

[3] Bernheim, B. Douglas, and Antonio Rangel. 2004. "Addiction and Cue-Triggered Decision Processes." *American Economic Review,* 94(5):1558-90.

[4] Bhattacharya, Jay, and Darius Lakdawalla. 2004. "Time-Inconsistency and Welfare." NBER Working Paper No. 10345.

[5] Bossert, Walter, Yves Sprumont, and Kotaro Suzumura. 2005. "Consistent Rationalizability." *Economica*, 72: 185-200.

[6] Caplin, Andrew, and John Leahy. 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *The Quarterly Journal of Economics,* 116(1): 55-79.

[7] Ehlers, Lars, and Yves Sprumont. 2006. "Weakened WARP and Top-Cycle Choice Rules." Mimeo, University of Montreal.

[8] Fon, Vincy, and Yoshihiko Otani. 1979. "Classical Welfare Theorems with Non-Transitive and Non-Complete Preferences." *Journal of Economic Theory*, 20: 409-418.

[9] Gale, David, and Andreu Mas-Colell. 1975. "An Equilibrium Existence Theorem for a General Model Without Ordered Preferences." *Journal of Mathematical Economics* 2: 9-15.

[10] Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica,* 69(6):1403-1435.

[11] Gul, Faruk, and Wolfgang Pesendorfer. 2006. "Random Expected Utility." *Econometrica*, forthcoming.

[12] Imrohoroglu, Ayse, Selahattin Imrohoroglu, and Douglas Joines. 2003. "Time-Inconsistent Preferences and Social Security." *Quarterly Journal of Economics*, 118(2): 745-784.

[13] Iyengar, S. S., and M. R. Lepper. 2000. "Why Choice is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology* 79, 995-1006.

[14] Kalai, Ehud, Ariel Rubinstein, and Ran Spiegler. 2002. "Rationalizing Choice Functions by Multiple Rationales." *Econometrica*, 70(6): 2481-2488.

[15] Koszegi, Botond, and Matthew Rabin. 2007. "Revealed Mistakes and Revealed Preferences." Unpublished.

[16] Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2):443-477

[17] Laibson, David, Andrea Repetto, and Jeremy Tobacman. 1998. "Self-Control and Saving for Retirement." *Brookings Papers on Economic Activity*, 1: 91-172.

[18] Mandler, Michael. 2006. "Welfare Economics with Status Quo Bias: A Policy Paralysis Problem and Cure." Mimeo, University of London.

[19] Mas-Colell, Andreu. 1974. "An Equilibrium Existence Theorem Without Complete or Transitive Preferences." *Journal of Mathematical Economics*, 1: 237-246.

[20] O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review*, 89(1):103-24

[21] Read, Daniel, and Barbara van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes*, 76(2): 189-205.

[22] Rigotti, Luca, and Chris Shannon. 2005. "Uncertainty and Risk in Financial Markets." *Econometrica*, 73(1): 203-243.

[23] Rubinstein, Ariel, and Yuval Salant. 2006. "A model of choice from lists." *Theoretical Economics,* 1: 3-17.

[24] Rubinstein, Ariel, and Yuval Salant. 2007. "(A,f) Choice with frames." Mimeo.

[25] Sen, Amartya K. 1971. "Choice Functions and Revealed Preference." *Review of Economic Studies*, 38(3): 307-317.

[26] Shafer, Wayne, and Hugo Sonnenschein, "Equilibrium in Abstract Economies Without Ordered Preferences." *Journal of Mathematical Economics*, 2: 345-348.

[27] Sugden, Robert. 2004. "The Opportunity Criterion: Consumer Sovereignty Without the Assumption of Coherent Preferences." *American Economic Review*, 94(4): 1014-33.

[28] Suzumura, Kotaro. 1976. "Remarks on the Theory of Collective Choice." *Economica*, 43: 381-390.

[29] Tversky, Amos, and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biase." *Science,* 185, 1124-1131.

Appendix

1. Proofs of results for the β, δ model

Proof of Theorem 4: Let

$$V_t(C_t) = \sum_{k=t}^{T} \delta^{k-t} u(c_k)$$

Given our assumptions, we have, for all C_t, $V_t(C_t) \geq U_t(C_t) \geq W(C_t)$, where the first inequality is strict if $c_k > 0$ for any $k > t$, and the second inequality is strict if $c_k > 0$ for any $k > t+1$.

Suppose the individual faces the GCS (X, R). Because the individual is dynamically consistent within each period, we can without loss of generality collapse multiple decision within any single period into a single decision. So a lifetime decision involves a sequence of choices, $r_1, ..., r_T$ (some of which may be degenerate), that generate a sequence of consumption levels, $c_1, ..., c_T$. The choice r_t must at a minimum resolve any residual discretion with respect to c_t. That choice may also impose constraints on the set of feasible future actions and consumption levels (e.g., it may involve precommitments). For any G, a sequence of feasible choices $r_1, ..., r_t$ leads to a continuation problem $G^C(r_1, ..., r_t)$, which resolves any residual discretion in $r_{t+1}, .., r_T$.

With these observation in mind, we establish three lemmas.

Lemma 1: *Suppose that, as of some period t, the individual has chosen $r_1, ..., r_{t-1}$ and consumed c_1^A,...,c_{t-1}^A, and that C_t^A remains feasible for $G^C(r_1, ..., r_{t-1})$. Suppose there is an equilibrium in which the choice from this continuation problem is C_t^B. Then $V_t(C_t^B) \geq U_t(C_t^B) \geq W_t(C_t^A)$.*

Proof: We prove the lemma by induction. Consider first the case of $t = T$. Then $V_T(C_T^B) = U_t(C_t^B) = u(c_T^B)$ and $W_T(C_T^A) = u(c_T^A)$. Plainly, if the individual is willing to choose c_T^B even though c_T^A is available, then $u(c_T^B) \geq u(c_T^A)$.

Now suppose the claim is true for $t+1$; we will prove it for t. By assumption, the individual has the option of making a choice r_t in period t that locks in c_t^A in period t, and that leaves C_{t+1}^A available.

Let $\widehat{C}_{t+1}$ be a continuation trajectory that the individual would choose from that point forward after choosing r_t. Notice that

$$\begin{aligned} U_t(c_t^A, \widehat{C}_{t+1}) &= u(c_t^A) + \beta\delta V_{t+1}(\widehat{C}_{t+1}) \\ &\geq u(c_t^A) + \beta\delta W_{t+1}(C_{t+1}^A) \\ &= W_t(C_t^A) \end{aligned}$$

Since the individual is willing to make a decision at time t that leads to the continuation consumption trajectory C_t^B, and since another period t decision will lead to the continuation consumption trajectory $(c_t^A, \widehat{C}_{t+1})$, we must have

$$U_t(C_t^B) \geq U_t(c_t^A, \widehat{C}_{t+1})$$

Thus, $U_t(C_t^B) \geq W_t(C_t^A)$, and we already know that $V_t(C_t^B) \geq U_t(C_t^B)$. Q.E.D.

Lemma 2: *Suppose $U_1(C_1^B) \geq W_1(C_1^A)$. Then there exists some G for which C_1^B is an equilibrium outcome even though C_t^A is available. If the inequality is strict, there exists some G for which C_1^B is the only equilibrium outcome even though C_t^A is available.*

Proof: We prove this lemma by induction. Consider first the case of $T = 1$. Note that $U_1(C_1^A) = u(c_1^A) = W_1(C_1^A)$. Thus, $U_1(C_1^B) \geq W_1(C_1^A)$ implies $U_1(C_1^B) \geq U_1(C_1^A)$. Let G consist of a single choice between C_1^A and C_1^B made at time 1. With $U_1(C_1^B) \geq U_1(C_1^A)$, the individual is necessarily willing to choose C_1^B; with strict inequality, he is unwilling to choose C_1^A.

