BEYOND REVEALED PREFERENCE: CHOICE-THEORETIC FOUNDATIONS FOR BEHAVIORAL WELFARE ECONOMICS*

B. DOUGLAS BERNHEIM AND ANTONIO RANGEL

We propose a broad generalization of standard choice-theoretic welfare economics that encompasses a wide variety of nonstandard behavioral models. Our approach exploits the coherent aspects of choice that those positive models typically attempt to capture. It replaces the standard revealed preference relation with an *unambiguous choice* relation: roughly, x is (strictly) unambiguously chosen over y (written xP^*y) iff y is never chosen when x is available. Under weak assumptions, P^* is acyclic and therefore suitable for welfare analysis; it is also the most discerning welfare criterion that never overrules choice. The resulting framework generates natural counterparts for the standard tools of applied welfare economics and is easily applied in the context of specific behavioral theories, with novel implications. Though not universally discerning, it lends itself to principled refinements.

I. INTRODUCTION

Interest in behavioral economics has grown in recent years, stimulated by accumulating evidence that the standard model of consumer decision-making may provide an inadequate positive description of behavior. Behavioral models are increasingly finding their way into policy evaluation, which inevitably raises questions concerning welfare. Unfortunately, there is as yet no consensus concerning the general principles that should govern such normative inquiries. In many cases, economists adopt *ad hoc* criteria for particular positive models, offering justifications based on loose and inevitably controversial intuition. The tight connection between choice and welfare that has governed normative economic analysis for more than half a century is also typically

*We would like to thank Colin Camerer, Andrew Caplin, Vincent Crawford, Robert Hall, Peter Hammond, Botond Koszegi, Preston McAfee, Paul Milgrom, three anonymous referees, and seminar participants at Stanford University, U. C. Berkeley, Princeton University, Rutgers University, University of Chicago GSB, M.I.T., Harvard University, the 2006 and 2008 NYU Methodologies Conferences, 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, the Summer 2007 PET Meetings, and the Fall 2007 NBER Public Economics Meetings, for useful comments. We are also indebted to Xiaochen Fan and Eduardo Perez for able research assitance. Bernheim gratefully acknowledges financial support from the NSF (SES-0452300 and SES-0752854). Rangel gratefully acknowledges financial support from the NSF (SES-0134618) and the Moore Foundation.

The Quarterly Journal of Economics, February 2009

severed. Indeed, many behavioral economists distinguish between *decision utility*, which rationalizes choice, and *true utility*, which encapsulates well-being. That distinction compels them either to make paternalistic judgments, or to adopt some alternative measure of experienced well-being. Despite attempts to define and measure true utility (e.g., Kahneman, Wakker, and Sarin [1997]; Kahneman [1999]), there are concerns regarding the feasibility of that approach, and many economists remain understandably hesitant to adopt normative principles that are not rooted in choice.[1]

Our objective is to generalize standard choice-based welfare analysis to settings with nonstandard decision makers.[2] How one approaches that task depends on one's interpretation of the standard framework. According to one interpretation, standard normative analysis respects the decision maker's true objectives, which her choices reveal. Unfortunately, it is often difficult to formulate coherent and normatively compelling rationalizations for nonstandard choice patterns.[3] As a result, discussions of welfare become mired in controversy, leading some to reject choice as a foundation for normative analysis (Sugden 2004).

According to a second interpretation of standard welfare analysis, welfare is *defined* in terms of choice rather than underlying objectives. The statement "x is strictly revealed preferred to y" (xPy) simply means that x (and not y) is chosen from the set $\{x, y\}$. Thus, one can determine whether xPy directly from choice patterns without relying on any underlying rationalization. Furthermore, one does not require a rationalization to justify normative judgments; arguably, choices provide appropriate guidance because they are choices, not because they reflect something else. From that perspective, preferences and utility are useful *positive* tools—clever analytic constructs that allow economists to systematize knowledge concerning behavior and predict choices where direct observations are absent—but they play no direct role in normative analysis. In particular, because a preference representation simply recapitulates the choice correspondence, it cannot resolve normative puzzles arising from nonstandard behavior, and any appearance to the contrary is deceiving.

1. The justifications for building a welfare framework around choice are familiar; see Bernheim (2009a, b) for a recent discussion.

2. A preliminary summary of this work appeared in Bernheim and Rangel (2007); see also Bernheim and Rangel (2008b) and Bernheim (2009b).

3. See, for example, Koszegi and Rabin (2008a), who argue that choices alone cannot identify preferences, or Bernheim (2009b).

We adopt the second interpretation of standard welfare analysis, and develop a generalized welfare criterion that respects choice directly, without reference to the decision maker's underlying objectives.[4] We thereby avoid the thorny problems associated with formulating and justifying rationalizations.[5] Naturally, operationalizing the principle of respect for choice is problematic when choices conflict. However, useful behavioral theories do not imply that choice conflicts are ubiquitous. On the contrary, those theories are generally motivated by the observation that choice patterns exhibit a substantial degree of underlying coherence. We take that observation as our central premise, and devise welfare criteria that respect the coherent aspects of choice. Specifically, we propose replacing the standard revealed preference relation with an *unambiguous choice* relation: roughly, x is (strictly) unambiguously chosen over y (written xP^*y) iff y is never chosen when x is available. That criterion instructs us to respect choice whenever it provides clear normative guidance, and to live with whatever ambiguity remains. Though P^* need not be transitive, it is always acyclic, and therefore suitable for rigorous welfare analysis. Moreover, among welfare criteria that never overrule choice by deeming an object improvable within a set from which it is chosen, P^* is always the most discerning.

Like standard welfare economics, our framework requires only information concerning the mapping from environments to choices. Because it encompasses any theory that generates a choice correspondence, it is applicable irrespective of the processes generating behavior, and regardless of whether one adopts a positive model that is preference-based, algorithmic, mechanistic, or heuristic. It generalizes standard choice-based welfare economics in two senses. First, the approaches are equivalent when

4. In this respect, our approach to behavioral welfare analysis contrasts with that of Green and Hojman (2007). They demonstrate that it is possible to rationalize apparently irrational choices as compromises among simultaneously held, conflicting preference relations, and they propose evaluating welfare based on unanimity among those relations. Unlike our framework, Green and Hojman's approach does not generally coincide with standard welfare analysis when behavior conforms to standard rationality axioms.

5. Thus, our concerns are largely orthogonal to issues examined in the literature that attempts to identify representations of nonstandard 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), Ehlers and Sprumont (2006), and Manzini and Mariotti (2007), as well as much of Green and Hojman (2007).

standard choice axioms hold. Second, for settings in which departures from those axioms are minor, our framework implies that one can approximate the appropriate welfare criterion by ignoring choice anomalies entirely. It generates natural counterparts for the standard tools of applied welfare analysis and permits a broad generalization of the first welfare theorem. It is easily applied in the context of specific positive theories and leads to novel normative implications for the familiar β, δ model of time inconsistency. For a model of coherent arbitrariness, it provides a choice-based (nonpsychological) justification for multiself Pareto optimality. Finally, though not universally discerning, it lends itself to principled refinements. Our analysis of refinements for the β, δ model provides novel ways to justify the judgments embedded in the long-run criterion and reconciles those judgments with the multiself Pareto criterion.

We begin in Section II by presenting a general framework for describing choices. Section III sets forth principles for evaluating individual welfare and applies them to specific positive models. Section IV generalizes compensating variation and consumer surplus. Section V generalizes Pareto optimality and examines competitive market efficiency as an application. Section VI demonstrates with generality that standard welfare analysis is a limiting case of our framework when behavioral anomalies are small. Section VII sets forth an agenda for refining our welfare criterion. Section VIII offers concluding remarks. Proofs appear in the Appendix.

II. A General Framework for Describing Choices

Let $\mathbb{X}$ denote the set of all possible choice objects. The elements of $\mathbb{X}$ need not be simple consumption bundles; for example, they could be lotteries, intertemporal outcome trajectories, or even consumption trajectories that depend on random and potentially welfare-relevant events.[6] Thus, despite our compact notation, the framework subsumes decision problems involving uncertainty, dynamics, and other features (discussed below).

A *constraint set* $X \subseteq \mathbb{X}$ is a collection of choice objects. When we say that the constraint set is X, we mean that, according to

6. As in the standard framework, welfare-relevant states of nature may not be observable to the planner. Thus, the framework subsumes cases in which such states are internal (e.g., hunger, or randomly occurring moods); see Gul and Pesendorder (2006).

the objective information available to the individual, the alternatives are the elements of X. The constraint set thus depends implicitly both on the objects among which the individual is actually choosing, and on the information available to him concerning those objects.

We define a *generalized choice situation* (GCS), $G = (X, d)$, as a constraint set, X, paired with an *ancillary condition*, d.[7] An ancillary condition is a feature of the choice environment that may affect behavior, but is not taken as relevant to a social planner's evaluation. Typical examples of ancillary conditions include the point in time at which a choice is made, the manner in which information or alternatives are presented, the labeling of a particular option as the "status quo," the salience of a default option, or exposure to an anchor. Notably, allowing an individual to choose between K GCSs, $(X_1, d_1), \ldots, (X_K, d_K)$, simply creates a new GCS, (X', d'), where $X' = \cup_{k=1}^{K} X_k$, and d' describes the mechanics of the decision (i.e., first choose a subset of X', and then choose an element of that subset under some specified condition).

Let $\mathcal{G}^*$ denote the set of all generalized choice situations contemplated by the positive theory of behavior for which we wish to develop a normative criterion. Thus, $\mathcal{G}^*$ is theory-specific. For example, standard consumer theory contemplates a domain with no ancillary conditions, whereas the theory of quasihyperbolic discounting contemplates a domain in which ancillary conditions specify the sequencing and timing of decisions (see Section III.E). The positive theory under consideration identifies a choice correspondence $C : \mathcal{G}^* \Rightarrow \mathbb{X}$, with $C(X, d) \subseteq X$ for all $(X, d) \in \mathcal{G}^*$, that governs the individual's behavior.[8] We interpret $x \in C(G)$ as an object that the individual is willing to choose when facing G.

When confronted with conflicting choice patterns, behavioral economists sometimes argue that certain choices are more welfare-relevant than others. In effect, they prune elements of $\mathcal{G}^*$ from the welfare-relevant domain, so that the remaining choices coherently reveal "true" objectives. We allow for pruning by defining a *welfare-relevant domain*, $\mathcal{G} \subseteq \mathcal{G}^*$, which identifies the choices from which we will take normative guidance. We will discuss

7. Rubinstein and Salant (2008) have independently formulated a similar notation for describing the impact of choice procedures on decisions; they refer to ancillary conditions as "frames."

8. For our purposes, the nature of the evidence used to recover the choice correspondence is of no consequence. The reader is free to assume that positive analysis relies exclusively on choice evidence, or that nonchoice evidence also plays a role.

potential objective criteria for pruning in Section VII. Meanwhile, our analysis will take $\mathcal{G}$ as given. Because our framework accommodates violations of standard choice axioms within $\mathcal{G}$, it permits one to demand more rigorous justifications for any deletions, even if the result is an enlarged domain that encompasses conflicting choices, such as $\mathcal{G}^*$.

Although our framework allows for a behavioral theory defined on a domain encompassing all conceivable choice situations (perhaps one that combines the features of more narrowly focused theories), we note that the prevalence of behavioral inconsistencies within that universal domain might render choice essentially useless as a normative guide. Were one to examine such a composite theory, it would be essential to identify a smaller welfare-relevant domain, for example by pruning GCSs that confuse or manipulate the decision maker.

We make two simple assumptions. The first pertains to the welfare-relevant domain, the second to the choice correspondence. We define $\mathcal{X}$ to include every constraint set X such that there is some ancillary condition d for which $(X, d) \in \mathcal{G}$.

ASSUMPTION 1. Every nonempty finite subset of $\mathbb{X}$ is contained in $\mathcal{X}$.

ASSUMPTION 2. $C(G)$ is nonempty for all $G \in \mathcal{G}$.

II.A. What Are Ancillary Conditions?

