AVERTISSEMENT

LIENS

Code la Propriété Intellectuelle – Articles L. 122-4 et L. 335-1 à L. 335-10
Loi n° 92-597 du 1er juillet 1992, publiée au *Journal Officiel* du 2 juillet 1992

http://www.cfcopies.com/V2/leg/leg-droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par l'Université Toulouse Capitole

École doctorale : **Sciences Economiques-Toulouse School of Economics**

Présentée et soutenue par

WIKMAN Peter

le 25 mai 2020

Essays on conventions in games and anticipation-dependent preferences

---

Discipline : **Sciences Economiques**

Unité de recherche : **TSE-R (UMR CNRS 5314 – INRA 1415)**

Directeur de thèse : Jörgen WEIBULL, Professeur, Stockholm School of Economics

JURY

Rapporteurs Klaus RITZBERGER, Professeur, Royal Holloway, University of London
Itzhak GILBOA, Professeur, Tel-Aviv University et HEC Paris

Suffragants Christian GOLLIER, Professeur, Université Toulouse 1 Capitole
Takuro YAMASHITA, Professeur, Université Toulouse 1 Capitole

Abstract

This thesis consists of three chapters. Two chapters fall into the field of game theory and one into the field of decision theory.

In the first chapter, I study strategic interaction when people are familiar with the setting they interact in. In such situations, social conventions often emerge and tend to dictate how people behave. Conventions in which people disregard alternatives outside of the convention not only help people coordinate their interactions but also simplify their decision-making. Motivated by this, I develop a novel game-theoretic concept that captures outcomes that are consistent with the existence of such self-enforcing conventions. The resulting solution concept is operational and allows for decomposing games into smaller self-contained games that can be studied in isolation.

In the second chapter, I ask whether behavior consistent with the just-described conventions can be given evolutionary interpretations. In such interpretations, the convention is the resulting pattern of behavior in a large population of individuals after they have interacted for some time, with their behavior adjusting over time in response to the payoffs that their actions have given in the past. These interpretations differ from the standard justification of solution concepts based on the assumption of rational individuals that have correct expectations about others' behavior. I find that indeed these conventions admit such interpretations, and, moreover, standard notions of evolutionarily stable behavior are often consistent with the adherence to such conventions.

In the last chapter, I develop a model of a decision-maker who evaluates outcomes as gains and losses relative to her recent expectations. The decision-maker forms her expectations of an uncertain future outcome by trading off the joy from anticipating a higher outcome with the risk of being disappointed by the outcome. These expectations are then taken as given when the outcome nears. Moreover, the decision-maker is loss averse in the sense that losses relative to these expectations are felt worse than same-sized gains are felt good. The main result is a complete description of the observable choices that are consistent with this behavior. More specifically, I provide necessary and sufficient conditions on choices in the form of axioms such that it is as-if the decision-maker acts as described by the model.

Résumé

Cette thèse se compose de trois chapitres. Deux chapitres relèvent du domaine de la théorie des jeux et un autre du domaine de la théorie de la décision.

Dans le premier chapitre, j'étudie des interactions stratégiques dans un contexte où les agents connaissent le cadre dans lequel ils interagissent. Souvent dans de telles situations, des conventions sociales émergent et tendent à dicter comment les agents se comportent. Les conventions dans lesquelles les gens ne tiennent pas compte des alternatives hors de convention aident les agents non seulement à coordonner leurs interactions mais aussi à simplifier leur prise de décision. Motivé par cela, je développe un nouveau concept en théorie des jeux qui capture des résultats compatibles avec l'existence de conventions qui s'enforcent par eux-mêmes. Le concept de solution qui en résulte est opérationnel et permet de décomposer des jeux en jeux autonomes plus petits qui peuvent être étudiés isolément.

Dans le deuxième chapitre, je me demande si un comportement conforme aux conventions qui viennent d'être décrites peut recevoir des interprétations évolutionnaires. Dans une telle interprétation, la convention est le modèle qui résulte du comportement d'individus dans une grande population, après qu'ils ont interagi pendant un certain temps, leur comportement s'ajustant au fil du temps en réponse au payment que leurs actions ont généré dans le passé. Ces interprétations diffèrent de la justification standard des concepts de solution basée sur l'hypothèse d'individus rationnels qui ont des attentes correctes sur le comportement des autres. Je prouve dans ce chapitre qu'en effet ces conventions admettent de telles interprétations, et les notions standard de comportement stable sur le plan de l'évolution sont souvent compatibles avec l'adhésion à de telles conventions.

Dans le dernier chapitre, je développe un modèle de décideur qui évalue les résultats comme des gains et des pertes par rapport à ses anticipations récentes. Le décideur forme ses anticipations d'un résultat futur et incertain, en échangeant la joie d'anticiper un résultat plus élevé avec le risque d'être déçu par le résultat. Ces anticipations sont alors considérées comme données lorsque la réalisation du résultat approche. De plus, le décideur est averse aux pertes dans le sens où les pertes par rapport à ces attentes sont ressenties pire que les gains de même taille. Le résultat principal est une description complète des choix observables qui sont cohérents avec ce comportement. Plus précisément, je fournis des conditions nécessaires et suffisantes sur les choix sous forme d'axiomes, comme si le décideur agissait comme décrit par le modèle.

Declaration

Financial support by Uppsala University, the Agence Nationale de la Recherche (ANR-11-IDEX-0002-02), and ERC grant No 714693 are gratefully acknowledged.

Acknowledgements

I would like to thank my advisor Jörgen Weibull for taking the chance on me. This thesis would not have been possible without his support and guidance. I have had a lot of fun discussing various game-theoretical issues with him throughout the last four years. I have also enjoyed working with Klaus Ritzberger and I am grateful for his help with the first chapter of this thesis. I am indebted to Takuro Yamashita for his support, especially towards the last year of my doctorate.

Furthermore, I would like to thank Taisuke Imai and Yves Le Yaouanq for inviting me to visit the Department of Economics at the Ludwig-Maximilians-Universität in Munich. I am also thankful to Todd Sarver for hosting me in the Department of Economics at Duke University.

Much can be said about Toulouse School of Economics, but the PhD student there are absolutely fantastic. I owe thanks to many friends there and elsewhere, especially Torkel Mattesson and Jeremy Roh. I also want to thank my family and in particular my parents.

Table of Contents

Chapter 1: Nash blocks

Abstract

A product set of pure strategies is a Nash block if it contains all best replies to the Nash equilibria of the game in which the players are restricted to the strategies in the block. This defines an intermediate block property, between curb (Basu and Weibull, 1991) and coarse tenability (Myerson and Weibull, 2015). While the new concept is defined without reference to the consideration-set framework that defines tenability, the framework can be used to characterize Nash blocks in terms of potential conventions when large populations of individuals recurrently interact. Although weaker than curb, Nash blocks nevertheless maintain most robustness properties of curb sets.

1.1 Introduction

In Nash's mass-action interpretation (Nash, 1950), individuals are randomly and repeatedly drawn from large populations to play a game against each other. It is assumed that an individual's behavior cannot influence the future behavior of other individuals, implying the absence of supergame effects. A Nash equilibrium is then seen as a stationary state in population frequencies over the individual's pure strategies. In such interactions, conventions for how to play the underlying game often emerge.[1] These conventions not only help the individuals to coordinate but also simplify their decision-making (see, e.g., Schelling (1960) and Lewis (1969)). However, as is well-known, simple examples show that many Nash equilibria are implausible as such conventions. The totally mixed Nash equilibrium in the following coordination game is a striking example[2]

	L	R
L	$1,1$	$0,0$
R	$0,0$	$1,1$

In this game, it seems unreasonable that the individuals would not be able to over time coordinate their expectations to settle on one of the strict equilibria.