Now suppose the claim is true for $T-1$; we will prove it for T. For $\varepsilon \geq 0$, define

$$c_2^\varepsilon \equiv u^{-1}\left[W_2(C_2^A) + \varepsilon\right],$$

and $C_2^\varepsilon = (c_2^\varepsilon, 0, ..., 0)$. Notice that $U_2(C_2^\varepsilon) = W_2(C_2^A) + \varepsilon$. Therefore, by the induction step, there exists a choice problem G' for period 2 forward (a $T-1$ period problem) for which C_2^ε is an equilibrium outcome (the only one for $\varepsilon > 0$) even though C_2^A is available. We construct G as follows. At time 1, the individual has two alternatives: (i) lock in C_1^B, or (ii) choose c_1^A, and then face G'. Provided we resolve any indifference at $t = 2$ in favor of choosing C_2^ε, the decision at time $t = 1$ will be governed by a comparison of $U_1(C_1^B)$ and $U_1(c_1^A, C_2^\varepsilon)$. But

$$\begin{aligned} U_1(c_1^A, C_2^\varepsilon) &= u(c_1^A) + \beta\delta u(c_{t+1}^\varepsilon) \\ &= u(c_1^A) + \beta\delta\left[W_2(C_2^A) + \varepsilon\right] \\ &= W_1(C_1^A) + \beta\delta\varepsilon \end{aligned}$$

If $U_1(C_1^B) = W_1(C_1^A)$, we set $\varepsilon = 0$. The individual is indifferent with respect to his period t choice, and we can resolve indifference in favor of choosing C_1^B. If $U_1(C_1^B) > W_1(C_1^A)$, we set $\varepsilon < \left[U_1(C_1^B) - W_1(C_1^A)\right]/\beta\delta$. In that case, the individual is only willing to pick C_1^B in period 1. Q.E.D.

Lemma 3: *Suppose $W_1(C_1^A) = U_1(C_1^B)$. If there is some G for which C_1^B is an equilibrium outcome even though C_1^A is available, then C_1^A is also an equilibrium outcome.*

Proof: Consider any sequence of actions $r_1^A, ..., r_T^A$ that leads to the outcome $c_1^A, ..., c_T^A$. As in the proof of Lemma 1, let $\widehat{C}_{t+1}$ be the equilibrium continuation consumption trajectory that the individual would choose from $t+1$ forward after choosing $r_1^A, ..., r_t^A$ and consuming $c_1^A, ..., c_1^t$. (Note that $\widehat{C}_1 = C_1^B$.) According to expression (**??**), $U_t(c_t^A, \widehat{C}_{t+1}) \geq W_t(C_t^A)$. Here we will show that if $W_1(C_1^A) = U_1(C_1^B)$ and C_1^B is an equilibrium outcome, then $U_t(c_t^A, \widehat{C}_{t+1}) = W_t(C_t^A)$. The proof is by induction.

Let's start with $t = 1$. Suppose $U_1(c_1^A, \widehat{C}_2) > W_1(C_1^A)$. By assumption, $W_1(C_1^A) = U_1(C_1^B)$. But then, $U_1(c_1^A, \widehat{C}_2) > U_1(C_1^B)$, which implies that the individual will not choose the action in period 1 that leads to C_1^B, a contradiction.

Now let's assume that the claim is correct for some $t-1$, and consider period t. Suppose $U_t(c_t^A, \widehat{C}_{t+1}) > W_t(C_t^A)$. Because $U_t(\widehat{C}_t) \geq U_t(c_t^A, \widehat{C}_{t+1})$ (otherwise the individual would not choose the action that leads to $\widehat{C}_t$ after choosing $r_1^A, ..., r_{t-1}^A$), we must therefore have $U_t(\widehat{C}_t) > W_t(C_t^A)$, which in turn implies $V_t(\widehat{C}_t) > W_t(C_t^A)$. But then

$$\begin{aligned} U_{t-1}(c_{t-1}^A, \widehat{C}_t) &= u(c_{t-1}^A) + \beta\delta V_t(\widehat{C}_t) \\ &> u(c_{t-1}^A) + \beta\delta W_t(C_t^A) \\ &= W_{t-1}(C_{t-1}^A) \end{aligned}$$

By the induction step, $U_{t-1}(c_{t-1}^A, \widehat{C}_t) = W_{t-1}(C_{t-1}^A)$, so we have a contradiction. Therefore, $U_t(c_t^A, \widehat{C}_{t+1}) = W_t(C_t^A)$.

Now we construct a new equilibrium for G for which C_1^A is the equilibrium outcome. We accomplish this by modifying the equilibrium that generates C_1^B. Specifically, for each every history of choices of the form $r_1^A, ..., r_{t-1}^A$, we change the individual's next choice to r_t^A; all other choices in the decision tree remain unchanged.

When changing a decision in the tree, we must verify that the new decision is optimal (accounting for changes at successor nodes), and that the decisions at all predecessor nodes remain optimal. When we change the choice following a history of the form $r_1^A, ..., r_{t-1}^A$, all of the predecessor nodes correspond to histories of the form $r_1^A, ..., r_k^A$, with $k < t-1$. Thus, to verify that the individual's choices are optimal after the changes, we simply check the decisions for all histories of the form $r_1^A, ..., r_{t-1}^A$, in each case accounting for changes made at successor nodes (those corresponding to larger t).

After any history $r_1^A, ..., r_{t-1}^A$, choosing r_t^A in period t leads (in light of the changes at successor nodes) to C_1^A, producing period t decision utility of $U_t(C_t^A)$. Since we have only changed decisions along a single path, no other choice at time t leads to period t decision utility greater than $U_t(\widehat{C}_t)$. For $t \geq 2$, we have established that $U_{t-1}(c_{t-1}^A, \widehat{C}_t) = W_{t-1}(C_{t-1}^A)$, from which it follows that $V_t(\widehat{C}_t) = W(C_t^A)$. But then we have $U_t(\widehat{C}_t) \leq V_t(\widehat{C}_t) = W(C_t^A) \leq U_t(C_t^A)$. Thus, the choice of r_t^A is optimal. For $t = 1$, we have $\widehat{C}_1 = C_1^B$, and we have assumed that $W_1(C_1^A) = U_1(C_1^B)$, so we have $U_1(C_1^A) \geq W_1(C_1^A) = U_1(C_1^B)$, which means that the choice r_1^A is also optimal. Q.E.D.

Using Lemmas 1 through 3, we now prove the theorem.

Proof of part (i): $C_1'R'C_1''$ iff $W_1(C_1') \geq U_1(C_1'')$

First let's suppose that $C_1'R'C_1''$. Imagine that, contrary to the theorem, $W_1(C_1') < U_1(C_1'')$. Then, according to Lemma 2, there is some G for which C_1'' is the only equilibrium outcome even though C' is available. That implies $\sim C_1'R'C_1''$, a contradiction.

Next suppose that $W_1(C_1') \geq U_1(C_1'')$. If the inequality is strict, then according to Lemma 1, C_1'' is never an equilibrium outcome when C_1' is available, so $C_1'RC_1''$. If $W_1(C_1') = U_1(C_1'')$, then according to Lemma 3, C_1' is always an equilibrium outcome when C_1'' is an equilibrium outcome and both are available, so again $C_1'RC_1''$.

Proof of part (ii): $C_1'P^*C_1''$ iff $W_1(C_1') > U_1(C_1'')$

First let's suppose that $C_1'P^*C_1''$. Imagine that, contrary to the theorem, $W_1(C_1') \leq U_1(C_1'')$. Then, according to Lemma 2, there is some G for which C_1'' is an equilibrium outcome even though C_1' is available. That implies $\sim C_1'P^*C_1''$, a contradiction.

Next suppose that $W_1(C_1') > U_1(C_1'')$. Then according to Lemma 1, C_1'' is never an equilibrium outcome when C_1' is available, so $C_1'P^*C_1''$.