For the GCS (X, d), how does one objectively draw a line between the characteristics of the objects in the constraint set X and aspects of the ancillary condition d? In principle, one could view virtually any feature of a decision problem as a characteristic of the available objects. Yet if we incorporated *every* feature of each decision problem into the descriptions of the objects, then each object would be available in one and only one decision problem, and choices would provide little in the way of useful normative guidance. Consequently, practical considerations *compel* us to adopt a more limited conception of an object's attributes.

One natural way to draw the required line is to distinguish between conditions that pertain exclusively to experience and conditions that pertain at least in part to choice.[9] Conditions that pertain exclusively to experience do not change when a decision is

9. For example, hunger at the time of choice would be an ancillary condition, while hunger at the time of consumption would not.

delegated from an individual to a social planner. Consequently, if the planner treats such conditions as characteristics of the available objects, he can still take guidance from the choices the individual would make. If the planner must provide the individual with either a red car or a green car, he can sensibly ask which one the individual would choose; the meaning of color does not change with the chooser. In contrast, a condition that pertains to choice necessarily changes when the decision is delegated, because it then references a different chooser. If a planner were to treat such conditions as characteristics of the available objects, he would be forced to acknowledge that delegation necessarily changes the objects, in which case he would no longer be able to take guidance from a hypothetical undelegated choice. If he wishes to take such guidance, he must therefore define objects' characteristics to exclude conditions of choice. Within our framework, one can classify those excluded conditions as ancillary; if indeed they affect behavior, then one simply concludes that choice offers ambiguous guidance concerning the delegated decision problem.

Consider the example of time inconsistency. Suppose alternatives x and y yield payoffs at time t; the individual selects x over y when choosing at time t, and y over x when choosing at $t-1$. Note that we could include the time of choice in the description of the objects: when choosing between x and y at time k, the individual actually chooses between "x chosen by the individual at time k" and "y chosen by the individual at time k" ($k = t, t-1$). With that formulation, the objects of choice are different at distinct points in time, so reversals involve no inconsistency. But then, when the decision is delegated, the objects become "x chosen by the planner at time k" and "y chosen by the planner at time k." Because that set of options is entirely new, neither the individual's choice at time t, nor his choice at time $t-1$, offers useful guidance. If we wish to construct a theory of welfare based on the choice correspondence 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.[10] Some might argue that the individual's choice at $t-1$ is the planner's best guide because it is at arm's length from the

10. Another option would be to define the goods as "x chosen at time k" and "y chosen at time k," omitting the phrases "by the individual" and "by the planner." A planner who looks to the individual's choices for guidance would then choose x at time t and y at time $t-1$. None of the existing work on time consistency adopts that standard. The reason is clear: these definitions of the objects ignore the fact

experience and hence does not trigger the psychological processes responsible for apparent lapses of self-control; others might insist that the choice at t is the best guide because reward is properly appreciated only in the moment and excessively intellectualized at arm's length. The first position argues for excluding the time t choice from the welfare-relevant domain $\mathcal{G}$; the second argues for excluding the time $t-1$ choice. Choice patterns alone cannot resolve that controversy. Including both choices in $\mathcal{G}$ and treating the time of choice as an ancillary condition permits us to recognize the conflict, remain agnostic, and embrace the implied ambiguity.

In many cases (e.g., when exposure to an arbitrary number influences choice), treating a condition of choice as a welfare-relevant characteristic of the available objects would seem to defy common sense; consequently, classifying it as an ancillary condition should be relatively uncontroversial. Other cases may be less clear. Different analysts may wish to draw different lines between the characteristics of choice objects and ancillary conditions, based either on the distinction between conditions of choice and experience discussed above, or on completely different criteria. We therefore emphasize that the tools we develop in this paper 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. Where differences in line-drawing lead to different normative conclusions, our framework usefully pinpoints the source of disagreement.

Drawing a 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 the approaches. First, with our approach, 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 ambiguous welfare comparisons where choice data conflict; in contrast, an experienced utility function admits no ambiguity.

II.B. Scope of the Framework

Our framework can incorporate nonstandard behavioral patterns in four separate ways. (1) It allows choice to depend

---

that the condition of choice pertains to the chooser. Specifically, the significance of making the choice at time t changes when the decision is delegated to the planner.

on ancillary conditions, thereby subsuming a wide range of behavioral phenomena. Specifically, the typical anomaly involves a constraint set, X, along with two ancillary conditions, d' and d'', for which $C(X, d') \neq C(X, d'')$. This is sometimes called a *preference reversal*, but in the interests of greater precision we call it a *choice reversal*. We listed some well-known examples at the outset of Section II. (2) Our framework does not impose any counterparts to standard choice axioms. Indeed, throughout most of this paper, we allow for *all* nonempty choice correspondences (Assumption 2), even ones for which choices are intransitive or depend on "irrelevant" alternatives (entirely apart from ancillary conditions). (3) Our framework subsumes the possibility that people can make choices from opportunity sets that are not compact (e.g., selecting "almost best" elements). (4) We can interpret a choice object $x \in \mathbb{X}$ more broadly than in the standard framework (e.g., as in Caplin and Leahy [2001], who axiomatize anticipatory utility by treating the time at which uncertainty is resolved as a characteristic of a lottery).

III. Individual Welfare

Welfare analysis typically requires us to judge whether one alternative represents an *improvement* over another, even when the new alternative is not necessarily the best one. For that purpose, we require a binary relation, call it Q, where xQy means that x improves upon y. Within the standard framework, the revealed preference relation serves that role.

When imposing standard choice axioms, one typically defines the weak and strict revealed preference relations in terms of choices from binary sets: $xR'y$ (xP^*y) is equivalent to the statement that $x \in C(\{x, y\})$ ($y \notin C(\{x, y\})$). Those definitions imply the following:[11]

$$(1) \qquad xRy \text{ iff, for all } X \in \mathcal{X} \text{ with } x, y \in X, y \in C(X) \text{ implies } x \in C(X),$$

$$(2) \qquad xPy \text{ iff, for all } X \in \mathcal{X} \text{ with } x, y \in X, \text{ we have } y \notin C(X).$$

Expressions (2) and (2) immediately suggest two natural generalizations of revealed preference. The first extends (2), the weak

11. The implications follow from WARP. Note that the definition of P, below, differs from the one proposed by Arrow (1959), which requires only that there be some $X \in \mathcal{X}$ with $x, y \in X$ for which $x \in C(X)$ and $y \notin C(X)$.

revealed preference relation:

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

The statement "$xR'y$" means that whenever x and y are both available, y isn't chosen unless x is as well. We will then say that x is *weakly unambiguously chosen over* y. Let P' denote the asymmetric component of R' ($xP'y$ iff $xR'y$ and $\sim yR'x$), and let I' denote the symmetric component ($xI'y$ iff $xR'y$ and $yR'x$). The statement "$xP'y$" means that whenever x and y are available, sometimes x is chosen but not y, and otherwise either both or neither are chosen. The statement "$xI'y$" means that, whenever x is chosen, so is y, and vice versa.

The second generalization of revealed preference extends (2), strict revealed preference:

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

The statement "xP^*y" means that whenever x and y are available, y is never chosen. We will then say that x is *strictly unambiguously chosen over* y (sometimes dropping "strictly" for the sake of brevity). As a general matter, P' and P^* may differ. However, if C maps each $G \in \mathcal{G}$ to a unique choice, they necessarily coincide. We note that Rubinstein and Salant (2008) have separately proposed a binary relation that is related to P' and P^*.[12]

There are many binary relations for which P^* is the asymmetric component; each is a potential generalization of weak revealed preference. The coarsest is, of course, P^* itself. The finest, R^*, is defined by the property that xR^*y iff $\sim yP^*x$.[13] 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. Let I^* be the symmetric component of R^* (xI^*y iff xR^*y and yR^*x). The statement "xI^*y" means that there is at least one GCS for which x is chosen with

12. The following is a description of Rubinstein and Salant's (2008) binary relation, using our notation. Assume C is always single-valued. Then $x \succ y$ iff $C(\{x, y\}, d) = x$ for all d such that $(\{x, y\}, d) \in \mathcal{G}$. In contrast to P' or P^*, the relation $\succ$ 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.

13. One binary relation, A, is *weakly coarser* than another, B, if xAy implies xBy. When A is weakly coarser than B, B is *weakly finer* than A.

y available, and at least one GCS for which y is chosen with x available.

When choices are invariant with respect to ancillary conditions and satisfy standard axioms, R' and R^* specialize to R, whereas P' and P^* specialize to P. Thus, our framework subsumes standard welfare economics as a special case.

III.A. Some Properties of the Welfare Relations

How are R', P', and I' related to R^*, P^*, and I^*? It is easy to check that xP^*y implies $xP'y$ implies $xR'y$ implies xR^*y, so that P^* is the coarsest of these relations and R^* the finest. Also, $xI'y$ implies xI^*y.

The relation R^* is always complete, but R' need not be, and there is no guarantee that any of the relations defined here are transitive. (See Example 1 below for an illustration of intransitivity involving P^*.) However, to conduct useful welfare analysis, one does not require transitivity. Our first main result establishes that there cannot be a cycle involving R', the direct generalization of weak revealed preference, if one or more of the comparisons involve P^*, the direct generalization of strict revealed preference.

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

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." [14] The theorem has an immediate and important corollary:

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

Like transitivity, acyclicity guarantees the existence of maximal elements for finite sets and allows us to both identify and measure unambiguous improvements. Thus, regardless of how poorly behaved the choice correspondence may be, P^* is always a viable welfare criterion. In contrast, it is easy to devise examples in which P' cycles.

14. In the context of standard decision theory, Suzumura's (1976) analogous consistency property plays a similar role. A preference relation R is *consistent* if $x_1 R x_2 \ldots R x_N$ with $x_i P x_{i+1}$ for some i implies $\sim x_N R x_1$ (where P is the asymmetric component of R). Theorem 1 has the following trivial corollary: If C maps each $G \in \mathcal{G}$ to a unique choice (so that P' coincides with P^*), then R' is consistent.

III.B. Individual Welfare Optima

We will say that it 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 $yP'x$. When a strict improvement is impossible, we say that x is a *weak individual welfare optimum*. When a weak improvement is impossible, we say that x is a *strict individual welfare optimum*.

The following two observations (which follow immediately from the definitions) characterize individual welfare optima.

Observation 1. Every $x \in C(X, d)$ for $(X, d) \in \mathcal{G}$ 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 guarantees the existence of weak welfare optima and assures us that our welfare criterion respects a natural "libertarian" principle: any action voluntarily chosen from a set X within the welfare-relevant choice domain, $\mathcal{G}$, is a weak optimum within X. Thus, according to the relation P^*, it is impossible to design an intervention that "improves" on a choice made by the individual within $\mathcal{G}$. Nevertheless, it may be possible to improve decisions made in any GCS that is not considered welfare-relevant (i.e., elements of $\mathcal{G}^*$ that are excluded from $\mathcal{G}$); see Section VII.[15] It may also be possible to improve upon market outcomes when market failures are present, just as in standard economics; see Section V.B.

The fact that we have established the existence of weak individual welfare optima without making any additional assumptions, for example, related to continuity and compactness, may at first seem surprising, but it simply reflects our assumption that the choice correspondence is well-defined over the set $\mathcal{G}$. Standard existence issues arise when the choice function is built up from other components. The following example clarifies these issues.

15. Many behavioral economists have proposed interventions which, they claim, would improve on individual choices; see, for example, Thaler and Sunstein's (2003) discussion of *libertarian paternalism*. Those claims reflect assumptions, often implicit, concerning which choices are and are not appropriate guides to welfare. We are sympathetic to the view that it may be possible and desirable to make such judgments in some settings; that is why we allow the welfare-relevant domain $\mathcal{G}$ to diverge from the full domain $\mathcal{G}^*$. However, as discussed in Section VII, we would prefer to see those judgments stated explicitly, and justified where possible based on objective criteria.