In the spirit of Myerson and Weibull (2015), this paper develops a set-valued concept that identifies Nash equilibria that are compatible with potential conventions in finite normal-form games. In such a game, a *block* is a nonempty set of pure strategies for each player, and a *block game*[3] is a game where each player is restricted to using strategies with support in the associated block. Myerson and Weibull (2015) interpret blocks as the basis for potential conventions, that is, sets of strategies individuals take into consideration when called to play the game in their player role. For such conventions to be self-enforcing, no individual should be able to do better by using strategies outside of the convention. This notion can be formalized in several ways. An elegant and operational formulation is due to Basu and Weibull (1991): a block is *closed under rational behavior*, or curb, if it contains all best replies to the mixed-strategy profiles with support in the block.[4]

[1] Here, a convention is best described by the words of Peyton Young: "A convention is a pattern of behavior that is customary, expected, and self-enforcing" (Young, 1993, p.57).

[2] Note that this equilibrium satisfies most refinements in the literature, such as perfection (Selten, 1975), properness (Myerson, 1978), strategic stability (Kohlberg and Mertens, 1986) and essentiality (Wu and Jiang, 1962).

[3] A perhaps more telling name would be normal-form subgame. However, this name is preoccupied by Mailath et al. (1993) who use it to refer to a block that represents a subgame in an extensive form consistent with the normal form.

[4] See Section 7 for a discussion of other related concepts.

As noted by Myerson and Weibull (2015), curb blocks are sometimes large and depend on features of the game that may be regarded as strategically inessential. Therefore, the authors weaken the above robustness requirement to only hold when the *overall population play constitutes a Nash equilibrium*. They formalized this within a framework in which every player role in a game is represented by a large population of boundedly rational individuals. Using this framework, the authors elaborate two block properties called *coarse* and *fine tenability*, respectively. Coarse tenability satisfies the weaker robustness requirement by requiring that the block contains at least one best reply to any population Nash equilibrium. Fine tenability relaxes this requirement by restricting the population distribution of boundedly rational individuals to be biased towards more rationality types. *Settled equilibria* are Nash equilibria with support in minimal coarsely and finely tenable blocks.

In this paper, I explore an intermediate block property, between curb and coarse tenability, called *Nash block*. A Nash block contains all best replies to all Nash equilibria of its block game. Thus, every Nash equilibrium of a Nash-block game is a Nash equilibrium of the full game. A *Nash-block settled equilibrium*, or NBE, is any Nash equilibrium with support in some *minimal* Nash block. Every finite game has at least one such equilibrium. The restriction to minimal blocks is motivated by the following observation. Assume that there exists some population dynamic that lead play to settle in a Nash block. If this block contains a proper subblock that is also a Nash block, it is likely that play over time would settle in it since it has the same external stability properties as the original Nash block. In the above coordination game, it is easy to verify that the two strict Nash equilibria constitutes the two minimal Nash blocks, $\{L\} \times \{L\}$ and $\{R\} \times \{R\}$ (this is also true for the other just-mentioned concepts).

As hinted by its definition, the Nash block concept is closely related to the curb concept. I show that most of the latter's robustness properties are inherited by Nash blocks. In particular, a strategy profile constitutes a singleton Nash block if and only if it is a strict Nash equilibrium. Moreover, every Nash block contains the support of an essential component (Jiang, 1963)—thus also a proper equilibrium (Myerson, 1978) and a strategically stable subset (Kohlberg and Mertens, 1986, Mertens, 1989, 1991). Despite these similarities, the Nash and the curb concept differ on an open set of games.

The Nash block concept is also related to coarse tenability. Although the former has a succinct definition without reference to population play, it can be characterized within Myerson and Weibull's (2015) population framework. This characterization provides a micro foundation for the concept and highlights its relationship with coarse tenability. While the latter requires the no individual does *strictly better* using strategies outside the convention (in equilibrium), the Nash block concept strengthens this condition to require that any individual

does *strictly worse* doing so. I show by way of example that coarsely (and finely) tenable blocks lack some of the robustness properties that Nash blocks inherit from the curb concept.

Moreover, Nash blocks offer a simple way to decompose games into self-contained block games that can be studied in isolation from the full game. For Nash equilibria with support in such blocks, this is without loss of generality in the following sense: global properties—such as robustness against payoff perturbations of an equilibrium component—is determined by local properties of the same component in the corresponding Nash-block game. I show this using index theory (Ritzberger, 1994).

The solution concept developed in this paper offers an operational approach to equilibrium selection that has good cutting power in games in which standard solutions concepts are known to perform poorly. These include games with cheap talk, signaling, or voting, which are typically associated with a plethora of Nash equilibria, many of which constitute implausible predictions.

Finally, I discuss a property that minimal Nash blocks share with minimal tenable blocks (but not curb blocks), namely, that such blocks are *not* invariant under the deletion of strictly dominated strategies. However, in contrast to minimal tenable blocks, minimal Nash blocks are invariant under the addition of such strategies. Therefore, in any given game it is easy to identify the set of such blocks that are invariant under both the removal and addition of strictly dominated strategies.

The rest of the paper is organized as follows. Section 2 contains a motiving example. In Section 3, notation and definitions are provided. The Nash block concept is introduced in Section 4, where some of its fundamental properties are provided and its cutting power in important classes of games is highlighted. Section 5 discusses the concept's relationship with coarse tenability. In Section 6, index theory is used to study properties of Nash equilibrium components with support in the same Nash block. In Section 7, it is shown that minimal Nash blocks may contain strictly dominated strategies. Finally, related set-valued concepts are discussed in Section 8 and Section 9 contains a discussion of the concept's relationship with explicitly modeled dynamics.

1.2 Motivating example

Although the coordination game in the introduction suggests that the curb concept can be useful in refining the set of Nash equilibria in some games, the robustness properties of curb blocks are not necessarily related to the Nash equilibrium concept per se.[5] This may

[5]For example, a Nash equilibrium refinement can be defined by considering Nash equilibria with support in *minimal* curb blocks.

have adverse effects on the selectiveness of the concept, as suggested by the following simple example. Consider the generic perfect information extensive-form game given in Figure 1.

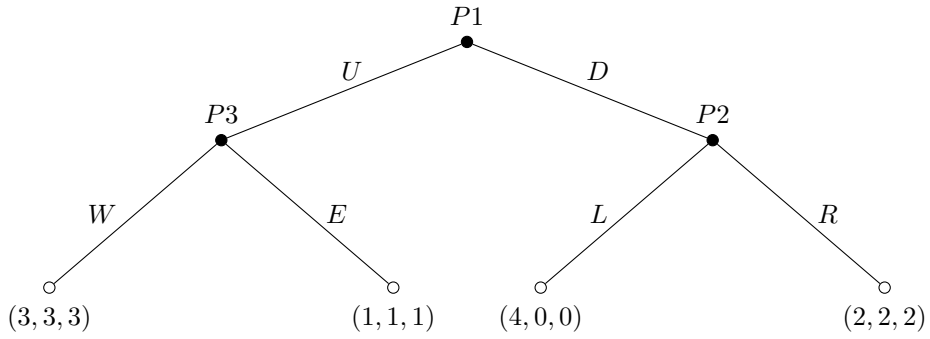

Figure 1.1: A generic perfect information extensive-form game.

Its normal form is given by

$Game\ 1:$		L	R
	U	$3,3,3$	$3,3,3$
	D	$4,0,0$	$2,2,2$

W

		L	R
	U	$1,1,1$	$1,1,1$
	D	$4,0,0$	$2,2,2$

E

This game has two Nash equilibrium components (disjoint, closed and connected sets of Nash equilibria): the component where 1 plays U, 2 assigns sufficiently little weight to L and 3 plays W; and the component where 1 plays D, 2 plays R and 3 assigns sufficiently little weight to W. The subgame perfect equilibrium (SPE) is (U, R, W).