Proof of part (iii): $P' = R'$. We demonstrate this part of the theorem by showing that $C_1'R'C_1''$ implies $C_1'P'C_1''$ (the opposite implication is immediate). From part (i), we know that $W_1(C_1') \geq U_1(C_1'')$. It follows that

$$U_1(C_1') \geq W_1(C_1') \geq U_1(C_1'') \geq W_1(C_1'') \tag{11}$$

If the second inequality in (11) is strict, we have $U_1(C_1') > W_1(C_1'')$, which implies $\sim C_1''R'C_1'$ by part (i). If the second inequality in (11) is not strict, then (with $C_1' \neq C_1''$), there must be some $k > 1$ for which either $c_k' > 0$ or $c_k'' > 0$, which means either $U_1(C_1') > W_1(C_1')$ or

$U_1(C_1'') > W_1(C_1'')$. In either case, expression (11) implies $U_1(C_1') > W_1(C_1'')$, which in turn implies $\sim C_1''R'C_1'$ by part (i). Therefore, $C_1'R'C_1''$ implies $\sim C_1''R'C_1'$, and thus $C_1'P'C_1''$.

Proof of part (iv): $C_1'R^*C_1''$ iff $U_1(C_1') \geq W_1(C_1'')$.

First suppose that $C_1'R^*C_1''$. Then by definition, there exists some G for which C_1' is an equilibrium outcome even though C_1'' is available. But then Lemma 1 implies $U_1(C_1') \geq W_1(C_1'')$.

Next suppose that $U_1(C_1') \geq W_1(C_1'')$. By Lemma 2, there exists some G for which C_1' is an equilibrium outcome even though C_1'' is available. But then, by definition, $C_1'R^*C_1''$.

Proof of part (v): R', P', and P^* are transitive.

First consider R' (and hence P'). Suppose that $C_1^1R'C_1^2R'C_1^3$. From part (i), we know that $W_1(C_1^1) \geq U_1(C_1^2)$ and $W_1(C_1^2) \geq U_1(C_1^3)$. Using the fact that $U_1(C_1^2) \geq W_1(C_1^2)$, we therefore have $W_1(C_1^1) \geq U_1(C_1^3)$, which implies $C_1^1R'C_1^3$.

Next consider P^*. Suppose that $C_1^1P^*C_1^2P^*C_1^3$. From part (ii), we know that $W_1(C_1^1) > U_1(C_1^2)$ and $W_1(C_1^2) > U_1(C_1^3)$. Using the fact that $U_1(C_1^2) \geq W_1(C_1^2)$, we therefore have $W_1(C_1^1) > U_1(C_1^3)$, which implies $C_1^1P^*C_1^3$. Q.E.D.

Proof of Theorem 11: First suppose that C_1^* solves $\max_{C_1 \in X_1} U_1(C_1)$. Consider $G \in \mathcal{G}_1$ such that the individual chooses the entire consumption trajectory from X_1 at $t = 1$. For that G, we have $C(G) = \{C^*\}$ (uniqueness of the choice follows from strict concavity of u). It follows that $\sim C_1P'C_1^*$ for all $C_1 \in X_1$. Accordingly, C_1^* is a strict individual welfare optimum (and hence a weak individual welfare optimum) in X_1.

Now consider any $\widehat{C}_1 \in X_1$ that does not solve $\max_{C_1 \in X_1} U_1(C_1)$. There must be some $C_1' \in X_1$ with $U_1(C_1') > U_1(\widehat{C}_1)$. But then there must also be some $C_1'' \in X_1$ with $U_1(C_1'') > U_1(\widehat{C}_1)$ and $c_1'' \neq c_1^*$. (We can construct C_1'' as follows. If $c_1' > 0$, simply reduce c_1' slightly. If $c_1' = 0$, simply increase c_1' by some small $\varepsilon > 0$ and reduce c_t' in some future period t by $\lambda^{t-1}\varepsilon$.) Now consider any G that contains the options $\widehat{C}_1$ and C_1''. Notice $G \in \mathcal{G}_1$; we cannot have $G \in \mathcal{G}_t$ for any $t > 1$, because a choice from G resolves some discretion at time

$t = 1$. But since $U_1(C_1'') > U_1(\widehat{C}_1)$ and $G \in \mathcal{G}_1$, the individual will not select $\widehat{C}_1$ from G. Thus, $C_1'' P^* \widehat{C}_1$. It follows that $\widehat{C}_1$ is not a weak individual welfare optimum (and hence not a strict individual welfare optimum).

Now fix $(c_1', ..., c_{t-1}')$ and suppose that C_1^* (with $c_k^* = c_k'$ for $k < t$) maximizes $\alpha U_t(C_t) + (1-\alpha)V_t(C_t)$ in $X_t(c_1', ..., c_{t-1}')$ for some $\alpha \in [0,1]$. For any other $C_1 \in X_t(c_1', ..., c_{t-1}')$, either (i) $U_t(C_t^*) > U(C_t)$, or (ii) $V_t(C_t^*) > V(C_t)$. In case (i), consider $G \in \mathcal{G}_t$ such that the individual chooses between C_1^* and C_1 (and nothing else) at time t. Since he will select C_1^* and not C_1, we have $\sim C_1 P' C_1^*$. In case (ii), consider $G \in \mathcal{G}_k$ for any $k < t$ such that the individual chooses between C_1^* and C_1 (and nothing else) at time k. Since he will select C_1^* and not C_1, we have $\sim C_1 P' C_1^*$. Accordingly, C_1^* is a strict individual welfare optimum (and hence a weak individual welfare optimum) in $X_t(c_1', ..., c_{t-1}')$.

Now consider any $\widehat{C}_1 \in X_1$ that does not maximize $\alpha U_t(C_t) + (1-\alpha)V_t(C_t)$ in $X_t(c_1', ..., c_{t-1}')$ for any $\alpha \in [0,1]$. Because u is strictly concave, the efficient frontier of the set $(U_t(C_t), V_t(C_t))$ for $C_1 \in X_t(c_1', ..., c_{t-1}')$ is strictly concave. All points on the frontier of that set maximize $\alpha U_t(C_t) + (1-\alpha)V_t(C_t)$ for some $\alpha \in [0,1]$. It follows that $\left(U_t(\widehat{C}_t), V_t(\widehat{C}_t)\right)$ cannot lie on the frontier of that set. Accordingly, there must be some $C_1' \in X_t(c_1', ..., c_{t-1}')$ with $U_t(C_t') > U_t(\widehat{C}_t)$ and $V_t(C_t') > V_t(\widehat{C}_t)$. Given the existence of C_1', there must also be some $C_1'' \in X_t(c_1', ..., c_{t-1}')$ with $U_t(C_t'') > U_t(\widehat{C}_t)$, $V_t(C_t'') > V_t(\widehat{C}_t)$, and $c_t'' \neq c_t^*$. (We can construct C_1'' as follows. If $c_t' > 0$, simply reduce c_t' slightly. If $c_t' = 0$, simply increase c_t' by some small $\varepsilon > 0$ and reduce c_k' in some future period $k > t$ by $\lambda^{k-t}\varepsilon$.) Note that $V_t(C_t'') > V_t(\widehat{C}_t)$ implies $U_n(C_n'') > U_n(\widehat{C}_n)$ for all $n < t$.

Now consider any G that contains the options $\widehat{C}_1$ and C_1''. Notice $G \in \mathcal{G}_n$ for $n \leq t$; we cannot have $G \in \mathcal{G}_n$ for any $n > t$, because a choice from G resolves some discretion at time t. But since $U_n(C_n'') > U_n(\widehat{C}_n)$ for all $n \leq t$, the individual will not select $\widehat{C}_1$ when C_1^* is available from any $G \in \mathcal{G}_n$. Thus, $C_1'' P^* \widehat{C}_1$. It follows that $\widehat{C}$ is not a weak individual welfare optimum (and hence not a strict individual welfare optimum) in $X_t(c_1', ..., c_{t-1}')$. Q.E.D.