Example 1. Suppose $\mathcal{G} = \{X_1, \ldots, X_4\}$ (plus singleton sets, for which choice is trivial), with $X_1 = \{a, b\}$, $X_2 = \{b, c\}$, $X_3 = \{a, c\}$, and $X_4 = \{a, b, c\}$ (there are no ancillary conditions). Imagine that $C(X_1) = \{a\}$, $C(X_2) = \{b\}$, $C(X_3) = \{c\}$, and $C(X_4) = \{a\}$. Then we have aP^*b and bP^*c; in contrast, aI^*c. Despite the intransitivity of P^*, 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), and b is a strict welfare optimum in X_2 (c is not a weak optimum). Notably, both a and c are strict welfare optima in X_3, despite the fact that only c is chosen from X_3; a survives because it is chosen over c in X_4, which makes a and c not comparable under P^*.

Now let us limit attention to $\mathcal{G}' = \{X_1, X_2, X_3\}$. In that case, Assumption 1 is violated ($\mathcal{G}'$ does not contain all finite sets) and P^* cycles ($aP^*bP^*cP^*a$). If we wish to create a preference or utility representation based on the data contained in $\mathcal{G}'$ so that we can project the individual's choice within the set X_4, the intransitivity will pose a difficulty. And if we try to prescribe a welfare optimum for 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.[16] 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.

The previous example illustrates that the alternatives chosen from a set need not be the only individual welfare optima within that set (specifically, a is an optimum in X_3, but is not chosen from X_3). Our next observation accounts for that possibility.

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.

The following example, based loosely on an experiment reported by Iyengar and Lepper (2000), illustrates why one can

16. Even so, individual welfare optima exist within every set that falls within the restricted domain. Here, 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.

reasonably treat an alternative as an individual welfare optimum within a set even though the decision maker never chooses it from that set. Suppose a subject chooses strawberry jam when only one other flavor is available (regardless of what it is, and assuming he also has the option to take nothing), but rejects all flavors (including strawberry) in favor of nothing when thirty are available. In the latter case, one could argue that taking nothing is his best alternative because he chooses it. But one could also argue that strawberry jam is his best alternative because he chooses it over all of his other alternatives when facing simpler, less overwhelming decision problems. Our framework recognizes that both judgments are potentially valid on the basis of choice patterns alone.

III.C. Further Justification for P^*

Though the binary welfare relations proposed herein are natural and intuitive generalizations of the standard welfare relations, one could in principle devise alternatives. Here we provide an additional justification for favoring P^*. We have seen that P^* never overrules choice, in the sense that any object chosen from a set X in some welfare-relevant condition is necessarily a weak individual welfare optimum within X. Here we show that all other relations with that desirable property are less discerning than P^*.

Consider a choice correspondence C defined on $\mathcal{G}$ 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:

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

Also, for $X \in \mathcal{X}$, let $D(X)$ be the set of ancillary conditions associated with X:

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

We will say that Q is an *inclusive libertarian relation* for a choice correspondence C if, for all $X \in \mathcal{X}$, the maximal elements under Q include all of the elements the individual would choose from X, considering all associated ancillary conditions (formally, $\cup_{d \in D(X)} C(X.d) \subseteq m_Q(X)$). Such a relation never overrules choice in the sense mentioned above, and all other relations overrule choice in some circumstance.

Observation 1 implies that P^* is an inclusive libertarian relation, but it is not the only one. For example, the null relation,

R^{Null} ($\sim x R^{\text{Null}} y$ for all $x, y \in \mathbb{X}$), falls into that category. Yet R^{Null} is less discerning than P^*. According to the following result, so are all other inclusive libertarian relations.

THEOREM 2. Consider any choice correspondence C, and any asymmetric inclusive libertarian relation $Q \neq P^*$. Then P^* is finer than Q. Thus, 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, $m_{P^*}(X) \subseteq m_Q(X)$).

Ideally, for a given choice correspondence C, one might wish to find a binary welfare relation Q such that, for all $X \in \mathcal{X}$, the maximal elements under Q coincide *exactly* with the elements the individual would choose from X, considering all associated ancillary conditions (formally, $\cup_{d \in D(X)} C(X.d) = m_Q(X)$). We will call any such Q a *libertarian relation* for C.[17] Because any libertarian relation is also an inclusive libertarian relation, Theorem 2 implies that a libertarian relation exists for a choice correspondence C if and only if P^* is libertarian.[18] Thus, whenever there exists some preference relation that rationalizes choice on $\mathcal{G}$, P^* provides such a rationalization.

In principle, for a given choice correspondence C, one might also wish to find a binary welfare relation Q such that, for all $X \in \mathcal{X}$, every maximal element under Q is chosen from X for some ancillary condition (formally, $m_Q(X)$ is nonempty, and $m_Q(X) \subseteq \cup_{d \in D(X)} C(X.d)$). We will call any such Q an *exclusive libertarian relation* for C. We focus on inclusive rather than exclusive libertarian relations for three reasons. First, in our view, if an individual is willing to choose x from the set X within the welfare-relevant domain, a choice-based welfare criterion should not declare that x is improvable within X. Second, there

17. In the absence of ancillary conditions, the statement that Q is a libertarian relation for C is equivalent to the statement that Q *rationalizes* C (see, e.g., Bossert, Sprumont, and Suzumura [2005]). As is well known, one must impose restrictive conditions on C to guarantee the existence of a rationalization. For instance, there is no rationalization (and hence no libertarian relation) for the choice correspondence described in Example 1. One naturally wonders about the properties that a generalized choice correspondence must have to guarantee the existence of a libertarian relation. See Rubinstein and Salant (2008) for an analysis of that issue.

18. Indeed, according to Theorem 2, if there is an inclusive libertarian relation Q for a choice correspondence C and a choice set X for which the set of maximal elements under Q coincides exactly with the set of chosen elements (that is, $\cup_{d \in D(X)} C(X) = m_Q(X)$), then the set of maximal elements under P^* also coincides exactly with the set of chosen elements.

are sometimes good reasons to treat objects not chosen from a set as individual welfare optima within that set (recall the jam example in Section III.B). Finally, exclusive libertarian relations do not exist for many choice correspondences.[19]

One might also consider a more direct interpretation of choice-based welfare economics: 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. However, that approach does not allow us to determine whether a change from one element of X to another is an *improvement*, except in cases where the individual would choose either the initial or final element from X. For that purpose we require a binary relation.

III.D. Relation to Multiself Pareto Optima

Under certain restrictive conditions, our notion of an individual welfare optimum coincides with the idea of a multiself Pareto optimum. That criterion is most commonly invoked in the literature on quasi-hyperbolic discounting, where it is applied to an individual's many time-dated "selves" (see, e.g., Laibson, Repetto, and Tobacman [1998]).

Suppose $\mathcal{G}$ is the Cartesian product of the set of constraint sets and a set of ancillary conditions ($\mathcal{G} = \mathcal{X} \times D$, where $d \in D$); in that case, we say that $\mathcal{G}$ is *rectangular*. Suppose also that, for each $d \in D$, choices correspond to the maximal elements of a well-behaved preference ranking R_d, and hence to the alternatives that maximize a utility function u_d.[20] One can then imagine that each ancillary condition activates a different "self" and apply the Pareto criterion across selves. We will say that y *weakly multiself Pareto dominates* x, written yMx, iff $u_d(y) \geq u_d(x)$ for all $d \in D$, with strict inequality for some d; it *strictly multiself Pareto dominates* x, written yM^*x, iff $u_d(y) > u_d(x)$ for all $d \in D$. Moreover, $x \in X \subset \mathbb{X}$ is a *weak (strict) multiself Pareto optimum* in X if there is no $y \in X$ such that yM^*x (yMx).

THEOREM 3. Suppose $\mathcal{G}$ is rectangular and choices for each $d \in D$ maximize a utility function u_d. Then $M^* = P^*$ and $M = P'$. It

19. It is easy to construct examples of choice correspondences that violate WARP for which no exclusive libertarian relation exists; see Example 5 in Bernheim and Rangel (2008a).

20. 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 those assumptions play no role in what follows.

follows that $x \in X$ is a weak (strict) multiself Pareto optimum in X iff it is a weak (strict) individual welfare optimum.

Thus, in certain narrow settings, our approach justifies the multiself Pareto criterion without invoking potentially controversial psychological assumptions, such as the existence of multiple coherent decision-making entities within the brain. That justification does *not* apply to quasihyperbolic consumers because $\mathcal{G}^*$ is not rectangular; however, it *does* justify the use of the multiself Pareto criterion for cases of "coherent arbitrariness," such as those studied by Ariely, Loewenstein, and Prelec (2003) (see Section III.E).

III.E. Applications to Specific Positive Models

Next we explore the implications of our framework in the context of two specific positive models: coherent arbitrariness and quasihyperbolic discounting. Partially coherent choice patterns should also provide adequate traction for choice-based normative analysis in other settings. A few additional examples include individuals who (a) sometimes exhibit choice reversals when alternatives are listed in different orders, (b) always notice stochastic dominance in lotteries but are otherwise susceptible to framing effects, or (c) do not notice small differences between alternatives, but choose coherently when differences are large.

Coherent Arbitrariness. Behavior is coherently arbitrary when some psychological anchor (for example, calling attention to a number) affects choice, but the individual nevertheless conforms to standard axioms for any fixed anchor (see Ariely, Loewenstein, and Prelec [2003], who construed this pattern as an indictment of the revealed preference paradigm). To illustrate, let us suppose that an individual consumes two goods, y and z, and that we have the following representation of decision utility:

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

with u and v strictly increasing, differentiable, and strictly concave. We interpret the ancillary condition, $d \in [d_L, d_H]$, as an anchor that influences decision utility.

Because $\mathcal{G}$ is rectangular, and because choices maximize $U(y, z \mid d)$ for each d, Theorem 3 implies that our welfare criterion is equivalent to the multiself Pareto criterion, where each d

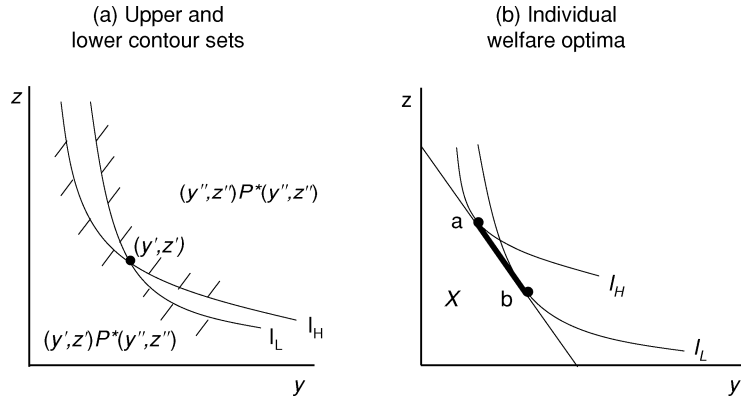

FIGURE I
Coherent Arbitrariness

indexes a different self. It follows that

$$(3) \qquad (y', z')R'(y'', z'') \text{ iff } u(y') + dv(z') \geq u(y'') + dv(z'') \text{ for } d = d_L, d_H.$$

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

Figure Ia shows two decision-indifference curves (that is, indifference curves derived from decision utility) passing through the bundle (y', z'), one for d_L (labeled I_L) and one for d_H (labeled I_H). All bundles (y'', z'') lying below both decision-indifference curves satisfy $(y', z')P^*(y'', z'')$; this is the analog of a lower contour set. All bundles (y'', z'') lying above both decision-indifference curves satisfy $(y'', z'')P^*(y', z')$; 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. As shown in Figure Ib, the individual 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 thick segment of the budget line between bundles a and b is uniquely chosen for some $d \in [d_L, d_H]$, so all these bundles are strict individual welfare optima. It is easy to prove that there are no other welfare optima, weak or strict.

As the gap between d_L and d_H shrinks, the set $\{(y'', z'') \mid (y'', z'') P^*(y', z')\}$ 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 generalize that observation in Section VI.

Dynamic Inconsistency. Consider the well-known β, δ model of quasihyperbolic 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 multiself Pareto criterion, which associates each moment in time with a different self, or the "long-run criterion," which assumes that well-being is described by exponential discounting at the rate δ. As we will see, our framework leads to a different criterion.