In this game, each player's (pure) strategy in the SPE is a best reply to any strategy profile. Therefore, since a curb block includes all best replies to the strategy profiles with support in the block, it must include the SPE. As 2 is indifferent between her strategies in this equilibrium, both L and R must be included in the block. In fact, the unique curb block is the whole strategy space as D is optimal for 1 if 2 plays L and 3 plays W and E is optimal for 3 if 1 plays D and 2 plays R. It is worthwhile to point out that this feature is in no way related to L and W being weakly dominated strategies (it is possible to extend the game by adding strategies for 1 such that this is not the case without affecting the minimal curb block).[6]

[6]To see this, add a decision node for 1 after 1 plays U and 3 plays W, and after 1 plays D and 2 plays R. Let these decision nodes give 1 a choice between the respective end-node in the full game and an outcome that is very bad for all players. The modified game is still a generic perfect information extensive-form game. In the corresponding normal-form game, neither L nor W are weakly dominated.

By contrast, this game has two Nash blocks; trivially, the just-described curb block, and also $T = \{U, D\} \times \{L, R\} \times \{W\}$. Therefore, the set of NBE coincides with the equilibrium component containing the SPE. To see why T is a Nash block, consider the block game in which 3 only considers W. This corresponds to the two-player extensive form obtained from the extensive form in Figure 1 where the outcome is $(3, 3)$ if 1 plays U. The Nash equilibria of this block game has 1 playing U with probability 1. Thus, E is never optimal against any equilibrium in this component. The reason why U has to be included for 1 in the Nash block—hence why T is the unique minimal Nash block—is that (U, L, W) is a Nash equilibrium in the (one-player) block game where 1 plays U, 2 chooses between L and R, and 3 plays W. As is easily seen, this is not an equilibrium of the full game.

It can be shown that the SPE constitutes the unique minimal coarsely (and finely) tenable block. Note that this implies that tenable blocks sometimes fail to include the support of all Nash equilibria generating the same outcome in the corresponding extensive form. As will be shown below, this is not the case for Nash blocks.

1.3 Notation and definitions

A finite normal-form game (a *game*, for short) is a triple $G = \langle N, S, u \rangle$, where $N = \{1, 2, \ldots, n\}$ is the finite set of players, $S = \times_{i \in N} S_i$ the finite and nonempty set of pure-strategy profiles, and $u : S \to \mathbb{R}^n$ the payoff functions, with $u_i(s)$ representing the payoff to player i under strategy profile s.

Let player i's mixed-strategy set be denoted by $\Delta_i(S_i)$, the unit simplex in $\mathbb{R}^{m_i}$, where m_i is the number of elements in S_i. Let $m = \prod_{i \in N} m_i$. The mixed-strategy space, $\square[S] \subset \mathbb{R}^m$, is the Cartesian product of the simplices $\Delta_i(S_i)$. Each payoff function's domain is extended in the usual way from S to $\square[S]$ by

$$u_i(x) = \sum_{s \in S} \left[\prod_{j \in N} x_j(s_j) \right] \cdot u_i(s).$$

Let $u_i(x_{-i}, s_i)$ denote the payoff that player i gets from playing $s_i \in S_i$ when all other players play according to $x_{-i} \in \times_{j \in N \setminus \{i\}} \Delta(S_j)$. The set of *pure best replies* of player i against x is given by $\beta_i(x) = \arg\max_{s_i \in S_i} u_i(x_{-i}, s_i)$ with $\beta(x) = \times_{i \in N} \beta_i(x)$.

A strategy profile x is a *Nash equilibrium* if $x_i(s_i) > 0$ implies $s_i \in \beta_i(x)$. The nonempty set of Nash equilibria of a game is denoted $\square[S]^{NE}$. This set is semi-algebraic and consists of finitely many disjoint, closed and path-connected sets, so-called *equilibrium components* (Jiang, 1963). A Nash equilibrium is *strict* if it is the unique best reply against itself. Strict

equilibria evidently constitute singleton equilibrium components.

A pure strategy $s_i \in S_i$ is *weakly dominated* if there exists a mixed-strategy $y_i \in \Delta(S_i)$ such that $u_i(x_{-i}, y_i) \geq u_i(x_{-i}, s_i)$ for all $x \in \square[S]$, with strict inequality for some $x \in \square[S]$. It is *strictly dominated* if there exists a $y_i \in \Delta(S_i)$ such that $u_i(x_{-i}, y_i) > u_i(x_{-i}, s_i)$ for all $x \in \square[S]$. For any $\varepsilon > 0$, a strategy profile x such that $x_i(s_i) > 0$ for all $s_i \in S_i$ and $i \in N$ is *ε-proper* if

$$u_i(x_{-i}, s_i) < u_i(x_{-i}, t_i) \quad \Rightarrow \quad x_i(s_i) \leq \varepsilon \cdot x_i(t_i).$$

A *proper equilibrium* (Myerson, 1978) is any limit of some sequence of ε-proper strategy profiles as $\varepsilon \to 0$. The set of such Nash equilibria is nonempty in every game.

For any game G, a *block* is any set $T = \times_{i \in N} T_i$ such that $\emptyset \neq T_i \subseteq S_i \; \forall i \in N$. The associated *block game* is defined by $G^T = \langle N, T, u^T \rangle$ where u^T denotes the restriction of u to T. The mixed-strategy space of any block game is embedded in the strategy space of the *full game* G by identifying $\square[T]$ with $\{x \in \square[S] : x_i(s_i) = 0 \; \forall s_i \notin T_i, \; \forall i \in N\}$. A strategy profile x has support in a block T if $x \in \square[T]$. The set of Nash equilibria of G^T is denoted $\square[T]^{NE}$.

The following block concept is due to Basu and Weibull (1991): A block T is *curb* if $\beta(x) \subseteq T$ for all $x \in \square[T]$. Every game admits at least one minimal curb block.

1.4 Nash blocks

The block property to be introduced here requires that each player's set of strategies contains all best replies to any Nash equilibrium of the associated block game.

Definition 1. A block T is a *Nash block* if $\beta(x) \subseteq T$ for all $x \in \square[T]^{NE}$.

Using this definition, the game G^T where T is a Nash block is said to be a *Nash-block game*. The set of Nash equilibria of any Nash-block game coincides with the set of Nash equilibria of the full game with support in the block.[7] Thus, the concept defines a selection from the set of Nash equilibria of the full game.

I restrict attention to *minimal* Nash blocks. Since the set of pure-strategy profiles, S, is finite and trivially a Nash block, at least one minimal Nash block always exists. By considering the set of Nash equilibria with support in such blocks, one obtains a point-valued solution concept, a refinement of Nash equilibrium. Every game has at least one such equilibrium.

Definition 2. A *Nash-block settled equilibrium*, or NBE, is any Nash equilibrium with support in some minimal Nash block.

[7] As shown by Myerson and Weibull (2015), this is also a property of every coarsely tenable block.

I now derive a few properties of Nash blocks and relate the concept to curb blocks. It is easy to see that every curb block is a Nash block and that the converse does not hold, as in Game 2 below. Even so, the Nash block concept inherits many stability properties from the more demanding concept.

First, as noted by Basu and Weibull (1991), the curb concept can be interpreted as a generalization of strict Nash equilibrium. To see this, note that a strategy profile is a strict Nash equilibrium if and only if it constitutes a singleton curb block. This property is also satisfied by the Nash block concept.