Proof of Theorem 12: First note that if C_1^* maximizes $U_1(C_1)$, then it is a strict (and hence a weak) robust multi-self Pareto optimum. This conclusion follows from the fact that $U_1(C_1) < U_1(C_1^*)$ for any feasible $C_1 \neq C_1^*$; regardless of how other selves are affected by a switch from C_1^* to C_1, the time $t = 1$ self is strictly worse off.

Next we argue that $\widehat{C}_1 \neq C_1^*$ is not a weak robust multi-self Pareto optimum (and therefore not a strict robust multi-self Pareto optimum either). We divide the possibilities into the following three cases.

(i) $\widehat{c}_1 < c_1^*$. In that case, if each Γ_t is sufficiently sensitive to c_1, we have $\widehat{U}_t(C_1^*) > \widehat{U}_t(\widehat{C}_1)$ for $t = 2, .., T$. Since we also know that $U_1(C_1^*) > U_1(\widehat{C}_1)$, $\widehat{C}_1$ is not a weak robust multi-self Pareto optimum.

(ii) $\widehat{c}_1 = c_1^*$. Note that there must be some $t > 0$ such that $c_t^* > 0$ (or we would not have $U_1(C_1^*) > U_1(\widehat{C}_1)$). Define C_1' as follows: $c_1' = c_1^* + \varepsilon$, $c_t' = c_t^* - \varepsilon\lambda^{t-1}$, and $c_k' = c_k^*$ for $k \neq 1, t$. For $\varepsilon > 0$ sufficiently small, we have $U_1(C_1') > U_1(\widehat{C}_1)$. If each Γ_t is sufficiently sensitive to c_1, we will also have $\widehat{U}_t(C_1') > \widehat{U}_t(\widehat{C}_1)$ for $t = 2, .., T$, which implies $\widehat{C}_1$ is not a weak robust multi-self Pareto optimum.

(iii) $\widehat{c}_1 > c_1^*$. In that case, there exists $t > 1$ for which $\widehat{c}_t < c_t^*$. Let

$$\Delta c_1 = \min\left\{\widehat{c}_1 - c_1^*, \lambda^{t-1}\left(c_t^* - \widehat{c}_t\right)\right\} > 0,$$

and let

$$\Delta c_t = \lambda^{t-1}\Delta c_1 > 0.$$

Note that

$$\widehat{c}_1 - \Delta c_1 \geq c_1^* \tag{12}$$

and

$$\widehat{c}_t \leq c_t^* - \Delta c_t \tag{13}$$

Define C_1' as follows: $c_1' = c_1^* + \Delta c_1 > c_1^*$, $c_t' = c_t^* - \Delta c_t < c_t^*$, and $c_k' = c_k^*$ for $k \neq 1, T$. Define C_1'' as follows: $c_1'' = \widehat{c}_1 - \Delta c_1 < \widehat{c}_1$, $c_t'' = \widehat{c}_t + \Delta c_t > \widehat{c}_t$, and $c_k'' = c_k^*$ for $k \neq 1, T$. (It is easy to check that C_1', $C_1'' \in X_1$.)

We now show that $U_1(C_1^{''}) > U_1(\widehat{C}_1)$. We know that $U_1(C_1^*) > U_1(C_1')$; therefore,

$$u(c_1^* + \Delta c_1) - u(c_1^*) < \beta\delta^{t-1}\left[u\left(c_t^*\right) - u(c_t^* - \Delta c_t)\right] \tag{14}$$

From (12) and the concavity of u, we know that

$$u(\widehat{c}_1) - u(\widehat{c}_1 - \Delta c_1) < u(c_1^* + \Delta c_1) - u(c_1^*) \tag{15}$$

Similarly, from (13) and the concavity of u, we know that

$$u\left(c_t^*\right) - u(c_t^* - \Delta c_t) < u\left(\widehat{c}_t + \Delta c_t\right) - u(\widehat{c}_t) \tag{16}$$

Combining inequalities (14), (15), and (16), we obtain:

$$u(\widehat{c}_1) - u(\widehat{c}_1 - \Delta c_1) < \beta\delta^{t-1}\left[u\left(\widehat{c}_t + \Delta c_t\right) - u(\widehat{c}_t)\right].$$

But that implies $U_1(C_1^{''}) > U_1(\widehat{C}_1)$, as desired.

Now define C_1^0 as follows: $c_1^0 = c_1^{''} - \varepsilon$, $c_T^0 = c_T^{''} + \varepsilon\lambda^{T-1}$, and $c_k^0 = c_k^{''}$ for $k \neq 1, T$. For $\varepsilon > 0$ sufficiently small, we have $U_1(C_1^0) > U_1(\widehat{C}_1)$. For $\Gamma_t(c_1, ..., c_{t-1}) \equiv 0$, we also have $\widehat{U}_t(C_1') > \widehat{U}_t(\widehat{C}_1)$ for $t = 2, .., T$, which implies $\widehat{C}_1$ is not a weak robust multi-self Pareto optimum. Q.E.D.

2. Proofs of convergence results

Our analysis will require us to say when one set is close to another. For any compact set A, let $N_r(A)$ denote the neighborhood of A or radius r (defined as the set $\cup_{x\in A} B_r(x)$, where $B_r(x)$ is the open ball of radius r centered at x). For any two compact sets A and B, let

$$\delta_U(A, B) = \inf\left\{r > 0 \mid B \subset N_r(A)\right\}$$

δ_U is the upper Hausdorff hemimetric. This metric can also be applied to sets that are not compact (by substituting the closure of the sets).

Consider a sequence of choice functions C^n defined on $\mathcal{G}$. Also consider a choice function $\widehat{C}$ defined on $\mathcal{X}^c$, the compact elements of $\mathcal{X}$, that reflects maximization of a continuous utility function, u. We will say that C^n weakly converges to $\widehat{C}$ if, for all $\varepsilon > 0$, there exists N such that for all $n > N$ and $(X, d) \in \mathcal{G}$, we have $\delta_U\left(\widehat{C}(\text{clos}(X)), C^n(X, d)\right) < \varepsilon$.

In addition to $U^n(x)$, $L^n(x)$, $\widehat{U}^*(u)$, and $\widehat{L}^*(u)$ (defined in the text), we also define $\widehat{U}(x) \equiv \{y \in X \mid u(y) > u(x)\}$ and $\widehat{L}(x) \equiv \{y \in X \mid u(y) < u(x)\}$.

We begin our proofs of the convergence results with a lemma.

Lemma 4: *Suppose that C^n weakly converges to $\widehat{C}$, where $\widehat{C}$ is defined on X^c and reflects maximization of a continuous utility function, u. Consider any values u_1 and u_2 with $u_1 > u_2$. Then there exists N' such that for $n > N'$, we have $yP^{n*}x$ for all $y \in \widehat{U}^*(u_1)$ and $x \in \widehat{L}^*(u_2)$.*

Proof: Since u is continuous, there exists $r' > 0$ such that $N_{r'}\left(\widehat{U}^*(u_1)\right)$ does not contain any point in $\widehat{L}^*(u_2)$. Moreover, since C^n weakly converges to $\widehat{C}$, there exists some N' such that for $n > N'$ and $(X, d) \in \mathcal{G}$, we have $\delta_U\left(\widehat{C}(\text{clos}(X)), C^n(X, d)\right) < r'$.

Now we show that if $n > N'$, then for all generalized choice sets that include at least one element of $\widehat{U}^*(u_1)$, no element of $\widehat{L}^*(u_2)$ is chosen. Consider any set X_1 containing containing at least one element of $\widehat{U}^*(u_1)$. We know that $\widehat{C}(\text{clos}(X_1)) \subseteq \widehat{U}^*(u_1)$, from which it follows that $N_{r'}\left(\widehat{C}(\text{clos}(X_1))\right)$ does not contain any element of $\widehat{L}^*(u_2)$. But then, for $n > N'$, there is no d with $(X_1, d) \in \mathcal{G}$ for which $C^n(X_1, d)$ contains any element of $\widehat{L}^*(u_2)$.