Suppose the consumer's task is to choose a consumption vector, $C_1 = (c_1, \ldots, c_T)$, where $c_t \geq 0$ denotes the level of consumption at time t. Let C_t denote the continuation consumption vector $(c_t, \ldots, c_T)$. Choices at time t maximize the function

$$U_t(C_t) = u(c_t) + \beta \sum_{k=t+1}^{T} \delta^{k-t} u(c_k), \tag{4}$$

where $\beta, \delta \in (0, 1)$. We assume perfect foresight concerning future decisions, so that behavior is governed by subgame perfect equilibria. We also assume $u(0)$ is finite; for convenience, we normalize $u(0) = 0$.[21] Finally, we assume $\lim_{c \to \infty} u(c) = \infty$.

To conduct normative analysis, we must recognize that the selection of an intertemporal consumption vector involves only one choice by a single decision maker. Critically, that statement remains valid even when the individual makes the decision over time in a series of steps (notwithstanding the common practice of modeling such problems as games between multiple time-dated selves); he still selects a single consumption trajectory. For this positive model, a GCS $G = (X, \tau)$ involves a set of lifetime

21. 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. One can show that if conceivable consumption is unbounded and u is unbounded both above and below, then no alternative in $\mathbb{R}_{++}$ is unambiguously chosen over any other alternative.

consumption vectors, X, and a decision tree, τ, for selecting an element of X. The decision tree describes the options available at each point in time (including precommitment opportunities), how those options depend on past actions, and how they affect the options that will be available in future periods. There are typically many different trees that allow the individual to select from any given X. Because some decisions depend on the points in time at which they are made, we may have $C(X, \tau) \neq C(X, \tau')$ for $\tau \neq \tau'$; that is why we treat τ as an ancillary condition.

For every possible constraint set X, $\mathcal{G}^*$ includes every conceivable pair (X, τ), where τ is the decision tree for selecting from X. Note that $\mathcal{G}^*$ is not rectangular: decision trees are tailored to constraint sets, and in any case the individual cannot chose consumption for period t using a tree that allows no choice until period $k > t$. Hence, Theorem 3, which identifies conditions that justify the multiself Pareto criterion, does not apply.

The following result completely characterizes R' and P^* for the β, δ model, assuming that the welfare-relevant domain $\mathcal{G}$ coincides with the full choice domain $\mathcal{G}^*$.[22]

THEOREM 4. Let $\mathcal{G} = \mathcal{G}^*$. Define $W_t(C_t) \equiv \sum_{k=t}^{T} (\beta\delta)^{k-t} u(c_k)$. Then $C_1' R' C_1''$ iff $W_1(C_1') \geq U_1(C_1'')$, and $C_1' P^* C_1''$ iff $W_1(C_1') > U_1(C_1'')$. Moreover, R' and P^* are transitive.

In effect, the theorem tells us that it is possible to design an intrapersonal game in which C_1'' is chosen when C_1' is feasible if and only if $W_1(C_1') \leq U_1(C_1'')$. Thus, to determine whether C_1', is unambiguously chosen over 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 $\beta\delta$ (that is, $W_1(C_1')$). 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''.[23] That observation explains the transitivity of the welfare

22. From the characterization of R', we can deduce that $C_1' I' C_1''$ iff $W_1(C_1') = U_1(C_1') = W_1(C_1'') = U_1(C_1'')$, which requires $c_k' = c_k'' = 0$ for $k > 2$. Thus, for comparisons involving consumption profiles with strictly positive consumption in the third period or later, P' coincides with R'. From the characterization of P^*, we can deduce that (i) $C_1' R^* C_1''$ iff $U_1(C_1') \geq W_1(C_1'')$, and (ii) $C_1' I^* C_1''$ iff $U_1(C_1') \geq W_1(C_1'')$ and $U_1(C_1'') \geq W_1(C_1')$.

23. 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''.

relation.[24] It also implies that any welfare improvement under P^* or P' must also be a welfare improvement under U_1, the decision utility at the first moment in time.

From Theorem 4, it follows that C_1 is a weak welfare optimum in X if and only if the decision utility that C_1 provides at $t = 1$ is at least as large as the highest available discounted value of u, using $\beta\delta$ as a time-consistent discount factor. Formally:

COROLLARY 2. For any consumption set X, C_1 is a weak welfare optimum in X iff $U_1(C_1) \geq \sup_{C_1' \in X} W_1(C_1')$. If $U_1(C_1) > \sup_{C_1' \in X} W_1(C_1')$, then C_1 is a strict welfare optimum in X.[25]

Notice that, for all C_1, $\lim_{\beta \to 1}[W_1(C_1) - U_1(C_1)] = 0$. 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 multiself 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 multiself Pareto optimum, even when $\beta = 1$.

Note that if the relevant time periods are short (e.g., days) and the value of β is noticeably less than one (e.g., 0.95), then the welfare criterion identified in Theorem 4 may be discerning only when applied to problems with short planning horizons (e.g., short-term procrastination, but not retirement). In Section VII, we discuss potential criteria for restricting the welfare-relevant domain $\mathcal{G}$, thereby generating more discerning criteria.

IV. TOOLS FOR APPLIED WELFARE ANALYSIS

In this section we show that the concept of compensating variation has a natural counterpart within our framework; the same is true of equivalent variation (for analogous reasons). We also illustrate how, under more restrictive assumptions, the generalized compensating variation of a price change corresponds to an analog of consumer surplus.

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

25. If $U_1(C_1) = \sup_{C_1' \in X} W_1(C_1')$, then C_1 may or may not be a strict welfare optimum.

IV.A. Compensating Variation

Let us assume that the individual's constraint set, $X(\alpha, m)$, 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 condition, and $(X(\alpha_0, 0), d_0)$ the initial GCS. We will consider a change in parameters to α_1 and in the ancillary condition to d_1, along with 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.[26]

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. In extending that definition to our framework, we encounter three ambiguities. The first arises when the individual is willing to choose more than one alternative either in the initial GCS $(X(\alpha_0, 0), d_0)$, or in the final GCS, $(X(\alpha_1, m), d_1)$. Unlike in the standard framework, comparisons may depend on the particular pair considered. We handle that ambiguity by insisting that compensation be adequate for all pairs of outcomes that could be chosen from the initial and final sets.

A second ambiguity arises from a potential form of nonmonotonicity. Without further assumptions, we cannot guarantee that, if the payment m is adequate to compensate an individual for some change, then any $m' > m$ is also adequate. We handle that issue by finding a level of compensation beyond which such reversals do not occur. (We discuss an alternative in Appendix D of Bernheim and Rangel [2008b].)

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? That ambiguity is an essential feature of welfare evaluations with inconsistent choice. Accordingly, we define two notions of compensating variation:

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

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

26. This formulation of compensating variation assumes $\mathcal{G}$ 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 varied with m.

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

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

In other words, all levels of compensation greater than the CV-A (smaller than CV-B) guarantee that everything selected in the new (initial) set is unambiguously chosen over everything selected from the initial (new) set.[27] It is easy to verify that $m^A \geq m^B$. Also, when $\alpha_1 = \alpha_0$ and $d_1 \neq d_0$, we always have $m^A \geq 0 \geq m^B$. Thus, the welfare effect of a change in the ancillary condition, by itself, is always ambiguous.

Theorem 1 guarantees that CV-A and CV-B are well-behaved welfare measures in the following sense: If the individual experiences a sequence of changes and is adequately compensated for each in the sense of the CV-A, no alternative he would select from the initial set is unambiguously chosen over any alternative he would select from the final set.[28] 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 he would select from the final set is unambiguously chosen over any alternative he would select from the initial set.

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.[29] However, we are not troubled by nonadditivity. 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) is not a serious conceptual difficulty.

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

28. 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, for some $\eta > 0$, m_2^A is the CV-A for a change from $(X(\alpha_1, m_1^A + \eta), d_1)$ to $(X(\alpha_2, m_1^A + \eta + m), d_2)$, then nothing that the individual would choose from $(X(\alpha_0, 0), d_0)$ is unambiguously chosen over anything that he would choose from $(X(\alpha_2, m_1^A + \eta + m_2^A + \varepsilon), d_2)$ for $\varepsilon > 0$.

29. 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.

IV.B. Consumer Surplus

Under more restrictive assumptions, the compensating variation of a price change corresponds to an analog of consumer surplus. Let us consider again the model of coherent arbitrariness, but assume a more restrictive form of decision utility (which involves no income effects, so that Marshallian consumer surplus would be valid in the standard framework):

$$U(y, z \mid d) = y + dv(z). \tag{5}$$

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

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. 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 $z_0 > z_1$. Because there are no income effects, z_1 will not change as the individual is compensated. The following result provides a simple formula for CV-A and CV-B:

THEOREM 5. Suppose decision utility is given by equation (5), and consider a change from (p_0, d_0) to (p_1, d_1). Let $m(d) = [p_1 - p_0]z_1 + \int_{z_1}^{z_0} [P(z, d) - p_0]dz$ (where z_k satisfies $P(z_k, d_k) = p_k$ for $k = 0, 1$, and $z_0 > z_1$). Then $m^A = m(d_H)$ and $m^B = m(d_L)$.

The first term in the expression for $m(d)$ is the extra amount the consumer pays for the first z_1 units. The second term involves the area between the demand curve and a horizontal line at p_0 between z_1 and z_0 when d is the ancillary condition. Figure IIa provides a graphical illustration of CV-A, analogous to ones found in most microeconomics textbooks: it is the sum of the areas labeled A and B. Figure IIb illustrates CV-B: it is the sum of the areas labeled A and C, minus the area labeled E. Note that 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. As the range of possible ancillary conditions narrows, CV-A and CV-B both converge to standard consumer surplus, a property that we generalize in Section VI.

For an application of this framework to a practical problem involving the salience of sales taxes, as well as for an extension to settings with income effects, see Chetty, Looney, and Kroft (2008).

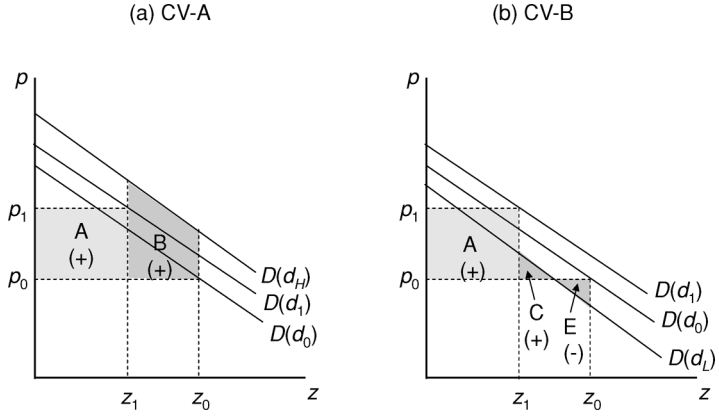

FIGURE II
CV-A and CV-B for a Change from (p_0, d_0) to (p_1, d_1)

V. WELFARE ANALYSIS INVOLVING MORE THAN ONE INDIVIDUAL

In this section we describe a natural generalization of Pareto optimality to settings with behavioral anomalies, and we illustrate its use by examining the efficiency of competitive market equilibria.

V.A. Generalized Pareto Optima

Suppose there are N individuals indexed $i = 1, \ldots, N$. 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 correspondence 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 correspondences on $\mathcal{G}_i \subseteq \mathcal{G}_i^*$ induce the relations R_i' and P_i^* over $\mathbb{X}$. Assume $\mathcal{X} \in \mathcal{X}_i$ for all $\mathcal{X} \subseteq \mathbb{X}$.

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.[30] 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

30. 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.

have incomplete and/or intransitive preferences (see, e.g., Fon and Otani [1979], Rigotti and Shannon [2005], or Mandler [2006]).[31]

Because 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 $(X, d) \in \mathcal{G}_i$.

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

Example 2. Consider a two-person exchange economy involving two goods, y and z. Suppose the choices of consumer 1 are described by the model of coherent arbitrariness discussed earlier, whereas consumer 2's choices respect standard axioms. In Figure III, we have drawn two standard contract curves. The one labeled T_H is formed by the tangencies between the consumers' indifference curves when consumer 1 faces ancillary condition d_H (such as the point at which I_{1H} touches I_2); the one labeled T_L is formed by the tangencies when consumer 1 faces ancillary condition d_L (such as the point at which I_{1L} touches I_2). The shaded area between those two curves is the generalized 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.[32] Like a standard contract curve, the generalized contract curve runs between the

31. 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 within X under the preference relation. In contrast, in our framework, an individual is not necessarily willing to select any element of X that is maximal within X under the individual welfare relation P^*. (Recall that P^* is an inclusive libertarian relation, but that it need not rationalize the choice correspondence.) However, for the limited purpose of characterizing socially efficient outcomes, choice is not involved, so that distinction is immaterial. Thus, as illustrated in an example below, existing results concerning the structure or characteristics of the Pareto efficient set with incomplete and/or intransitive preferences apply in our setting.