Observation 1. *A pure strategy profile is a strict Nash equilibrium if and only if it constitutes a singleton Nash block.*

A second related observation is that, even if an individual attaches a small probability to the possibility that other individuals will not play a NBE, all her best replies are still in the block. Hence, the Nash block property is robust against perturbations of mixed strategies.[8]

Proposition 1. *If T is a Nash block, then there exists an open set U such that $\square[T]^{NE} \subset U$ and $\beta(x) \subseteq T$ for all $x \in U$.*

Proof. Given a Nash block T, for any $x \in \square[T]^{NE}$, $i \in N$, and $t_i \in \beta_i(x) \subseteq T_i$, we have $u_i(x_{-i}, t_i) > u_i(x_{-i}, s_i)$ for all $s_i \notin T_i$. Fix $x \in \square[T]^{NE}$. Since each payoff function u_i is continuous and each S_i finite, there exists a neighborhood $U_{i,x,t} \subset \mathbb{R}^m$ including x such that $u_i(y_{-i}, t_i) > u_i(y_{-i}, s_i)$ for all $y \in U_{i,x,t} \cap \square[S]$ and $s_i \notin T_i$. For all $x \in \square[T]^{NE}$ and $i \in N$, let the finite intersection of such sets define $U_{i,x} = \cap_{t_i \in \beta_i(x)} U_{i,x,t}$, implying that $u_i(y_{-i}, t_i) > u_i(y_{-i}, s_i)$ for all $y \in U_{i,x} \cap \square[S]$, $t_i \in \beta(x)$, and $s_i \notin T_i$. Define $U = \cup_{x \in \square[T]^{NE}} \cap_{i \in N} U_{x,i}$ which defines a neighborhood of $\square[T]^{NE}$. By construction, $\beta(x) \subseteq T$ for all $x \in U \cap \square[S]$. □

Third, Nash blocks 'respect' Nash equilibrium components; each such component is either disjoint from or contained in $\square[T]$ of any Nash block T. As shown by Ritzberger and Weibull (1995), curb blocks also have this property.

Proposition 2. *If T is a Nash block and ζ is a Nash equilibrium component, then either $\zeta \subseteq \square[T]^{NE}$ or $\zeta \cap \square[T]^{NE} = \emptyset$.*

Proof. The Nash equilibrium correspondence (assigning to each game a set of Nash equilibria)

[8] Ritzberger and Weibull (1995) show that this is also the case for any strategy profile in $\square[T]$ if T is a curb block.

is semi-algebraic. This implies that any Nash equilibrium component is closed and path-connected. For any block T with $X = \square[T]^{NE}$, let $x \in \zeta \cap X$ and $y \in \zeta \setminus X$ for a Nash equilibrium component ζ in G. By path-connectedness, there exists a continuous function $\gamma : [0,1] \to X$ with $\gamma(0) = x$ and $\gamma(1) = y$. As $\zeta \cap X$ is a closed set, by continuity there exists a $t \in [0,1)$ and an $\bar{\varepsilon}' \in (0,1)$ such that $\gamma(t) \in \zeta \cap X$ and $\gamma(t+\varepsilon) \in \zeta \setminus X$ for any $\varepsilon \in (0, \bar{\varepsilon}')$. By Proposition 1, there exists an $\bar{\varepsilon} \in (0,1)$ such that $\beta(\gamma(t+\varepsilon)) \subseteq \beta(\gamma(t))$ for any $\varepsilon \in (0, \bar{\varepsilon})$. Moreover, by assumption $\gamma(t)$ is a Nash equilibrium for all $t \in [0,1]$, and for any $\varepsilon \in (0, \min\{\bar{\varepsilon}, \bar{\varepsilon}'\})$ there exists a $s_i \notin T_i$ for at least one $i \in N$ with $\gamma(t+\varepsilon)(s_i) > 0$. As $s_i \in \beta_i(\gamma(t))$, I conclude that T is not a Nash block.

□

Example 2 provides a generic normal-form game in which there is a minimal Nash block that is not contained in any minimal curb block.

Example 2. Consider the game

$$\begin{array}{rccc} & L & C & R \\ U & 4,1 & 1,4 & 0,0 \\ M & 3,3 & 2,2 & 0,0 \\ B & 5,0 & -3,0 & 1,1 \end{array} \qquad \textit{Game 2}:$$

This game has three Nash equilibria given by $x = (\frac{1}{4}U + \frac{3}{4}M, \frac{1}{2}L + \frac{1}{2}C)$, $y = (B, R)$, and $z = (\frac{1}{14}U + \frac{3}{14}M + \frac{5}{7}B, \frac{1}{5}L + \frac{1}{5}C + \frac{3}{5}R)$. It has two minimal Nash blocks; $T = \{B\} \times \{R\}$ and $T' = \{U, M\} \times \{L, C\}$. Thus, both y and x are NBE while the completely mixed Nash equilibrium z is not. The unique minimal curb block is T.

According to the curb concept, T' is unstable since if 1 would assign high probability to the event that 2 will choose L, then 1's unique optimal strategy, B, is outside the block. Moreover, it does not suffice to simply add B to T', since the optimal strategy for 2 against B is R. By contrast, T' is a Nash block since all strategies that are optimal against the unique equilibrium x of the block game $G^{T'}$ are in T'.

Although strictly weaker than curb and being defined in terms of properties of Nash equilibria, the concept has cutting power in a variety of important classes of games for which it is well-known that established solutions concepts admits, arguably, implausible equilibria. For example, Laslier and Straeten (2004) analyze a class of games of electoral competition between parties that only care about being elected and compete by proposing different policies, and where the electorate only cares about the implemented policy. The

authors show that perfection cannot rule out any equilibrium outcome. By contrast, it can be shown that the Nash block concept eliminates all equilibria except those in which the parties propose the optimal policy given their private signal about the state of the world.

In a simple sender-receiver game due to Balkenborg et al. (2015), the sender observes the state of the world and sends a message to the receiver that, after having received the message, implements an action. Although both players' incentives are aligned, there exist a continuum of perfect equilibria in which the players fail to coordinate. Here, only the equilibria in which full coordination is achieved are NBE. Finally, only the 'intuitive' equilibrium in the signaling game due to Cho and Kreps (1987) is NBE.

1.5 Nash blocks and tenability

In this section, I analyze the Nash block concept's relationship with tenable strategy blocks. Myerson and Weibull's (2015) theory embeds the full game in a so-called consideration-set game. Such a meta-game endows every player role with a large population of boundedly rational individuals. For a block to be a potential convention, it is required that no individual should be able to do better by choosing a strategy outside the block when most individuals use strategies in the block. A coarsely tenable block formalizes such a convention when the overall population play constitutes a Nash equilibrium.

Formally, a game G is given a large population of individuals for each player role $i \in N$. One individual from each population is from time to time randomly drawn to play the game in her player role. Every individual is boundedly rational as she only considers a subset of the strategies available to her. Such a set of strategies is called the individual's consideration set, or her type. The type space for each player role i is given by $\Theta_i = \mathcal{C}(S_i)$, where $\mathcal{C}(S_i)$ is the collection of all nonempty subsets $C_i \subseteq S_i$. Let μ_i define a probability distribution on $\mathcal{C}(S_i)$ where $\mu_i(C_i) \in [0,1]$ is the probability that the individual drawn to play in role i is of the type $\theta_i = C_i \in \mathcal{C}(S_i)$. A vector $\mu = (\mu_1, \ldots, \mu_n) \in \times_{i \in N} \Delta(\mathcal{C}(S_i))$ is called a type distribution, and the draws from each population are statistically independent.