Since we have assumed that $\{a, b\} \in \mathcal{X}$ for all $a, b \in X$, it follows immediately that $yP^{n*}x$ for all $y \in \widehat{U}^*(u_1)$ and $x \in \widehat{L}^*(u_2)$. Q.E.D.

Proof of Theorem 5: The proof proceeds in three steps. For each, we fix a value of $\varepsilon > 0$.

Step 1: Suppose that C^n weakly converges to $\widehat{C}$. Then for n sufficiently large, $\widehat{L}^*(u(x^0) - \varepsilon) \subseteq L^n(x^0)$.

Let $u_1 = u(x^0)$ and $u_2 = u(x^0) - \varepsilon$. By Lemma 4, there exists N' such that for $n > N'$, we have $yP^{n*}x$ for all $y \in \widehat{U}^*(u_1)$ and $x \in \widehat{L}^*(u_2)$. Taking $y = x^0$, for $n > N'$ we have $x^0 P^{n*} x$ (and therefore $x \in L^n(x^0)$) for all $x \in \widehat{L}^*(u_2)$

Step 2: Suppose that C^n weakly converges to $\widehat{C}$. Then for n sufficiently large, $\widehat{U}(u(x^0)+\varepsilon) \subseteq U^n(x^0)$.

Let $u_1 = u(x^0)+\varepsilon$ and $u_2 = u(x^0)$. By Lemma 4, there exists N'' such that for $n > N''$, we have $yP^{n*}x$ for all $y \in \widehat{U}^*(u_1)$ and $x \in \widehat{L}^*(u_2)$. Taking $x = x^0$, for $n > N''$ we have $yP^{n*}x^0$ (and therefore $y \in U^n(x^0)$) for all $x \in \widehat{U}^*(u_1)$. Q.E.D.

In the statement of Theorem 6, we interpret d_1 is a function of the compensation level, m, rather than a scalar. With that interpretation, the theorem subsumes cases in which $\mathcal{G}$ is not rectangular.

Proof of Theorem 6: It is easy to verify that our notions of CV-A and CV-B for $\widehat{C}$ coincide with the standard notion of compensating variation. That is, $\widehat{m}_A = \widehat{m}_B = \widehat{m}$; the infinum (supremum) of the payment that leads the individual to choose something better than (worse than) the object chosen from the initial opportunity set equals the payment that exactly compensates for the change. Therefore, our task is to show that $\lim_{n\to\infty} m_A^n = \widehat{m}_A$, and $\lim_{n\to\infty} m_B^n = \widehat{m}_B$. We will provide the proof for $\lim_{n\to\infty} m_A^n = \widehat{m}_A$; the proof for $\lim_{n\to\infty} m_B^n = \widehat{m}_B$ is completely analogous.

Step 1: Consider any m such that $y\widehat{P}^*x$ for all $x \in \widehat{C}(X(\alpha_0, 0))$ and $y \in \widehat{C}(X(\alpha_1, m))$. We claim that there exists N_1 such that for $n > N_1$, we have $yP^{n*}x$ for all $x \in C^n(X(\alpha_0, 0), d_0)$ and $y \in C^n(X(\alpha_1, m), d_1(m))$.

Define $u_1 = \frac{1}{3}u(w) + \frac{2}{3}u(z)$ and $u_2 = \frac{2}{3}u(w) + \frac{1}{3}u(z)$ for $w \in \widehat{C}(X(\alpha_0, 0))$ and $z \in \widehat{C}(X(\alpha_1, m))$. Since $u_1 > u_2$, Lemma 1 implies there exists N_1' such that for $n > N_1'$, we have $yP^{n*}x$ for all $y \in \widehat{U}^*(u_1)$ and $x \in \widehat{L}^*(u_2)$.

Next, notice that since u is continuous, there exists $r_1 > 0$ such that $N_{r_1}\left(\widehat{C}(X(\alpha_0, 0))\right) \subset \widehat{L}^*(u_2)$, and $N_{r_1}\left(\widehat{C}(X(\alpha_1, m))\right) \subset \widehat{U}^*(u_1)$. Moroever, there exists N_1'' such that for $n > N_1''$, we have $C^n(X(\alpha_0, 0), d_0) \subset N_{r_1}\left(\widehat{C}(X(\alpha_0, 0))\right)$ and $C^n(X(\alpha_1, m), d_1(m)) \subset N_{r_1}\left(\widehat{C}(X(\alpha_1, m))\right)$. Consequently, for $n > N_1''$, we have $C^n(X(\alpha_0, 0), d_0) \subset \widehat{L}^*(u_2)$ and $C^n(X(\alpha_1, m), d_1(m)) \subset \widehat{U}^*(u_1)$. It follows that, for $n > N_1 = \max\{N_1', N_1''\}$, we have $yP^{n*}x$ for all $x \in C^n(X(\alpha_0, 0), d_0)$ and $y \in C^n(X(\alpha_1, m), d_1(m))$.

Step 2: Consider any m such that $y\widehat{P}^*x$ for all $y \in \widehat{C}(X(\alpha_0, 0))$ and $x \in \widehat{C}(X(\alpha_1, m))$. We claim that there exists N_2 such that for $n > N_2$, we have $yP^{n*}x$ for all $y \in C^n(X(\alpha_0, 0), d_0)$ and $x \in C^n(X(\alpha_1, m), d_1(m))$.

Define $u_1 = \frac{1}{3}u(w) + \frac{2}{3}u(z)$ and $u_2 = \frac{2}{3}u(w) + \frac{1}{3}u(z)$ for $z \in \widehat{C}(X(\alpha_0, 0))$ and $w \in \widehat{C}(X(\alpha_1, m))$. Since $u_1 > u_2$, Lemma 1 implies there exists N_2' such that for $n > N_2'$, we have $yP^{n*}x$ for all $y \in \widehat{U}^*(u_1)$ and $x \in \widehat{L}^*(u_2)$.

Next, notice that since u is continuous, there exists $r_2 > 0$ such that $N_{r_2}\left(\widehat{C}(X(\alpha_0, 0))\right) \subset \widehat{U}^*(u_1)$, and $N_{r_2}\left(\widehat{C}(X(\alpha_1, m))\right) \subset \widehat{L}^*(u_2)$. Moroever, there exists N_2'' such that for $n > N_2''$, we have $C^n(X(\alpha_0, 0), d_0) \subset N_{r_2}\left(\widehat{C}(X(\alpha_0, 0))\right)$ and $C^n(X(\alpha_1, m), d_1(m)) \subset N_{r_2}\left(\widehat{C}(X(\alpha_1, m))\right)$. Consequently, $C^n(X(\alpha_0, 0), d_0) \subset \widehat{U}^*(u_1)$ and $C^n(X(\alpha_1, m), d_1(m)) \subset \widehat{L}^*(u_2)$. It follows that, for $n > N_2 = \max\{N_2', N_2''\}$, we have $yP^{n*}x$ for all $x \in C^n(X(\alpha_1, m), d_1(m))$ and $y \in C^n(X(\alpha_0, 0), d_0)$.

Step 3: $\lim_{n\to\infty} m_A^n = \widehat{m}_A$.