32. Notably, in another setting with incomplete preferences, Mandler (2006) demonstrates with generality that the Pareto-efficient set has full dimensionality.

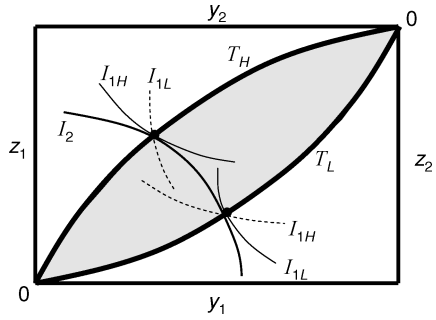

FIGURE III
The Generalized Contract Curve

southwest and northeast corners of the Edgeworth box, so there are many intermediate Pareto optima. If the behavioral effects of the ancillary conditions were smaller, the generalized contract curve would be thinner; in the limit, it would converge to a standard contract curve. (Section VI generalizes that point.)

Our next result establishes with generality (and with no further assumptions) that, just as in Figure III, for any set X, one can start with *any* alternative $x \in X$ and find a Pareto optimum over which no individual unambiguously chooses x.[33]

THEOREM 6. For every $x \in X$, the nonempty set $\{y \in X \mid \forall i, \sim x P_i^* y\}$ includes at least one weak generalized Pareto optimum in X.

V.B. 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. Let x^n denote the consumption vector of consumer n, z^n the endowment vector of consumer n, $\mathbb{X}^n$ the consumption set for consumer n, and y^f 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. Let Y denote the aggregate production set. We will say that an allocation $x = (x^1, \ldots, x^N)$ is

33. The proof of Theorem 6 is more subtle than one might expect; in particular, there is no guarantee that any individual's welfare optimum within the set $\{y \in X \mid \forall i, \sim x P_i^* y\}$ is a generalized Pareto optimum within X.

feasible if $\sum_{n=1}^{N}(x^n - z^n) \in Y$ and $x^n \in X^n$ for all n. Trade occurs at a price vector π subject to ancillary conditions $d = (d^1, \ldots, d^N)$, where d^n pertains to consumer n. The price vector π implies a budget constraint $B^n(\pi) = \{x^n \in \mathbb{X}^n \mid \pi x^n \leq \pi z^n\}$ for consumer n.

We assume that profit maximization governs the choices of firms. Consumer n's behavior is described by a choice correspondence $C^n(X^n, d^n)$, 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, $\widehat{x} = (\widehat{x}^1, \ldots, \widehat{x}^N)$, a production allocation, $\widehat{y} = (\widehat{y}^1, \ldots, \widehat{y}^F)$, and a set of ancillary conditions, $\widehat{d} = (\widehat{d}^1, \ldots, \widehat{d}^N)$, 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 $y^f \in Y^f$.[34]

Fon and Otani (1979) established the efficiency of competitive equilibria in exchange economies when consumers have incomplete and/or intransitive preferences (see also Rigotti and Shannon [2005] and Mandler [2006]). The efficiency of behavioral competitive equilibria in exchange economies (a much more general statement) follows as a corollary of their theorem.[35] A similar argument establishes efficiency for production economies.

THEOREM 7. If all choices are welfare-relevant ($\mathcal{G}_n = \mathcal{G}^*_n$), then the allocation associated with any behavioral competitive equilibrium is a weak generalized Pareto optimum.[36]

The generality of Theorem 7 is worth emphasizing: it establishes the efficiency of competitive equilibria within a framework

34. One could endogenize the ancillary conditions by supplementing this definition with additional equilibrium requirements. However, Theorem 7 would still apply.

35. 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 $P^*_1, \ldots, P^*_N$. 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. 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).

36. One can also show that a behavioral competitive equilibrium is a strict generalized Pareto optimum under the following additional assumption (which is akin to nonsatiation): 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.

that imposes almost no restrictions on consumer behavior, thereby allowing virtually any conceivable choice pattern, including all anomalies documented in the behavioral literature. Note, however, that the theorem plainly need not hold if firms pursue objectives other than profit maximization. Thus, we see that the first welfare theorem is driven by assumptions concerning the behavior of firms, not consumers.

Naturally, behavioral competitive equilibrium can be inefficient in the presence of sufficiently severe but otherwise standard market failures. In addition, even a perfectly competitive behavioral equilibrium may be inefficient when judged by a welfare relation derived from a restricted welfare-relevant choice domain ($\mathcal{G}_n \subset \mathcal{G}_n^*$). This observation alerts us to the fact that, in behavioral economies, there is a new class of potential market failures involving choice situations that have been pruned from $\mathcal{G}_n^*$. Our analysis of addiction (Bernheim and Rangel 2004) exemplifies that possibility.

VI. Standard Welfare Analysis as a Limiting Case

Several of the examples in the preceding sections suggest that, for settings in which departures from standard choice axioms are minor, one can approximate the appropriate welfare criterion by ignoring choice anomalies and applying the standard normative framework. We now establish that point with generality. Our analysis requires some technical machinery. First we add a mild assumption concerning the choice domain:

Assumption 3. $\mathbb{X}$ (the set of potential choice objects) is compact, 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 correspondences C^n, $n = 1, 2, \ldots$, defined on $\mathcal{G}$. Also consider a choice correspondence $\widehat{C}$ 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 $\widehat{C}(\text{clos}(X))$.[37]

Note that we allow for the possibility that the set X is not compact. In that case, our definition of convergence implies that

37. Technically, this involves uniform convergence in the upper Hausdorff hemimetric; see Appendix C.

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 $C^n(X) = [1 - 1/n, 1)$, whereas $\widehat{C}(\text{clos}(X)) = \{1\}$. The convergence of $C^n(X)$ to $\widehat{C}(\text{clos}(X))$ is intuitive: for a given n, the individual satisfices, but as n increases, he chooses 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 8. Suppose 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 $\widehat{L}^*(u(x^0) - \varepsilon) \subseteq L^n(x^0)$.

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 $\widehat{L}^*(u(x^0))$.

Our next result establishes that, under innocuous assumptions concerning $X(\alpha, m)$ and u, the CV-A and the CV-B converge generally to the standard compensating variation.

THEOREM 9. Suppose 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 $X(\alpha, m)$ is compact for all α and m, and continuous in m.[38]

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

Also assume $\max_{x \in X(\alpha,m)} u(x)$ is weakly increasing in m for all α, and strictly increasing if $\widehat{C}(X(\alpha, m)) \subset \text{int}(\mathbb{X})$. Consider a change from (α_0, d_0) to (α_1, d_1). Let $\widehat{m}$ be the standard compensating variation given $\widehat{C}$, and suppose $\widehat{C}(X(\alpha_1, \widehat{m})) \subset \text{int}(\mathbb{X})$.[39] Let m_A^n be the CV-A, and m_B^n the CV-B given C^n. Then $\lim_{n\to\infty} m_A^n = \lim_{n\to\infty} m_B^n = \widehat{m}$.

Our final convergence result establishes that generalized Pareto optima converge to standard Pareto optima.[40] The statement of the theorem requires the following notation: for any domain $\mathcal{G}$, choice set X, and collection of choice correspondences (one for each individual) $C_1, \ldots, C_N$ defined on $\mathcal{G}$, let $W(X; C_1, \ldots, C_N, \mathcal{G})$ denote the set of weak generalized Pareto optima within X. (When ancillary conditions are absent, we engage in a slight abuse of notation by writing the set of weak Pareto optima as $W(X; C_1, \ldots, C_N, \mathcal{X})$).

THEOREM 10. Consider any sequence of choice correspondence profiles, $(C_1^n, \ldots, C_N^n)$, 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 $x^n \in W(X; C_1^n, \ldots, C_N^n, \mathcal{G})$, all limit points of convergent subsequences lie in $W(\text{clos}(X), \widehat{C}_1, \ldots, \widehat{C}_N, \mathcal{X}^c)$.

Theorem 10 has an immediate corollary for a single decision maker:

COROLLARY 3. Suppose 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 convergent subsequences maximize u in $\text{clos}(X)$.

Theorems 8, 9, and 10 are important for three reasons. First, they justify the common view that the standard welfare framework must be approximately correct when behavioral anomalies

39. This statement assumes that $\widehat{m}$ is well defined. Without further restrictions, there is no guarantee that any finite payment will compensate for the change from α_0 to α_1.

40. It follows from Theorem 10 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.

are small. A formal justification for that view has been absent. 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. Unless we have established the correct criteria for such settings, we have no benchmark against which to gauge the performance of the standard criterion, even when choice anomalies are tiny. Our framework overcomes that problem by providing welfare criteria for all situations. Our results imply that small choice anomalies have only minor implications for welfare. Thus, we have formalized the intuition that a little bit of positive falsification is unimportant from a *normative* perspective.

Second, our convergence results imply that the debate over the significance of choice anomalies need not be resolved prior to adopting a framework for welfare analysis. If our framework is adopted and the anomalies ultimately prove to be small, one will obtain virtually the same answer as with the standard framework.

Third, our convergence results suggest that our welfare criterion will always be reasonably discerning provided behavioral anomalies are not too large. That observation 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.

VII. Refining the Welfare Relations

It is straightforward to verify that R' and P^* become weakly finer as the welfare-relevant domain ($\mathcal{G}$) shrinks and weakly coarser as it expands. Intuitively, if choices between two alternatives x and y are unambiguous over some domain, they are also unambiguous over a smaller domain.[41] Consequently, if one is concerned that R' and P^* are insufficiently discerning, one can potentially refine those relations by excluding GCSs from the welfare-relevant domain. Justifying such refinements generally requires one to officiate between apparent choice conflicts. Many existing discussions of behavioral welfare economics amount to informal arguments concerning officiation; for example, one choice is sometimes taken to be more indicative of "true preferences" than another. Our framework permits one to introduce and formalize such arguments within the context of identifying $\mathcal{G}$.

41. Notice, however, that the same principle does not hold for P' or R^*. If we have $xI'y$ for some domain $\mathcal{G}$, we might nevertheless have $xP'y$ for a more inclusive domain, $\mathcal{G}'$. Similarly, if we have xP^*y (so that $\sim yR^*x$) for some domain $\mathcal{G}$, we might nevertheless have yR^*x for a more inclusive domain, $\mathcal{G}'$.

For a choice-based normative framework, it is natural to consider the possibility of self-officiating through metachoices (that is, choices between choices). The case of time inconsistency illustrates some of the conceptual problems with that approach. Assume an individual would choose x over y for time t at time t, but would choose y over x for time t at time $t-1$. Any metachoice between those choices must occur at time $t-1$ or earlier. Therefore, just like the decision at $t-1$, all metachoices are made at arm's length from the reward. But an arm's-length choice clearly cannot objectively resolve whether another arm's-length choice (the one at time $t-1$) or an in-the-moment choice (the one at time t) is a more appropriate normative guide.

More generally, a metachoice is simply another GCS. Within our framework, consideration of metachoices therefore amounts to expanding the welfare-relevant domain $\mathcal{G}$, which makes the relations R' and P^* weakly coarser, potentially enlarging (and never shrinking) the set of weak individual welfare optima.[42] The welfare relations can become finer only if we also exclude the "defeated" GCS, which would implicitly require us to elevate the status of one type of choice (the metachoice) over another (the original choice). But that elevated status necessarily reflects an arbitrary judgment. We might seek a choice-based justification for that judgment by considering a second-level metachoice (between the original metachoice and the excluded GCS), but that path leads inevitably to consideration of higher and higher level metachoices, with no logical stopping point. In addition, unless one is willing to impose additional structure, there is no guarantee that metachoices will be decisive; for example, they may be cyclic, or kth-level metachoices may conflict with $(k+1)$th-level metachoices for all k. Thus, it is hard to imagine a compelling choice-based justification for deference to metachoices.