Each type distribution μ defines a game of incomplete information $G^\mu = \langle N, \times_{i \in N} F_i, u^\mu \rangle$, called a consideration-set game. A pure strategy for player $i \in N$ is given by a function $f_i : \mathcal{C}(S_i) \to S_i$ such that $f_i(C_i) \in C_i$ for all $C_i \in \mathcal{C}(S_i)$. The set of all such functions is denoted F_i, and the simplex of mixed strategies is denoted $\Delta(F_i)$ with generic element τ_i. A consideration-set game is connected to the full game as each mixed-strategy profile $\tau \in \square[F] = \times_{i \in N} \Delta(F_i)$ induces a corresponding mixed-strategy profile $\tau^\mu \in \square[S]$ in G. The conditional probability distribution over the strategies in S_i, induced by a strategy used by some type $\theta_i = C_i$, is denoted $\tau_{i|C_i}$. Hence, the probability that player i will use a pure

strategy s_i, given a strategy induced by τ_i, is

$$\tau_i^{\mu}(s_i) = \sum_{C_i \in \mathcal{C}(S_i)} \mu_i(C_i) \cdot \tau_{i|C_i}(s_i).$$

The expected payoff to each player i is given by $u_i^{\mu}(\tau) = u_i(\tau^{\mu})$. This defines the vector of expected payoff functions $u^{\mu} : \square[F] \to \mathbb{R}^n$ in G^{μ}. Every consideration-set game admits at least a Nash equilibrium, and it is straightforward to show that the projections of Nash equilibria of G^{μ} to G converges to Nash equilibria of G^T as $\mu_i(T) \to 1$ for all i.

A block T is *coarsely tenable* if there exists an $\varepsilon \in (0,1)$ such that $T \cap \beta(\tau^{\mu}) \neq \emptyset$ for every type distribution μ with $\mu_i(T_i) > 1 - \varepsilon \; \forall i \in N$ and every Nash equilibrium τ of G^{μ}. A *coarsely settled equilibrium* is any Nash equilibrium that has support in some minimal coarsely tenable block. For any block T and any $\varepsilon \in (0,1)$, a type distribution μ is $\varepsilon - proper$ on T if for every player $i \in N$

$$\begin{cases} (a) \;\; \mu_i(T_i) > 1 - \varepsilon, \\ (b) \;\; \mu_i(C_i) > 0 \quad \forall C_i \in \mathcal{C}(S_i), \\ (c) \;\; T_i \neq C_i \subset D_i \in \mathcal{C}(S_i) \quad \Rightarrow \quad \mu_i(C_i) \leq \varepsilon \cdot \mu_i(D_i). \end{cases}$$

A block T is *finely tenable* if there exists an $\bar{\varepsilon} \in (0,1)$ such that $T \cap \beta(\tau^{\mu}) \neq \emptyset$ for every type distribution μ that is $\bar{\varepsilon} - proper$ on T and every Nash equilibrium τ of G^{μ}. Myerson and Weibull (2015) show that every finely tenable block contains the support of a proper equilibrium. Therefore, a *finely settled equilibrium* is any proper equilibrium that has support in some finely tenable block. Finally, a *fully settled equilibrium* is any Nash equilibrium that is both coarsely and finely settled. Such an equilibrium exists in every game.

Throughout the examples given in this paper, minimal coarsely and finely tenable blocks coincide.

The Nash block concept is closely related to coarse tenability. In fact, it is possible to characterize the concept within the consideration-set framework that defines tenability. This not only highlights the relationship between the concepts but also provides a behavioral micro foundation for Nash blocks.

Call a block *strict coarsely tenable* if, when almost all individuals only consider strategies within the block, they do *strictly* better than any individual using strategies outside the block (given that the overall population play constitutes a Nash equilibrium). Formally:

Definition 3. A block T is *strict coarsely tenable* if there exists an $\varepsilon \in (0,1)$ such that $\beta(\tau^{\mu}) \subseteq T$ for every type distribution μ with $\mu_i(T_i) > 1 - \varepsilon \; \forall i \in N$ and every Nash equilibrium τ of G^{μ}.

Evidently, every strict coarsely tenable block is coarsely tenable. The below characterization then implies that every Nash block is coarsely tenable.

Proposition 3. *T is a Nash block if and only if it is strict coarsely tenable.*

Proof. ($\Leftarrow$) Suppose T is strict coarsely tenable. The set of Nash equilibria of a consideration-set game with $\mu_i(T_i) = 1 \; \forall i \in N$ then induces the set $\square[T]^{NE}$ in G. As a strict coarsely tenable block contains all best replies to the induced strategy profiles from the just mentioned consideration-set game, T is a Nash block.

($\Rightarrow$) Suppose T is a Nash block. By Proposition 1, there exists a neighborhood $U \subseteq \square[S]$ such that $\square[T]^{NE} \subset U$ and $\beta(x) \subseteq T$ for all $x \in U$. For any $\varepsilon \in (0,1)$, let X^ε denote the closed and nonempty set of strategy profiles induced by the set of Nash equilibria of a consideration-set game with $\mu_i(T_i) = 1 - \varepsilon$ for all $i \in N$. By continuity, there exists an $\bar{\varepsilon} \in (0,1)$ such that $X^\varepsilon \subset U$ for all $\varepsilon \in (0, \bar{\varepsilon})$. Thus, there exists an $\bar{\varepsilon} \in (0,1)$ such that $X^\varepsilon \subseteq U$ for all $\varepsilon \in (0, \bar{\varepsilon})$ completing the proof.

□

Corollary 1. *Every Nash block is coarsely tenable.*

In generic normal-form games, it is straightforward to show that every coarsely tenable block is a Nash block. This follows because, in such games, all Nash equilibria are quasi-strict: for any quasi-strict Nash equilibrium x, if $s_i \in \beta_i(x)$ then $x_i(s_i) > 0$ (see, e.g., van Damme (1991)).

Proposition 4. *Let G be a generic normal-form game. Then T is a Nash block if and only if it is coarsely tenable.*

As illustrated in Example 3, this result does not extend beyond generic normal-form games.

Example 3. Consider the generic extensive-form game given in Figure 2.

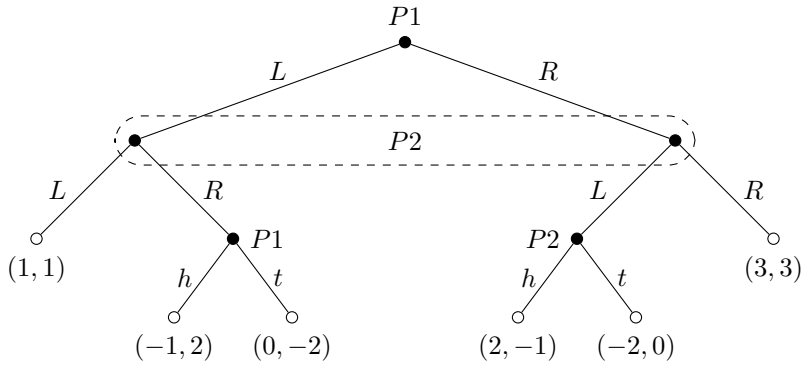

Figure 1.2: A coordination game with outside options in case of coordination failure.

It is an elaboration of a simple coordination game where one of the two pure Nash equilibria has been destabilized by giving one of the players an outside option if they fail to coordinate. The purely reduced normal-form representation of this extensive form is

$$\begin{array}{rccc} & Lh & Lt & R \\ Lh & 1,1 & 1,1 & -1,2 \\ Lt & 1,1 & 1,1 & 0,-2 \\ R & 2,-1 & -2,0 & 3,3 \end{array} \qquad Game\ 3:$$

This game has three Nash equilibrium components:

$$\Theta = \{(pLh + (1-p)Lt, qLh + (1-q)Lt) : p, q \in [0, 3/4]\},$$

the strict Nash equilibrium $x = (R, R)$, and the mixed Nash equilibrium $y = (\frac{1}{2}Lt + \frac{1}{2}R, \frac{1}{2}Lt + \frac{1}{2}R)$. The sole minimal Nash block is $T^1 = \{R\}^2$, the support of the NBE x. By contrast, the two minimal coarsely tenable blocks are T^1 and $T^2 = \{Lt\}^2$, thus, both x and $z = (Lt, Lt)$ are coarsely settled. The latter block is coarsely tenable as Lh is weakly dominated implying that it is redundant in any block including Lt.