Suppose not. Then the sequence m_A^n has a limit point $m_A^* \neq \widehat{m}_A$. Suppose first that $m_A^* > \widehat{m}_A$. Consider $m' = (m_A^* - \widehat{m}_A)/2$. Since u satisfies non-satiation and $m' > \widehat{m}_A$, we know by step 1 that there exists N_1 such that for $n > N_1$, we have $yP^{n*}x$ for all $x \in C^n(X(\alpha_0, 0), d_0)$ and $y \in C^n(X(\alpha_1, m'), d_1(m'))$. This in turn implies that $m_A^n \leq m' < m_A^*$ for all $n > N_1$, which contradicts the supposition that m_A^* is a limit point of m_A^n. The case of $m_A^* < \widehat{m}_A$ is similar except that we rely on step 2 instead of step 1. Q.E.D.

Proof of Theorem 9: Suppose not. Without loss of generality, assume that x^n converges to a point $x^* \notin W(\mathrm{clos}(X), \widehat{C}_1, ..., \widehat{C}_J, \mathcal{X}^c)$ (if necesary, take a convergent subsequence of the original sequence). Then there must be some $x^0 \in X$, some $\varepsilon > 0$, and some N' such that, for all $n > N'$, we have $x^n \in \widehat{L}_i^*(u(x^0) - \varepsilon)$ for all i. By Theorem 4, there exists N'' such that for $n > N''$, we have $\widehat{L}_i^*(u(x^0) - \varepsilon) \subseteq L_i^n(x^0)$ for all i. Hence, for all $n > \max\{N', N''\}$, we have $x^n \in L_i^n(x^0)$ for all i. But in that case, $x^n \notin W(X; C_1^n, ..., C_J^n, \mathcal{G})$, a contradiction. Q.E.D.

3. An alternative definition of compensating variation

Without futher structure, we cannot rule out the existence of compensation levels smaller than the CV-A for which everything selected in the new set is unambiguously chosen over everything selected from the initial set. Nor can we rule out compensation levels larger than the CV-B for which everything selected form the initial set is unambiguously chosen over everything selected from the new set. This observation suggests the following alternative definitions of compensating variation:.

Definition: CV-A$'$ is the level of compensation $m^{A'}$ that solves

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

Definition: CV-B$'$ is the level of compensation $m^{B'}$ that solves

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

In principle, the CV-A$'$ could be smaller than the CV-A (but not larger), and the CV-B$'$ could be larger than the CV-B (but not smaller). It is straightforward to demonstrate the equivalence of CV-A and CV-A$'$ under the following monotonicity assumption: If, for some $y \in X$, α, d, and m, we have $y \notin C(X, d')$ for all $(X, d') \in \mathcal{G}$ containing at least one alternative in $C(X(\alpha, m), d)$, then for all $m' > m$ we also have $y \notin C(X, d')$ for all $(X, d') \in \mathcal{G}$ containing at least one alternative in $C(X(\alpha, m'), d)$. A complementary assumption guarantees the equivalence of CV-B and CV-B$'$.

When the monotonicity assumption does not hold, the CV-A$'$ can be either larger or smaller than the CV-B$'$. Thus, unlike the CV-A and the CV-B, the CV-A$'$ and the CV-B$'$ cannot always be interpreted, respectively, as upper and lower bounds on required compensation.