Can we devise other compelling criteria for excluding GCSs from the welfare-relevant domain, $\mathcal{G}$? The remainder of this section discusses several alternatives.

VII.A. Refinements Based on Imperfect Information Processing

Suppose there is some GCS, $G=(X,d)$, in which the individual incorrectly perceives the constraint set as $Y \neq X$. We submit that it is appropriate to delete that GCSs from the

42. Expanding $\mathcal{G}$ can shrink the set of *strict* individual welfare optima for a constraint set X, but only if there are two optimal elements of X such that the individual is never willing to choose one but not the other when both are available.

welfare-relevant domain $\mathcal{G}$.[43] Even with its deletion, ambiguities in R' and P^* may remain, but those relations nevertheless become (weakly) finer and hence more discerning.

Why would the individual believe himself to be choosing from the wrong set? 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, or he may have learned from his past experiences more slowly than the objective information would permit. Accordingly, we propose using nonchoice evidence, including findings from psychology, neuroscience, and neuroeconomics, to identify and delete *suspect* GCSs in which those types of informational processing failures occur.[44]

The following simple example motivates the use of evidence from neuroscience.[45] An individual is offered a choice between alternatives x and y. He chooses x when the alternatives are described verbally, and y when they are described partly verbally and partly in writing. Which choice is the best guide for public policy? If we learn that the information was provided in a dark room, we would be inclined to respect the choice of x, rather than the choice of y. We would reach the same conclusion if an opthamologist certified that the individual was blind, or, more interestingly, if a brain scan revealed that his visual processing circuitry was impaired. In all these cases, nonchoice evidence sheds light on the likelihood that the individual successfully processed information that was in principle available to him, thereby properly identifying the choice set X.

Our work on addiction (Bernheim and Rangel 2004) illustrates this agenda. Citing evidence from neuroscience, we argue as follows. First, the brain's value forecasting circuitry includes a specific neural system that measures empirical correlations between cues and potential rewards. Second, the repeated use of an

43. In principle, if we understood the individual's cognitive processes sufficiently well, we might be able to identify his perceived choice set Y, and reinterpret the choice as pertaining to Y rather than to X. While it may be possible to accomplish that task in some instances (see, e.g., Koszegi and Rabin [2008b]), we suspect that, in most cases, it is beyond the current capabilities of economics, neuroscience, and psychology.

44. Thus, our analysis speaks to the current debate over the role of nonchoice evidence in economic analysis. See Gul and Pesendorfer (2008), as well as various other papers appearing in Caplin and Schotter (2008).

45. 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.

addictive substance causes that system to malfunction in the presence of cues that are associated with its use. Whether or not that system *also* plays a role in hedonic experience, the choices made in the presence of those cues are therefore predicated on improperly processed information, and welfare evaluations should be guided by choices made under other conditions (e.g., precommitments).

In many situations, simpler forms of evidence may suffice. For example, if an individual characterizes a choice as a mistake on the grounds that he neglected or misunderstood information, or if a simple test of his knowledge reveals that he ignored critical information, then one might justifiably declare the choice suspect. Other considerations, such as the complexity of a GCS, could also come into play.

Even in the absence of hard evidence, reasonable people may tend to agree that certain GCSs are not conducive to full and accurate information processing. We propose classifying such GCSs as *provisionally suspect*, and proceeding as described above. Anyone who questions a provisional classification can examine the sensitivity of welfare statements to the inclusion or exclusion of the pertinent GCSs. Moreover, any serious disagreement concerning the classification of a particular GCS could in principle be resolved through a narrow and disciplined examination of evidence pertaining to information processing failures.

Note that this refinement agenda entails only a mild modification of our choice-based perspective on welfare. Significantly, we do not propose the use of any information as either a substitute for or alternative to choice patterns. Within this framework, all evaluations ultimately respect at least some of the individual's choices, and must be consistent with all unambiguous choice patterns.

What Is a Mistake? The concept of a *mistake* does not exist within the context of standard choice-theoretic welfare economics. Within our framework, one can define a mistake as a choice made in a suspect GCS that is contradicted by choices in nonsuspect GCSs. According to that definition, the individual's mistake lies in his understanding of his constraint set, not in the choice he makes given that understanding.

In Bernheim and Rangel (2004), we mentioned the example of American visitors to the United Kingdom, who 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. We naturally sense that the pedestrian is not attending to pertinent information and/or options, and that his inattention leads to consequences that he would otherwise wish to avoid. Accordingly, we classify the associated GCS as provisionally suspect on the grounds that the behavior is probably mistaken (in the sense defined above), and instead examine choice situations for which the pedestrian noticeably attends to traffic patterns.

Paternalism. In extreme cases, all or most of an individual's potential GCSs may be suspect, in which case choice provides an insufficient basis for welfare analysis. Possible examples include people suffering from Alzheimer's disease, other forms of dementia, or severe injuries to some of the brain's information-processing circuitry. Likewise, we might regard decisions by young children as inherently suspect. Thus, our framework carves out a role for paternalism. It also suggests a strategy for formulating paternalistic judgments: construct the welfare relations after replacing deleted choices with proxies. Such proxies might be derived from the behavior of decision makers whose decision processes are not suspect, but who are otherwise similar (e.g., with respect to their choices for any nonsuspect GCSs that they have in common, and/or their hedonic responses to specific consequences). For individuals who experience episodes that simultaneously involve both abnormal hedonic responses and impaired decision-making circuitry (e.g., unpleasant and psychologically paralyzing anxiety attacks), it would not be appropriate to substitute the choices of a functional decision maker with normal hedonic responses. Instead, one could construct choice proxies by modeling the relationship between choices and hedonic responses for an individual with functional decision-making circuitry and predicting the choices he would make if he had the *same* abnormal hedonic responses.

VII.B. Refinements Based on Coherence

In some instances, it may be possible to partition behavior into coherent patterns and isolated anomalies. One might then adopt the position that welfare analysis should ignore the anomalies entirely. That argument suggests another potential refinement strategy: identify subsets of GCSs within which choices are coherent (in the sense that standard axioms hold); then construct welfare relations based on those GCSs and ignore other choices. The main difficulty with this *coherence criterion* is that all behavior is

coherent within a sufficiently narrow scope (e.g., every choice is coherent taken by itself). How does one judge whether that scope is too narrow? Despite our inability to offer a general and precise definition, there are nevertheless contexts in which coherence has a natural interpretation.

Take the problem of intertemporal consumption allocation for a β, δ consumer (Section III.E). Consider a *single-choice GCS* (in which the decision is completely resolved through full precommitment at a single point in time) that conflicts with a *staged-choice GCS* (in which it is made in a series of steps). Much of the extant literature adopts the position that a decision for the single-choice GCS reflects a single coherent perspective whereas a decision for the staged-choice GCS does not. That view invites an application of the coherence criterion: exclude staged-choice GCSs from the welfare-relevant domain, denoted $\mathcal{G}_c$, while retaining single-choice GCSs, which cohere within time-indexed subsets. We will explore the implications of that refinement to illustrate the potential power of the coherence criterion.

Because the proposed refinement does not officiate between conflicting single-choice GCSs, it fails to resolve all ambiguity. Nevertheless, within our framework, it yields a discerning welfare criterion. Our next result characterizes individual welfare optima under the resulting welfare relations, R_c' and P_c^*, for conventional intertemporal budget constraints.[46] Define $\lambda \equiv 1/(1+r)$, where r is the rate of interest. Also, define the constraint set X_1 as consisting of all non-negative consumption vectors $(c_1, \ldots, c_T)$ for which $\sum_{k=1}^{T} \lambda^{k-1} c_t \leq w_1$. Assume that initial wealth, w_1, is strictly positive, and that $u(c)$ is continuous and strictly concave.

THEOREM 11. Based on R_c' and P_c^*, the consumption vector C_1^* is an individual welfare optimum in X_1 (both weak and strict) iff C_1^* maximizes $U_1(C_1)$.

According to Theorem 11, welfare optimality within X_1 under R^c is completely governed by the individual's perspective at the first moment in time.[47] The special status of $t = 1$, which we noted following Theorem 4, is amplified when attention is restricted to

46. The characterization holds for any convex constraint set satisfying free disposal that permits continuous trade-offs between consumption in disparate periods. For weak optima, one can dispense with convexity.

47. Once period t arrives, an optimal path remains optimal within the set of paths that remain feasible, but there are other individual welfare optima within that set, and they need not maximize U_1 either overall or within the set of paths that remain feasible. See Bernheim and Rangel (2008a) for details.

$\mathcal{G}_c$. Thus, even though the coherence criterion does not resolve all choice conflicts, it justifies the judgements embedded in long-run criterion (exponential discounting at the rate δ) for certain environments, assuming the first period is short.

The coherence criterion is also equivalent to a novel and appealing variant of multiself Pareto optimality. As conventionally applied, that concept suffers from a conceptual deficiency: it assumes the time t self does not care about the past (see, e.g., Laibson, Repetto, and Tobacman [1998]).[48] Because one cannot choose past consumption, that assumption (as well as any other specific alternative) is arguably untestable and unwarranted. Given our ignorance concerning backward looking preferences, it is more appropriate to adopt a notion of multiself Pareto efficiency that is robust with respect to a wider range of possibilities.

Imagine that if the individual could choose both past and future consumption in period t, he would maximize the decision-utility function $\widehat{U}_t(C_1, \Gamma_t) = U(C_t) + \Gamma_t(c_1, \ldots, c_{t-1})$, where $U(C_t)$ is the objective function for the β, δ setting (equation (4)); $\widehat{U}_t(C_1, \Gamma_t)$ appends the backward-looking function Γ_t. We say that C_1 is a *weak robust multiself Pareto optimum* iff it is a weak multiself Pareto optimum for all possible $(\Gamma_2, \ldots, \Gamma_T)$.[49]

THEOREM 12. For any set X, a consumption vector C_1 is a weak individual welfare optimum (based P_c^*) iff it is a weak robust multiself Pareto optimum.[50]

Intuitively, if the welfare-relevant domain were rectangular, P_c^* would coincide with the strict multiself Pareto relation (Theorem 3). We can make it rectangular by hypothetically extending the choice correspondence C to include choices involving past consumption. Deleting those hypothetical choices makes the welfare relation more discerning and does not enlarge the set of weak individual welfare optima. Thus, the set of weak individual welfare optima under P_c^* must lie within the set of multiself Pareto optima for every conceivable pattern of backward-looking choices. In light of Theorem 12, Theorem 11 is also intuitive: The time $t = 1$ perspective dominates robust multiself Pareto

48. Other assumptions concerning backward-looking preferences appear in the literature; see, for example, Imrohoroglu, Imrohoroglu, and Joines (2003).

49. We omit Γ_1 because there is no consumption prior to period 1.

50. The proof establishes a stronger property: P_c^* is equivalent to a strict robust multiself Pareto dominance relation. One can also show that R_c' is equivalent to a weak robust multiself Pareto dominance relation.

comparisons because we lack critical information (backward-looking preferences, Γ_t) concerning all other perspectives.

Theorems 11 and 12 also explain why the multiself perspective justifies using $U_1(C_1)$ to evaluate the welfare of a *time-consistent* decision maker. The appropriateness of that standard is not obvious, because time-consistent behavior does not rule out divergences between $U_1(C_1)$ and backward-looking preferences at any time $t > 1$. However, if we allow for such divergences, acknowledge that we cannot shed light on them through choice experiments, and invoke the robust multiself Pareto criterion, we are led back to $U_1(C_1)$.

VII.C. Refinements Based on Other Criteria

If people process information more completely and accurately when making straightforward choices, a *simplicity criterion* could have merit. That criterion would presumably favor one-shot binary decision problems. Unfortunately, if we construct P^* exclusively from data on binary decisions, acyclicity is not guaranteed (recall Example 1). However, in certain settings, this procedure does generate coherent welfare relations. Consider again the β, δ model of quasihyperbolic discounting. Fixing the date of choice at time t, behavior within the set of one-shot binary decision problems fully "reveals" the decision-utility function U_t, as does behavior within the set of single-choice GCSs. Therefore, officiating in favor of one-shot binary decision problems is equivalent to officiating in favor of single-choice GCSs; both approaches lead to the welfare relations R'_c and P^*_c.