Since every coarsely tenable block contains the support of a proper equilibrium (Myerson and Weibull, 2015), every game has a NBE that is proper. Using another machinery, it is possible to show that every Nash block also contains a set of Nash equilibria satisfying other demanding refinements in the literature. This is the topic of the upcoming section.

1.6 Index invariance

To provide robustness properties of sets of NBE, I here utilize results from the literature on index theory applied to Nash equilibrium components, as introduced by Ritzberger (1994). In particular, it is shown that it suffices to analyze the associated Nash-block game to determine the index of a Nash equilibrium component with support in a Nash block.

I first provide an informal description of index theory for Nash equilibrium components.[9] To compare Nash equilibrium components of a given game, the theory adds a differential structure on the mixed-strategy space. Given this structure, it is possible to use concepts from differential topology to classify equilibrium components using an index. This index allows for inferring global properties—such as whether an equilibrium is proper—without computing perturbations of, e.g., mixed strategies.

More specifically, the differential structure on $\square[S]$ is induced by the system of equations obtained from the necessary Karush-Kuhn-Tucker conditions for a strategy to be a best reply against a given mixed-strategy profile. This system is then interpreted as a vector field.

In generic normal-form games, the determinant of the Jacobian of this vector field is non-zero, or regular, evaluated at any equilibrium. For such games, the index of a Nash equilibrium component (which, thus, is a singleton) is the sign of the determinant of -1 times the Jacobian evaluated at the equilibrium. Hence, it is either $+1$ or -1. To assign indices to non-regular components (that may be set-valued), the vector field is slightly perturbed so as to resemble the vector field of a generic game. The index is then defined as the sum of the indices of the 'Nash equilibria' of the perturbed vector field that are within an isolating neighborhood of the component.

The first result obtained using this machinery establishes that the index of a Nash equilibrium component with support in a Nash block is the same as the index of the same component in the associated Nash-block game. This result builds on Ritzberger (2002, Proposition 6.8), who showed that the index of an equilibrium component is invariant with respect to deletion of a strategy that is never a best reply against the component (see also McLennan (2016, Theorem 5)).

Proposition 5. *If a Nash equilibrium component ζ has support in a Nash block T, then the index of ζ in G is also the index of ζ in the Nash-block game G^T.*

Proof. Let ζ be a Nash equilibrium component in G. The proof of Proposition 6.8 in Ritzberger (2002, p.327-328) implies that if $s_i \notin \beta_i(x)$ for $x \in \zeta$, then the index of ζ is the

[9] See Ritzberger (2002) for a textbook treatment.

same in G as in the block game $G^{S'} = \langle N, S', u \rangle$ for $S' = S_{-i} \times (S_i \setminus \{s_i\})$. As S is a finite set, any Nash-block game can be obtained by reducing the strategy space of G by removing a finite number of strategies that are not best replies to the set of Nash equilibria with support in the corresponding block. Thus, the index of this equilibrium components stays the same. □

The above result implies that global properties of any equilibrium component consisting of NBE can be analyzed by restricting attention to the associated Nash-block game. Moreover, due to the Poincaré-Hopf theorem, it is possible to say something about the robustness properties of sets of Nash equilibria with support in Nash blocks. This theorem has the remarkable implication that the index sum across all equilibrium components in any game is $+1$. This, in turn, implies that the index sum across the equilibrium components with support in the same Nash block is $+1$.[10]

Corollary 2. *If T is a Nash block, then the index sum across all Nash equilibrium components with support in T is $+1$.*

Proof. By the Poincaré-Hopf theorem, the index sum across all equilibrium components in a game G is $+1$. Moreover, every block game is a finite normal-form game independent of the structure of the full game, thus, the same holds true for all such games. An application of Proposition 5 completes the proof. □

Corollary 2 implies that every game has a set of NBE containing an essential component (Jiang, 1963) and an M-stable (strategically stable in the sense of Mertens (1989, 1991)) set. Roughly speaking, a Nash equilibrium component is essential if a nearby component exists in all games with nearby payoffs. A minimal connected set of Nash equilibria is strategically stable if it is robust against arbitrary small perturbations of mixed strategies and satisfies an additional technical condition.

Corollary 3. *Every Nash block contains the support of an essential component and an M-stable set.*

The corollary follows from Ritzberger (1994, Theorem 4) and Demichelis and Ritzberger (2003, Theorem 2), who showed that a component with non-zero index contains the claimed

[10] Two other result for *generic* finite normal-form games also hold for the set of equilibria of any generic Nash-block game: (i) the number of Nash equilibrium components is finite and odd (Harsanyi, 1973b) and (ii) if the game has $m \geq 1$ pure Nash equilibria, then it has at least $m - 1$ mixed Nash equilibria (Gül et al., 1993).

sets. As the index sum across all components with support in a Nash block is $+1$, there is at least a component with non-zero index.

The last observation pertains to the cutting power of minimal Nash blocks.

Observation 2. *If a game admits more than one minimal Nash block, then there exists at least one Nash equilibrium component that does not have support in any of them.*

This observation a straightforward implication of the Poincaré-Hopf theorem together with Corollary 2 since the index sum across all components in n Nash blocks is $+n$.

As illustrated in Example 4 below, the above index properties are not inherited by coarsely (and thus finely) tenable blocks.[11] In fact, there exist coarsely tenable blocks consisting of a single pure Nash equilibrium that lives in a component with index 0. Clearly, this equilibrium has index $+1$ in the (trivial) block game it generates.

Example 4. Consider the game

$$\begin{array}{rccc} & L & C & R \\ \textit{Game 4}: \quad U & 0,0 & 2,0 & 3,3 \\ D & 2,2 & 0,2 & 1,-1 \end{array}$$

It has two Nash equilibrium components; the strict Nash equilibrium $x = (U, R)$, and the non-convex component Γ consisting of the union of the connected components

$$\begin{aligned} A &= \left\{ \left(\frac{1}{2}U + \frac{1}{2}D, \frac{1}{2}L + p\frac{1}{2}C + (1-p)\frac{1}{2}R \right) : p \in [0,1] \right\} \\ B &= \{(D, pL + (1-p)C) : p \in [1/2, 1]\} \\ C &= \left\{ \left(qU + (1-q)D, \frac{1}{2}L + \frac{1}{2}C \right) : q \in [0, 1/2] \right\}. \end{aligned}$$

Since x is strict, its index is $+1$. This implies that Γ has index zero.

The sole minimal Nash block is $T = \{U\} \times \{R\}$, implying that x is an NBE. However, there exists another minimal coarsely tenable block $T' = \{D\} \times \{L\}$. The latter is coarsely tenable since L and C are payoff equivalent for 2, implying that both are never included in the same minimal coarsely tenable block. Note that the unique strategy profile with support in T' belongs to the component with index 0. As x is the unique Nash equilibrium of any game in which 2's payoff from (D, C) is increased by $\varepsilon > 0$, the component with index 0 is not essential.[12]

[11] Of course, since every curb block is a Nash block, the former satisfies all of them.

[12] An open question is whether every coarsely tenable block includes the support of an M-stable set.