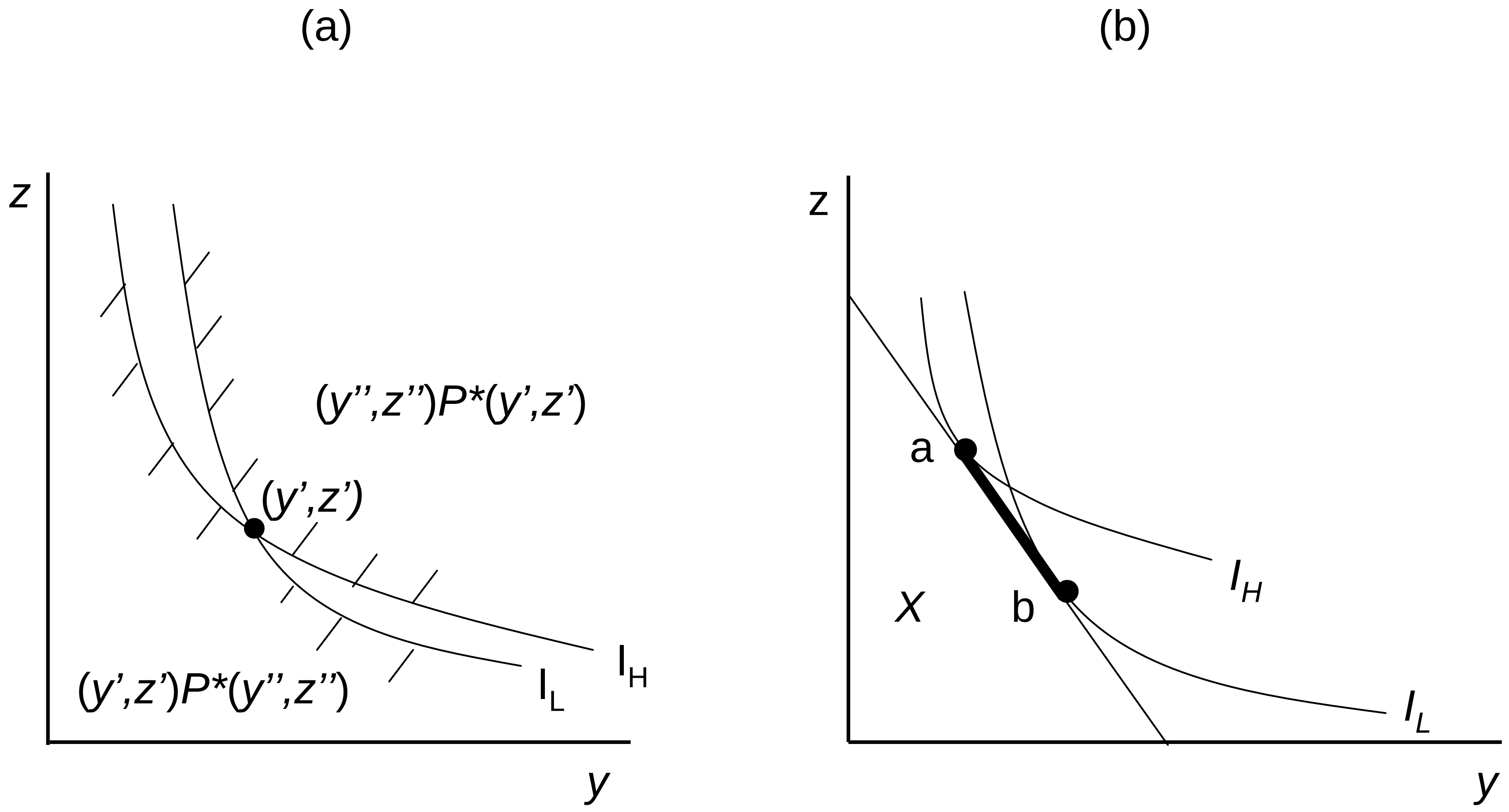

Figure 1: Coherent arbitrariness

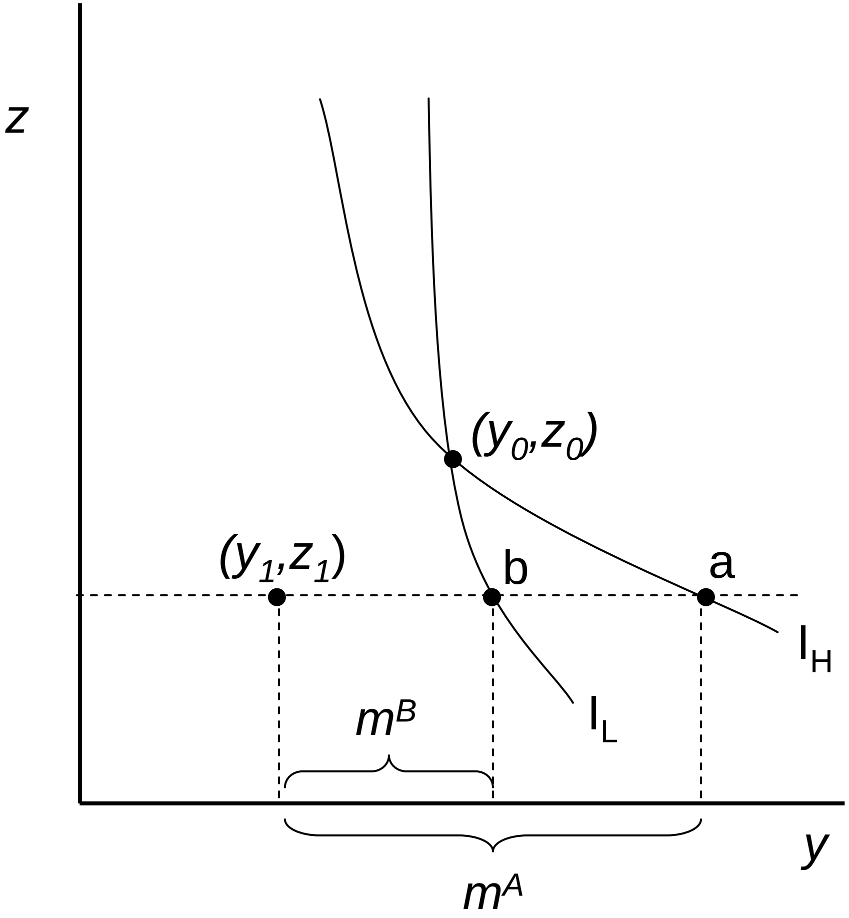

Figure 2: CV-A and CV-B for Example 7

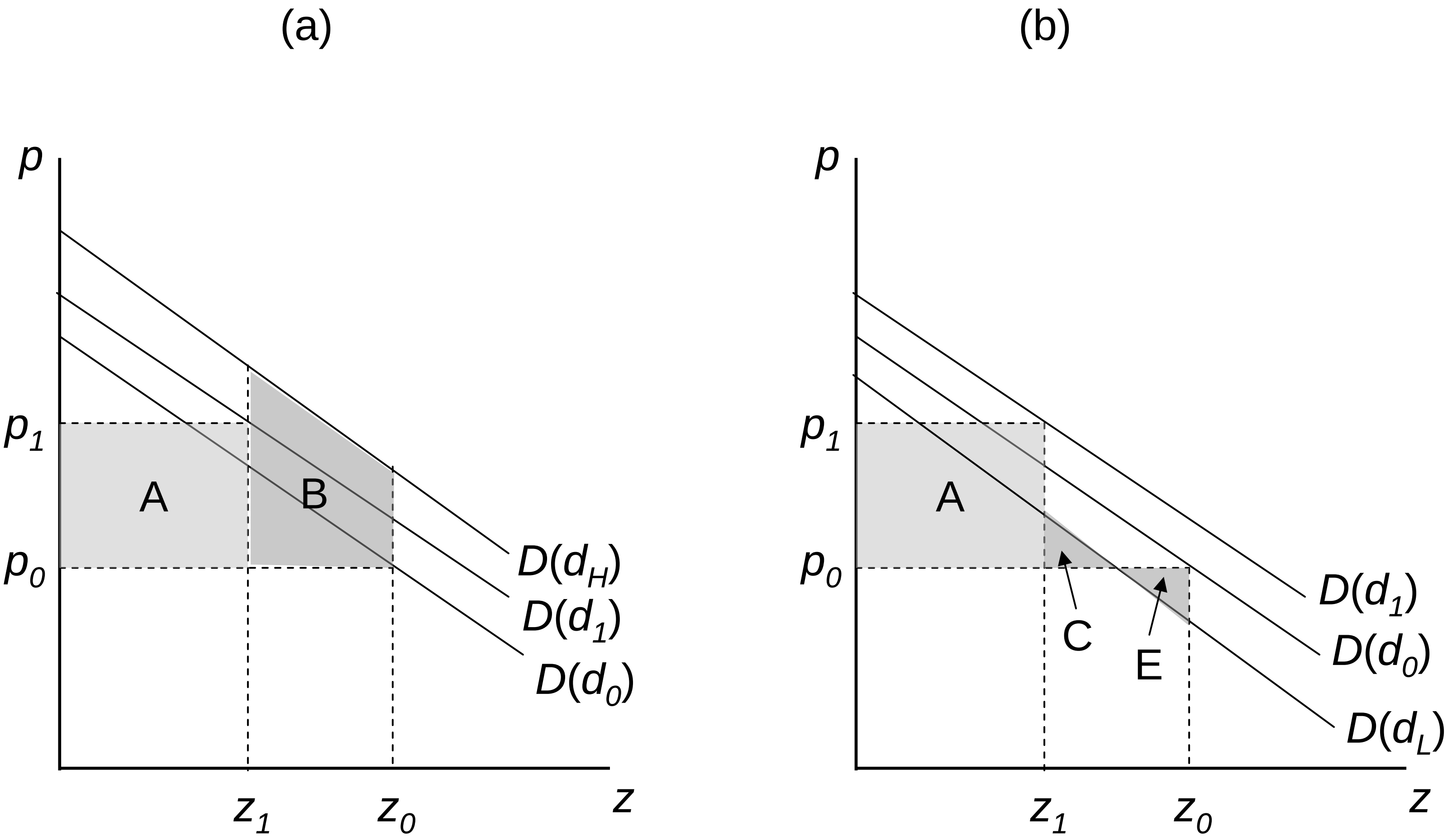

Figure 3: CV-A and CV-B for a price change

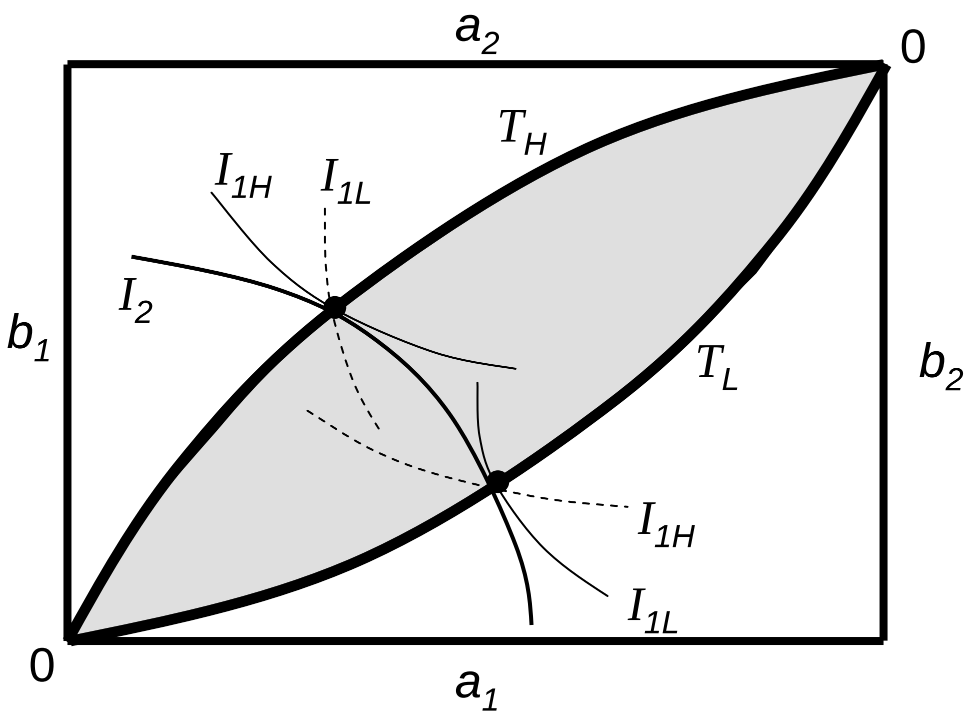