One could also apply a *preponderance criterion*: if someone ordinarily chooses x over y and rarely chooses y over x, disregard the exceptions and follow the rule. That criterion is sometimes invoked (at least implicitly) in the literature on quasi-hyperbolic (β, δ) discounting to justify use of the long-run perspective: trade-offs between rewards in periods t and $t + k$ are governed only by δ from the perspective of all periods $s < t$, and by both β and δ only from the perspective of period t.

We see two conceptual problems with the preponderance criterion. First, there are potentially many competing notions of frequency. Because it is possible to proliferate variations of ancillary conditions, one cannot simply count GCSs. In the quasi-hyperbolic setting, a count of time-dated perspectives would favor the long-run criterion. However, an application of preponderance based

on the frequency with which GCSs are encountered (an index of familiarity) might 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 quasi-hyperbolic setting, one could argue that people may appreciate their needs most accurately when those needs are immediate and concrete, rather than distant and abstract.

VIII. Conclusion

In this paper, we have proposed a choice-theoretic framework for behavioral welfare economics, one that accommodates choice conflicts and other nonstandard behavioral patterns without requiring economists to take a stand on whether individuals have true utility functions or on how well-being might be measured. Our approach exploits coherent aspects of choice by replacing the standard revealed preference relation with an *unambiguous choice* relation. That relation is always acyclic, and therefore suitable for rigorous welfare analysis. It is also the most discerning welfare criterion that never overrules choice.

Like standard welfare economics, our framework requires only information concerning the mapping from environments to choices. Because it encompasses any theory that generates a choice correspondence, it is applicable irrespective of the processes generating behavior, or of the positive model used to describe behavior. Thus, it potentially opens the door to greater integration of economics, psychology, and neuroeconomics. It generalizes standard choice-based welfare economics in two senses. First, the approaches are equivalent when standard choice axioms are satisfied. Second, for settings in which departures from those axioms are minor, our framework implies that one can approximate the appropriate welfare criterion by ignoring choice anomalies entirely. It generates natural counterparts for the standard tools of applied welfare analysis, such as compensating variation, consumer surplus, Pareto optimality, and the contract curve, and permits a broad generalization of the of the first welfare theorem. It is easily applied in the context of specific positive theories; indeed, elements have been incorporated into recent work by Chetty, Looney, and Kroft (2008) and Burghart, Cameron, and

Gerdes (2007). Finally, though not universally discerning, it lends itself to principled refinements.

APPENDIX

This Appendix is divided into three sections. The first contains proofs of miscellaneous theorems (Theorems 1, 2, 3, 5, 6, and 7). The second pertains to the β, δ model (Theorems 4, 11, and 12), and the third to convergence properties (Theorems 8, 9, and 10).

A. Proofs of Miscellaneous Theorems

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

Proof of Theorem 2. 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 $y \in C(X, d)$. Because Q is an inclusive libertarian relation, we must then have $y \in m_Q(X)$. But because $x \in X$, 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. QED

Proof of Theorem 3. 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.

The final statement concerning optima follows immediately from the equivalence of the binary relations. QED

Proof of Theorem 5. To calculate the CV-A, we must find the infimum of the values of m that satisfy

$$U(M - p_1z_1 + m', z_1 \mid d) > U(M - p_0z_0, z_0 \mid d) \quad \text{for all } m' \geq m \quad \text{and} \quad d \in [d_L, d_H].$$

Notice that this requires

$$m \geq [p_1z_1 - p_0z_0] + d[v(z_0) - v(z_1)] \quad \text{for all } d \in [d_L, d_H].$$

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

$$\begin{aligned} m^A &= [p_1z_1 - p_0z_0] + d_H[v(z_0) - v(z_1)] \\ &= [p_1z_1 - p_0z_0] + \int_{z_1}^{z_0} d_H v'(z)\, dz \\ &= [p_1 - p_0]z_1 + p_0z_1 - p_0[z_0 - z_1] - p_0z_1 + \int_{z_1}^{z_0} d_H v'(z)\, dz \\ &= [p_1 - p_0]z_1 + \int_{z_1}^{z_0} [d_H v'(z) - p_0]dz. \end{aligned}$$

The derivation of the expression for m^B is analogous. QED

Proof of Theorem 6. Consider the following set:

$$U^*(x, X) = \{y \in X \mid \forall i,\ \sim xP_i^*y \text{ and } \nexists M \geq 1 \text{ and } a_1, \ldots, a_M \text{ s.t. } xP_i^*a_1P_i^*a_2 \ldots a_MP_i^*y\}.$$

Because P_i^* is acyclic, $U^*(x, X)$ contains x and is therefore nonempty. It is also apparent that $U^*(x, X) \subseteq \{y \in X \mid \forall i,\ \sim xP_i^*y\}$. 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 exist $a_1, \ldots, a_N$ s.t. $xP_k^*a_1P_k^*a_2 \ldots a_NP_k^*wP_k^*z$. But that implies $z \notin U^*(x, X)$, a contradiction.

Now we prove the theorem. Take any individual i. Choose any $z \in C_i(U^*(x, X), d)$ for some d with $\big(U^*(x, X), d\big) \in \mathcal{G}$. 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 because $w, z \in U^*(x, X)$ and wP_i^*z, we have $z \notin C_i(U^*(x, X), d)$, a contradiction. QED

Proof of Theorem 7. 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

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

Moreover, because $\widehat{w}$ is feasible, we know that $\sum_{n=1}^{N} (\widehat{w}^n - z^n) \in Y$, or equivalently that there exists $v = (v^1, \ldots, v^F)$ 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. Because $v^f \in Y^f$, this contradicts the assumption that $\widehat{y}^f$ maximizes firm f's profits given $\widehat{\pi}$. QED

B. 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, $V_t(C_t) \geq U_t(C_t) \geq W_t(C_t)$, where the first inequality is strict if $c_k > 0$ for some $k > t$, and the second inequality is strict if $c_k > 0$ for some $k > t+1$.

Suppose the individual faces the GCS (X, τ). Because the individual is dynamically consistent within each period, we can without loss of generality collapse multiple decisions within any single period into a single decision. So a lifetime decision involves a sequence of choices, $r_1, \ldots, r_T$ (some of which may be degenerate), that generate a sequence of consumption levels, $c_1, \ldots, c_T$. 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 $r_1, \ldots, r_t$ leads to a continuation problem $G^C(r_1, \ldots, r_t)$, which resolves any residual discretion in $r_{t+1}, \ldots, r_T$.

With these observation in mind, we establish three lemmas.

LEMMA 1. Suppose that, as of some period t, the individual has chosen $r_1, \ldots, r_{t-1}$ and consumed $c_1^A, \ldots, c_{t-1}^A$, and that C_t^A remains feasible for $G^C(r_1, \ldots, r_{t-1})$. 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

$$
\begin{aligned}
U_t\big(c_t^A, \widehat{C}_{t+1}\big) &= u\big(c_t^A\big) + \beta\delta V_{t+1}\big(\widehat{C}_{t+1}\big) \\
&\geq u\big(c_t^A\big) + \beta\delta W_{t+1}\big(C_{t+1}^A\big) \\
&= W_t\big(C_t^A\big).
\end{aligned} \tag{6}
$$

Because the individual is willing to make a decision at time t that leads to the continuation consumption trajectory C_t^B, and because 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 $V_t(C_t^B) \geq U_t(C_t^B)$. QED

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_1^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_1^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

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

and $C_2^\varepsilon = (c_2^\varepsilon, 0, \ldots, 0)$. (Existence of c_2^ε is guaranteed because $W_2(C_2^A) + \varepsilon$ is strictly positive, and u^{-1} is defined on the nonnegative reals.) 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

$$\begin{aligned} U_1(c_1^A, C_2^\varepsilon) &= u(c_1^A) + \beta\delta u(c_2^\varepsilon) \\ &= u(c_1^A) + \beta\delta\big[W_2(C_2^A) + \varepsilon\big] \\ &= W_1(C_1^A) + \beta\delta\varepsilon. \end{aligned}$$

If $U_1(C_1^B) = W_1(C_1^A)$, we set $\varepsilon = 0$. The individual is indifferent with respect to his period 1 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. QED

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 $r_1^A, \ldots, r_T^A$ that leads to the outcome $c_1^A, \ldots, c_T^A$. 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 $r_1^A, \ldots, r_t^A$ and consuming $c_1^A, \ldots, c_t^A$. (Note that $\widehat{C}_1 = C_1^B$.) According to expression (6), $U_t(c_t^A, \widehat{C}_{t+1}) \geq W_t(C_t^A)$. 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 us 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 $r_1^A, \ldots, r_{t-1}^A$), 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

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

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 history of choices of the form $r_1^A, \ldots, r_{t-1}^A$, 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 $r_1^A, \ldots, r_{t-1}^A$, all of the predecessor nodes correspond to histories of the form $r_1^A, \ldots, r_k^A$, 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 $r_1^A, \ldots, r_{t-1}^A$, in each case accounting for changes made at successor nodes (those corresponding to larger t).

After any history $r_1^A, \ldots, r_{t-1}^A$, 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 $U_t(C_t^A)$. Because 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. QED

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

Step 1. $C_1' R' C_1''$ 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 $C_1' R' C_1''$. 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 $C_1' R' C_1''$.

Step 2. $C_1' P^* C_1''$ 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 $C_1' P^* C_1''$.

Step 3. R' and P^* are transitive.

First consider R'. Suppose that $C_1^1 R' C_1^2 R' C_1^3$. 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 $C_1^1 R' C_1^3$.

Next consider P^*. Suppose that $C_1^1 P^* C_1^2 P^* C_1^3$. 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 $C_1^1 P^* C_1^3$. QED

Proof of Theorem 11. 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. Then $\mathcal{G}_c = \mathcal{G}_1 \cup \mathcal{G}_2 \cup \cdots \cup \mathcal{G}_T$. 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).

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 $C(G) = \{C_1^*\}$ (uniqueness of the choice follows from strict concavity of u). It follows that $\sim C_1 P' C_1^*$ 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 $c_1'' \neq \widehat{c}_1$. (If $c_1' \neq \widehat{c}_1$, then $C_1'' = C_1'$. If $c_1' = \widehat{c}_1$, 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 X that contains the options $\widehat{C}_1$ and C_1''. Notice that $(X, 1) \in \mathcal{G}_1$; moreover, $(X, t) \notin \mathcal{G}_t$ for all $t > 1$, because a choice from X resolves some discretion at time $t = 1$. But because $U_1(C_1'') > U_1(\widehat{C}_1)$, the individual will not select $\widehat{C}_1$ from $(X, 1)$. 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). QED

Proof of Theorem 12. We begin by defining the strict robust multiself Pareto relation, M_R^*:

$C_1' M_R^* C_1''$ iff there exists $(\Gamma_2, \ldots, \Gamma_T)$ such that $U_1(C_1') > U_1(C_1'')$ and $\widehat{U}_t(C_1', \Gamma_t) > \widehat{U}_t(C_1', \Gamma_t)$ for all $t = 2, \ldots, T$.

Note that $C_1' M_R^* C_1''$ implies $\sim C_1'' M_R^* C_1'$. For any constraint set X, the set of weak robust multiself Pareto optimal clearly coincides with the set of maximal elements under M_R^*. Thus, we prove the theorem by demonstrating that P^* and M_R^* are equivalent.