1.7 Dominated strategies

A perhaps surprising property is that a minimal Nash block may not survive the deletion of strictly dominated strategies. Furthermore, such strategies may even be *included* in minimal Nash blocks. This failure of invariance is also inherited by minimal coarsely and finely tenable blocks.[13] I here provide the intuition behind this property which is illustrated in Example 5 below. Thereafter, I discuss the Nash block concept's invariance under the addition of strictly dominated strategies and its implications.

It is useful to begin with a couple of simple observations. First, Myerson and Weibull (2015, p.954) observe that a block containing all strategies that are not weakly dominated is coarsely tenable. The same observation holds for Nash blocks if weak is replaced by strict dominance. Second, if a Nash block includes a strategy that is strictly dominated in its block game, it is not minimal. This follows from the simple observation that a strictly dominated strategy is never a best reply to any strategy profile. Thus, it can be excluded from the block without loss.

However, a strictly dominated strategy may be undominated in a Nash-block game if the strategy that dominates it is not in the block.[14] This implies that a strictly dominated strategy can, if included in such a block, stop a strategy profile from being a Nash equilibrium of the block game. Without this strategy, the block is not a Nash block as the block game includes a Nash equilibrium that is not an equilibrium in the full game (since, by assumption, the inclusion of the dominated strategy eliminates it).

By definition, every strictly dominated strategy must have at least one (possibly mixed) strategy that dominates it. And although the support of this strategy could replace the strategy it dominates in the block, the resulting block may still not be minimal. To see this, consider a minimal Nash block that contains a strictly dominated strategy. Replace this strategy by the support of the strategy that dominates it. This block is *not* minimal if there exist a subblock including the just-added strategies that constitutes a minimal Nash block.

The above reasoning is illustrated in the upcoming example. In this example, the addition of a strictly dominated strategy allows for a new minimal Nash block and increases the set of NBE. The game is based on an elaboration of a simple coordination game, introduced by Myerson and Weibull (2015), in which the 'miscoordination end-nodes' (in its extensive form) are replaced by zero-sum subgames.

[13] In contrast, Basu and Weibull (1991) have shown that minimal curb blocks only contain strategies that survive the iterated removal of strictly dominated strategies.

[14] It might be the case that a strictly dominated strategy is dominated by a mixed-strategy profile. Then, the same observation holds true if any of the strategies in the support of the mixed strategy profile is missing from the block.

Example 5. Consider two versions of the following game

		Lh'	Lt'	Rh'	Rt'	D	A'
	Lh	$1,1$	$1,1$	$-2,2$	$2,-2$	$-1,-1$	$0,0$
	Lt	$1,1$	$1,1$	$2,-2$	$-2,2$	$-1,-1$	$0,0$
Game 5 :	Rh	$2,-2$	$-2,2$	$1,1$	$1,1$	$-3,3$	$5,5$
	Rt	$-2,2$	$2,-2$	$1,1$	$1,1$	$-3,3$	$5,5$
	A	$3,0$	$-6,-1$	$-4,-4$	$-4,-4$	$-8,2$	$7,7$

one version where D is available for 2 and one where it is deleted. Note that D is strictly dominated by A'.

This game has three Nash equilibrium components:

$$\Theta = \{(pLh + (1-p)Lt, qLh' + (1-q)Lt') : p, q \in [1/4, 3/4]\},$$

$$\Omega = \{(1/12)([pLh + (1-p)Lt + Rh + Rt], [5Lh + 5Lt + 2A']) : p \in [1, 7]\},$$

and $\{x\} = (A, A')$. Of course, the set of Nash equilibria does not depend on whether D is included in 2's strategy set. By contrast, as will be seen this is not true for the set of settled equilibria.

In the version of the game where D is available for 2, it is possible to show that there exist two minimal Nash blocks, $T' = \{A\} \times \{A'\}$ and $T = \{Lh, Lt, Rh', Rt', A\} \times \{Lh', Lt', Rh, Rt, D\}$. Thus, the set of NBE is given by $\{x\}$ and Ω. Notice that D is undominated in the block game G^T. In this game, the set of Nash, coarsely tenable and finely tenable blocks coincide.

Consider now the version of Game 5 where D is deleted. In this game, T' is the unique minimal Nash block, and $\{x\}$ is the set of NBE. To see this, consider the block $T^* = T_1 \times (T_2 \setminus D)$, that is, T excluding D for 2. It is not a Nash block as the set of Nash equilibria of the corresponding block game includes a Nash equilibrium component where A is the unique optimal strategy for 2.[15] However, adding A to T^* does *not* generate a minimal Nash block as the resulting block properly contains the Nash block T'.

It is easy to show that Nash blocks are invariant under the addition of strictly dominated strategies. That is, the introduction of such a strategy can never make a Nash block seize to exist. This property extends to the addition of strategies that are never best replies to any Nash equilibrium of a Nash-block game.

[15] The set of Nash equilibria of G^{T^*} is given by $\square[T^*]^{NE} = \Theta \cup \{(pRh' + (1-p)Rt', qRh + (1-q)Rt) : p, q \in [1/4, 3/4]\}$, where $x \in \square[T^*]^{NE} \setminus \Theta$ implies $\beta_2(x) = \{A'\}$ in G.

Observation 3. *Let T be a minimal Nash block in G. Then T is also a minimal Nash block in any game G' obtained from G by the addition of strategies that are never best replies against the set of NBE with support in T.*

The above observation implies that every game has a minimal Nash block, and therefore a set of NBE, that is invariant against the addition and deletion of strictly dominated strategies. The set of such NBE is easy to identify: reduce the strategy space of a game by iteratively deleting strictly dominated strategies and then applying the Nash block concept. The reduced game contains all the minimal Nash blocks (which exist in the full game) that are invariant against the addition and removal of dominated strategies.

It is interesting to note that most established Nash equilibrium refinements are not invariant to the addition of strictly dominated strategies. This failure of invariance has been defended by Kohlberg and Mertens (1986) when discussing strategic stability on the grounds that strategic stability "depends on the whole given situation. So, when some implausible alternatives are deleted, the analysis has already taken their unlikeliness into account. However, adding possibilities that were physically not present previously cannot and should not have been anticipated" (Kohlberg and Mertens, 1986, p.1017).

In the setting analyzed in this paper, where Nash equilibrium is interpreted as the outcome of a dynamic process and individuals tend to ignore strategies that are unconventional (and therefore never use them), it might often be hard to exactly pin down what is meant by "the whole given situation." A more pragmatic approach is to concede, in agreement with McLennan (2016), that "[e]conomic modeling requires strategic simplification. A model necessarily specifies only a few features of the world. The social scientist hopes that the selected features are the critical ones..." (McLennan, 2016, p.26-27). Confining the analysis of a game to one of its Nash-block games can be interpreted as such a 'strategic simplification.' If such an approach is taken, it is desirable that 'inessential' strategies do not affect the criterion used to predict potential outcomes. In this view, the above observation is an important robustness requirement that shows that the concept is *not*, in the words of McLennan (2016, p.27), "excessively sensitive to minor details of model specification."

Note that the invariance under the addition of strictly dominated strategies depends crucially on the requirement that *all* best replies to the Nash equilibria of the Nash-block game are included in the block. Therefore, coarse and fine tenability does *not* have this robustness property.