Figure 4: The generalized contact curve

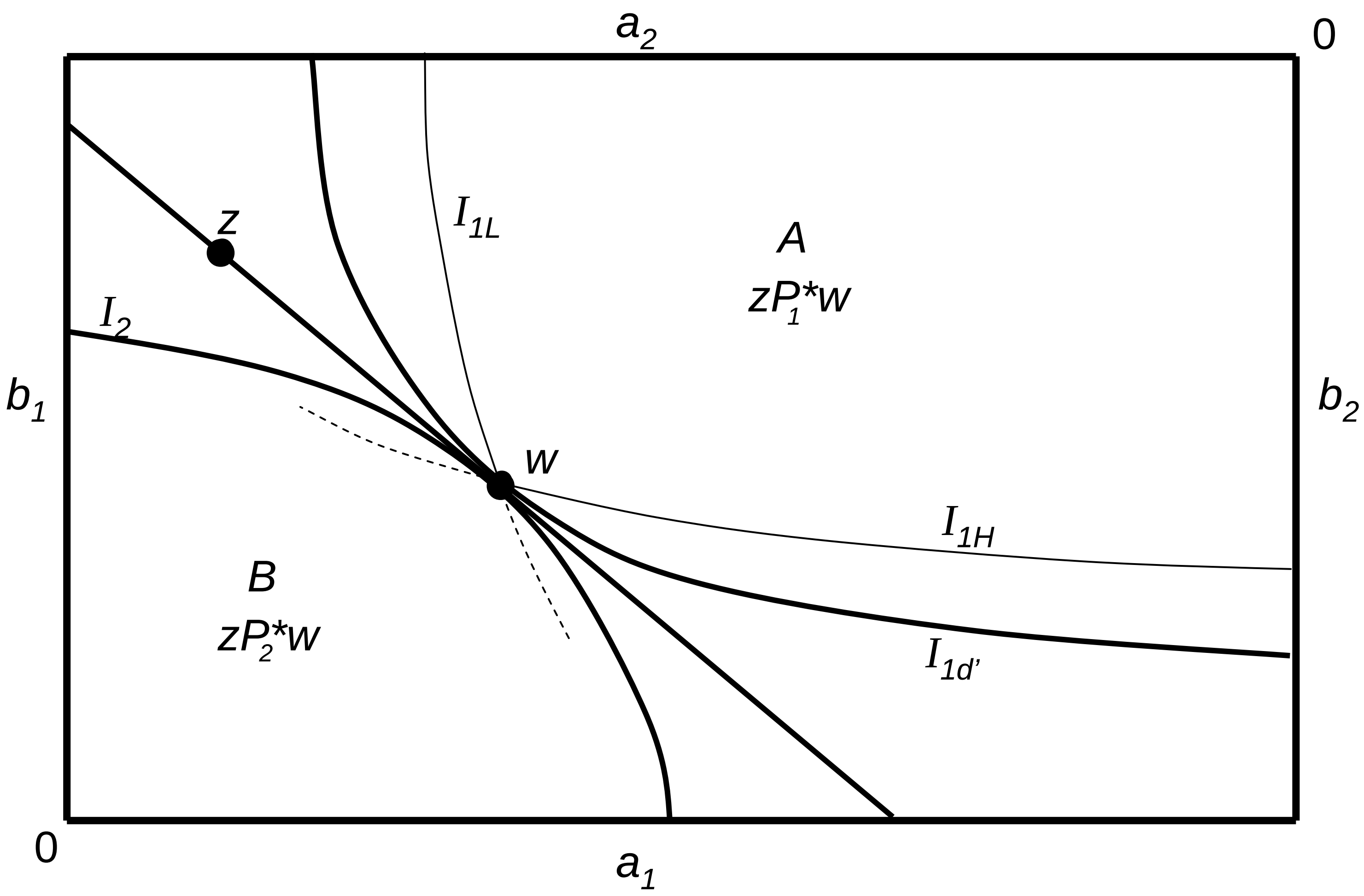

Figure 5: The efficiency of a behavioral competitive equilibrium in a simple economy

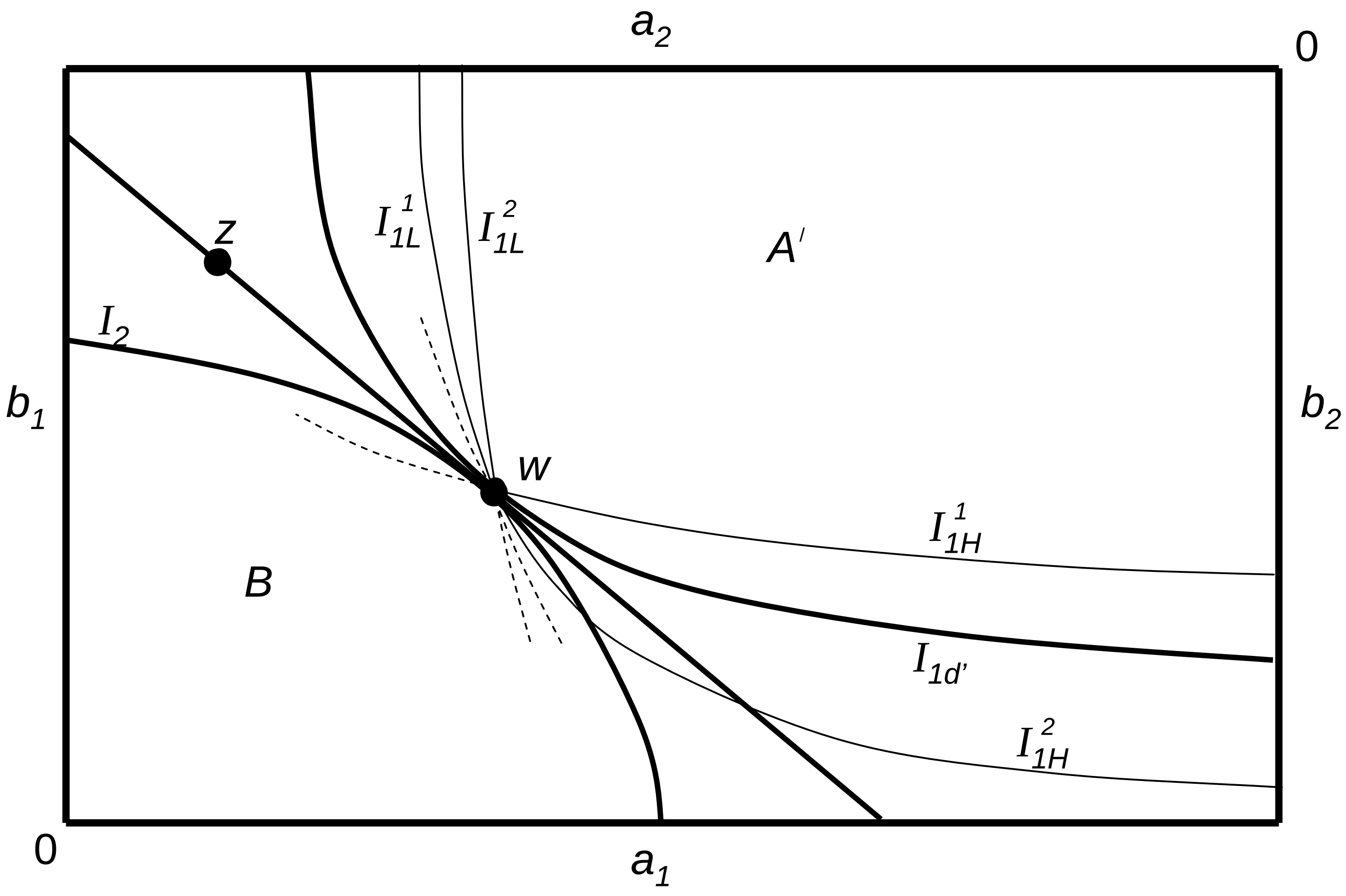

Figure 6: Market failure in a simple economy