First suppose $C_1' P_c^* C_1''$. Let k denote the earliest period in which C_1' and C_1'' differ. Because C_1' is strictly chosen over C_1'' in periods 1 through k, we must have

$$u(c_k') + \beta \sum_{t=k+1}^{T} \delta^{t-k} u(c_t') > u(c_k'') + \beta \sum_{t=k+1}^{T} \delta^{t-k} u(c_t'')$$

and, if $k > 1$,

$$\sum_{t=k}^{T} \delta^{t-k} u(c_t') > \sum_{t=k}^{T} \delta^{t-k} u(c_t'').$$

Now choose arbitrary functions $\Gamma_1, \ldots, \Gamma_k$, and for $s > k$ choose Γ_s such that

$$\begin{aligned} &\Gamma_s(c_1', \ldots, c_{s-1}') - \Gamma_s(c_1'', \ldots, c_{s-1}'') \\ &\quad > \left[u(c_s') + \beta \sum_{t=s+1}^{T} \delta^{t-s} u(c_t') \right] - \left[u(c_s'') + \beta \sum_{t=s+1}^{T} \delta^{t-s} u(c_t'') \right]. \end{aligned}$$

Then we necessarily have $U_1(C_1') > U_1(C_1'')$, and $\widehat{U}_t(C_1', \Gamma_t) > \widehat{U}_t(C_1'', \Gamma_t)$ for $t = 2, \ldots, T$, from which it follows that $C_1' M_R^* C_1''$.

Now suppose $\sim C_1' P_c^* C_1''$. Again let k denote the earliest period in which C_1' and C_1'' differ. Because C_1' is not strictly chosen over C_1'' in all periods 1 through k, we must have either

$$u(c_k') + \beta \sum_{t=k+1}^{T} \delta^{t-k} u(c_t') \leq u(c_k'') + \beta \sum_{t=k+1}^{T} \delta^{t-k} u(c_t'')$$

or (in the case of $k > 1$ only)

$$\sum_{t=k}^{T} \delta^{t-k} u(c_t') \leq \sum_{t=k}^{T} \delta^{t-k} u(c_t'').$$

If either $k = 1$ and the first inequality holds, or $k > 1$ and the second inequality holds, then $U_1(C_1') \leq U_1(C_1'')$, which implies $\sim C_1' M_R^* C_1''$. If $k > 1$ and the first inequality holds, then $\widehat{U}_k(C_1', \Gamma_k) \leq \widehat{U}_k(C_1'', \Gamma_k)$ for all Γ_k, so again $\sim C_1' M_R^* C_1''$. QED

C. 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 of 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

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

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 correspondences C^n defined on $\mathcal{G}$. Also consider a choice correspondence $\widehat{C}$ 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 $\delta_U(\widehat{C}(\text{clos}(X)), C^n(X, d)) < \varepsilon$.

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

Now we show that if $n > N'$, then for all generalized choice situations 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 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 $N_{r'}(\widehat{C}(\text{clos}(X_1)))$ 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 $\widehat{L}^*(u_2)$.

Because 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 $x \in \widehat{L}^*(u_2)$. QED

Proof of Theorem 8. The proof proceeds in two 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 $x \in \widehat{U}^*(u_1)$. QED

In the statement of Theorem 9, 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 (see footnote 26).

Proof of Theorem 9. 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 under the conditions stated in the theorem. That is, $\widehat{m}_A = \widehat{m}_B = \widehat{m}$; the infimum (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 $y \in \widehat{C}(X(\alpha_1, m))$. (Because $\widehat{C}(X(\alpha, \widehat{m})) \subset \text{int}(\mathbb{X})$, we know that $\arg\max_{z \in X(a,m)} u(z)$ is strictly increasing in m at $m = \widehat{m}$, so such an m necessarily exists.) We claim that there exists N_1 such that for $n > N_1$ and $m' \geq m$, we have $yP^{n*}x$ for all $x \in C^n(X(\alpha_0, 0), d_0)$ and $y \in C^n(X(\alpha_1, m), d_1(m))$. (It follows that m_A^n exists for $n > N_1$.)

Define $u_1 = (1/3)u(w) + (2/3)u(z)$ and $u_2 = (2/3)u(w) + (1/3)u(z)$ for $w \in \widehat{C}(X(\alpha_0, 0))$ and $z \in \widehat{C}(X(\alpha_1, m))$. Because $u_1 > u_2$, Lemma 4 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 because u is continuous (and therefore uniformly continuous on the compact set $\mathbb{X}$), there exists $r_1 > 0$ such

that $N_{r_1}(\widehat{C}(X(\alpha_0, 0))) \subset \widehat{L}^*(u_2)$, and $N_{r_1}(\widehat{C}(X(\alpha_1, m'))) \subset \widehat{U}^*(u_1)$ for all $m \geq m'$. Moreover, there exists N_1'' such that for $n > N_1''$, we have $C^n(X(\alpha_0, 0), d_0) \subset N_{r_1}(\widehat{C}(X(\alpha_0, 0)))$ and $C^n(X(\alpha_1, m'), d_1(m')) \subset N_{r_1}(\widehat{C}(X(\alpha_1, m')))$ for all $m' \geq m$. Consequently, for $n > N_1''$, we have $C^n(X(\alpha_0, 0), d_0) \subset \widehat{L}^*(u_2)$ and $C^n(X(\alpha_1, m'), d_1(m')) \subset \widehat{U}^*(u_1)$ for all $m' \geq m$. It follows that, for $n > N_1 = \max\{N_1', N_1''\}$ and $m \geq m'$, 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 $u_1 = (1/3)u(w) + (2/3)u(z)$ and $u_2 = (2/3)u(w) + (1/3)u(z)$ for $z \in \widehat{C}(X(\alpha_0, 0))$ and $w \in \widehat{C}(X(\alpha_1, m))$. Because $u_1 > u_2$, Lemma 4 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 because u is continuous, there exists $r_2 > 0$ such that $N_{r_2}(\widehat{C}(X(\alpha_0, 0))) \subset \widehat{U}^*(u_1)$, and $N_{r_2}(\widehat{C}(X(\alpha_1, m))) \subset \widehat{L}^*(u_2)$. Moreover, there exists N_2'' such that for $n > N_2''$, we have $C^n(X(\alpha_0, 0), d_0) \subset N_{r_2}(\widehat{C}(X(\alpha_0, 0)))$ and $C^n(X(\alpha_1, m), d_1(m)) \subset N_{r_2}(\widehat{C}(X(\alpha_1, m)))$. 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. Recall from step 1 that m_A^n exists for sufficiently large n. The sequence m_A^n must therefore have at least one limit point $m_A^* \neq \widehat{m}_A$. Suppose first that $m_A^* > \widehat{m}_A$. Consider $m' = (m_A^* + \widehat{m}_A)/2$. Because u satisfies nonsatiation 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. QED

Proof of Theorem 10. Suppose not. Without loss of generality, assume that x^n converges to a point $x^* \notin W(\text{clos}(X), \widehat{C}_1, \ldots, \widehat{C}_N, \mathcal{X}^c)$ (if necessary, 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 8, 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, $x^n \notin W(X; C_1^n, \ldots, C_N^n, \mathcal{G})$, a contradiction. QED

STANFORD UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH
CALIFORNIA INSTITUTE OF TECHNOLOGY AND NATIONAL BUREAU OF ECONOMIC RESEARCH

REFERENCES

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

Arrow, Kenneth J., "Rational Choice Functions and Orderings," *Economics*, 26 (1959), 121–127.

Bernheim, B. Douglas, "Neuroeconomics: A Sober (But Hopeful) Appraisal," *AEJ Microeconomics*, forthcoming, 2009a.

——, "Behavioral Welfare Economics," *Journal of the European Economics Association*, forthcoming, 2009b.

Bernheim, B. Douglas, and Antonio Rangel, "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, 94 (2004), 1558–1590.

——, "Toward Choice-Theoretic Foundations for Behavioral Welfare Economics," *American Economic Review Papers and Proceedings*, 97 (2007), 464–470.

——, "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," NBER Working Paper No. 13737, 2008a.

——, "Choice-Theoretic Foundations for Behavioral Welfare Economics," in *Foundations of Positive and Normative Economics*, Andrew Caplin and Andrew Schotter, eds. (Oxford University Press, forthcoming, 2008b).

Bossert, Walter, Yves Sprumont, and Kotaro Suzumura, "Consistent Rationalizability," *Economica*, 72 (2005), 185–200.

Burghart, Daniel R., Trudy Ann Cameron, and Geoffrey R. Gerdes, "Valuing Publicly Sponsored Research Projects: Risks, Scenario Adjustments, and Inattention," *Journal of Risk and Uncertainty,* 35 (2007), 77–105.

Caplin, Andrew, and John Leahy, "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics*, 116 (2001), 55–79.

Caplin, Andrew, and Andrew Schotter, eds., *Foundations of Positive and Normative Economics* (Oxford University Press, forthcoming, 2008).

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," mimeo, University of California, Berkeley, 2008.

Ehlers, Lars, and Yves Sprumont, "Weakened WARP and Top-Cycle Choice Rules," mimeo, University of Montreal, 2006.

Fon, Vincy, and Yoshihiko Otani, "Classical Welfare Theorems with Non-Transitive and Non-Complete Preferences," *Journal of Economic Theory*, 20 (1979), 409–418.

Gale, David, and Andreu Mas-Colell, "An Equilibrium Existence Theorem for a General Model without Ordered Preferences," *Journal of Mathematical Economics*, 2 (1975), 9–15.

Green, Jerry, and Daniel Hojman, "Choice, Rationality, and Welfare Measurement," mimeo, Harvard University, 2007.

Gul, Faruk, and Wolfgang Pesendorfer, "Random Expected Utility," *Econometrica*, 74(1) (2006), 121–146.

——, "The Case for Mindless Economics," in *Foundations of Positive and Normative Economics*, Andrew Caplin and Andrew Schotter, eds. (Oxford University Press, forthcoming, 2008).

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

Iyengar, S. S., and M. R. Lepper, "Why Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology,* 79 (2000), 995–1006.

Kahneman, Daniel, "Objective Happiness," in *Well-Being: The Foundations of Hedonic Psychology*, Daniel Kahneman, Ed Diener, and Norbert Schwarz, eds. (New York, NY: Russell Sage Foundation, 1999).
Kahneman, D., P. Wakker, and R. Sarin, "Back to Bentham? Explorations of Experienced Utility," *Quarterly Journal of Economics*, 112 (1997), 375–406.
Kalai, Gil, Ariel Rubinstein, and Ran Spiegler, "Rationalizing Choice Functions by Multiple Rationales," *Econometrica*, 70 (2002), 2481–2488.
Koszegi, Botond, and Matthew Rabin, "Choices, Situations, and Happiness," *Journal of Public Economics*, forthcoming, 2008a.
——, "Revealed Mistakes and Revealed Preferences," in *Foundations of Positive and Normative Economics*, Andrew Caplin and Andrew Schotter, eds. (Oxford University Press, forthcoming, 2008b).
Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics,* 112 (1997), 443–477.
Laibson, David, Andrea Repetto, and Jeremy Tobacman, "Self-Control and Saving for Retirement," *Brookings Papers on Economic Activity*, 1 (1998) 91–172.
Mandler, Michael, "Welfare Economics with Status Quo Bias: A Policy Paralysis Problem and Cure," mimeo, University of London, 2006.
Manzini, Paola, and Marco Mariotti, "Rationalizing Boundedly Rational Choice," mimeo, University of London, 2007.
Mas-Colell, Andreu, "An Equilibrium Existence Theorem without Complete or Transitive Preferences," *Journal of Mathematical Economics*, 1 (1974) 237–246.
O'Donoghue, Ted, and Matthew Rabin, "Doing It Now or Later," *American Economic Review*, 89 (1999), 103–124
Rigotti, Luca, and Chris Shannon, "Uncertainty and Risk in Financial Markets," *Econometrica*, 73 (2005), 203–243.
Rubinstein, Ariel, and Yuval Salant, "A Model of Choice from Lists," *Theoretical Economics*, 1 (2006), 3–17.
——, "(*A,f*) Choice with Frames," *Review of Economic Studies*, forthcoming, 2008.
Shafer, Wayne, and Hugo Sonnenschein, "Equilibrium in Abstract Economies without Ordered Preferences," *Journal of Mathematical Economics*, 2 (1975), 345–348.
Sugden, Robert, "The Opportunity Criterion: Consumer Sovereignty without the Assumption of Coherent Preferences," *American Economic Review*, 94 (2004), 1014–1033.
Suzumura, Kotaro, "Remarks on the Theory of Collective Choice," *Economica*, 43 (1976), 381–390.
Thaler, Richard, and Cass R. Sunstein, "Libertarian Paternalism," *American Economic Review Papers and Proceedings*, 93 (2003), 175–179.