1.8 Related literature

In this section, I present other related ideas and discuss their relationship with Nash blocks. A concept that can be reformulated as a block property was introduced by Kalai and Samet (1984). In any game in which no player has any payoff-equivalent strategies among her undominated strategies, the set of mixed-strategy profiles $\square[T]$ is an *absorbing retract* if there exists an open set U containing $\square[T]$ such that $\beta(x) \cap T \neq \emptyset$ for all $x \in U$. A *persistent retract* is a minimal absorbing retract. A *persistent equilibrium* is any Nash equilibrium with support in a persistent retract. It is easy to show that if T is a curb block, then $\square[T]$ is an absorbing retract, and if $\square[T]$ is an absorbing retract, then T is coarsely tenable. Another closely related idea is so-called prep sets, as introduced by Voorneveld (2004). A prep set is a block T such that $\beta(x) \cap \square[T] \neq \emptyset$ for all $x \in \square[T]$. If $\square[T]$ is an absorbing retract then T is a prep set. However, not all prep sets are coarsely tenable and vice versa.

Neither prep sets nor absorbing retracts are implied by or imply that the associated block is a Nash block. In Game 1 above, the two concepts coincide with the minimal coarsely tenable block, thus is a subset of the minimal Nash block. In Game 2, there exists a Nash block that is neither in a subblock of a prep set nor in a subblock that spans an absorbing retract. Game 3 provides an example of a reduced normal-form game, obtained from a generic extensive form, where there exists an absorbing retract that is not spanned by a subblock of a minimal Nash block. Moreover, in Game 4, there is a singleton coarsely tenable block that spans an absorbing retract which is not a Nash block. Finally, in the version of Game 5 when the strictly dominated strategy is added, there is a Nash block that is neither a prep set nor spans an absorbing retract. In the version of this game where the dominated strategy is removed, all the concepts considered so far in this paper coincide.

Voorneveld (2004, 2005) shows that curb blocks, prep sets and absorbing retracts coincide in generic normal-form games (T is, e.g., a prep set if and only if $\square[T]$ is an absorbing retract). As seen in Game 2, although coarsely tenable and Nash blocks coincide in generic normal-form games, they are generically distinct from curb blocks.

Kuzmics et al. (2013) and Balkenborg et al. (2015) analyze refinements of the best-reply correspondence that are upper hemi-continuous, closed- and convex-valued. They use these refined correspondences to provide weaker variations of curb blocks and prep sets. However, their correspondences coincide with the usual best-reply correspondence on generic normal-form games. Thus, these concepts generically differ from Nash blocks.

Finally, a related idea is p-best response sets by Tercieux (2006a) (see also Tercieux (2006b) for an analysis of a weaker requirement). A block is a p-best response set if it contains all best replies to all beliefs putting at least probability p on the block. However, in his paper,

beliefs are not constrained to treat other players' strategy choices as statistically independent. In two-player games this generalization is vacuous, implying that any curb block is a p-best response set for some $p < 1$ (Ritzberger and Weibull, 1995).

1.9 Discussion

I have here developed a block concept that captures candidates for potential conventions in a setting in which individuals are repeatedly and randomly drawn from large populations to play a game against each other, as in Nash's mass action interpretation of his equilibrium concept. While not explored here, the concept captures some notion of dynamic stability when explicitly modeled. As shown by Demichelis and Ritzberger (2003), a necessary condition for an equilibrium component to be stable with respect to any 'natural' dynamic process, in the sense that the individuals adjust their strategies toward those that generate higher payoffs, is that its index agrees with its Euler characteristic. Without going into detail on what the Euler characteristic is, in generic normal-form games, and in normal forms of generic two-player extensive-form games, this condition is fulfilled for any component that corresponds to a unique Nash equilibrium component of a Nash-block game.

It would be interesting to further explore connections between Nash blocks and explicit models of population dynamics. For example, Kuzmics et al. (2013) show that a sufficient condition for the set of mixed strategies with support in a block to be asymptotically stable under the best-reply dynamics is that the block is curb (see, e.g., Young (1993) for an analysis of stochastic dynamics). In such dynamic population models, the robustness properties of Nash blocks suggest that sets of NBE with support in the same minimal Nash block could be good predictors.

Chapter 2: Evolutionary stability and tenable strategy blocks

Abstract

I analyze relationships between evolutionary stability and tenable strategy blocks (Myerson and Weibull, 2015). I find that in two-player games, if a strategy profile is robust against equilibrium entrants, or REE (Swinkels, 1992b), then it has support in a minimal coarsely tenable block and is fully settled in the sense of Myerson and Weibull. As a result, coarse tenability captures van Damme's (1989) version of forward induction in the same way as REE does (Hauk and Hurkens, 2002). I provide two new evolutionary stability definitions and show that they completely characterize the tenable block concepts. Moreover, in symmetric two-player games, established notions of evolutionary stability are shown to imply symmetric versions of these concepts.

2.1 Introduction

When individuals are recurrently and randomly matched with each other to play a game, they have the opportunity to coordinate their interactions over time. In a societal context, in which the population of individuals can be taken to be (infinitely) large, such coordination is often achieved through norms or conventions and usually involve individuals neglecting freely available actions that are deemed unconventional (Hurwicz, 2008).

Recently, Myerson and Weibull (2015) developed a theory that permits the endogenous formation of such conventions in finite normal-form games. In particular, they formalized a meta game in which every player role is endowed with a large population of boundedly rational individuals, similar to Nash's mass-action interpretation (Nash, 1950). These individuals are boundedly rational in the sense that they do not consider all the strategies available to them. The authors take the basis of a convention to be a strategy *block*; a nonempty set of pure strategies for each player role. Such a convention is said to be *coarsely tenable* if, when almost all individuals in every population only consider the block strategies, no one can do better by using a strategy outside the block, given that the overall population play constitutes a Nash equilibrium. A weaker version of such a convention is said to be *fine tenability* if the above criterion is relaxed to hold only for populations where the individuals are "biased" towards more rational types.

While the associated solution concepts derived from minimal tenable blocks, so-called *settled equilibria*, are generically distinct from all established solution concepts and offer cutting power in important classes of games, Myerson and Weibull do not provide any formal justification for why the overall population play should equilibrate. However, the authors suggest that tenability could be given evolutionary interpretations in which the justification for equilibrium differs starkly from those proposed in traditional game theory. In such interpretations, a tenable block is seen as necessary robustness condition when almost everyone plays a conventional, or "incumbent," strategy and only a few individuals play the new unconventional, or "entrant," strategy. A population equilibrium is then viewed as an outcome of trial and error where more successful individual behavior tends to be more prevalent.

In this paper, I set out to provide such evolutionary interpretations of tenability. I do this in two ways: First, I explore the concepts' formal connections with already established notions of evolutionary stability, and, second, I characterize the tenability concepts as evolutionary stability properties.

Maynard Smith and Price's (1973) notion of evolutionary stability is defined for symmetric games and features a single population of individuals that are uniform random matched with

each other. Here, the incumbents constitute a population of individuals utilizing a given mixed strategy. Such a strategy is *evolutionarily stable* if it does better on average than any small population share of unconventional individuals, or entrants—utilizing a different strategy—when the entire population is made up of mostly incumbents and a small share entrants.

Swinkels (1992b) weakens this stability criterion to hold only for entrants that are "rational" in the sense that their strategy is a best reply to the resulting population mix, when a small share of them have entered the population. He calls this notion *robustness against equilibrium entrants*, or REE, which is extended to a set-valued concept called *equilibrium evolutionarily stable*, or EES, set. Both concepts are defined for n-player games in a multipopulation framework with random matching.

The main result in this paper establishes that in arbitrary two-player games, the support of a REE is a unique Nash equilibrium in a minimal coarsely tenable block consisting of its best replies. Thus, coarse and fine tenability can be seen as a generalization of REE in two-player games.

As an application of this result, I analyze coarse tenability's ability to capture a version of forward induction reasoning in outside option games (Kohlberg and Mertens, 1986). In such a game, player 1 first decides between 'enter' or 'stay out.' If 1 decides to stay out, the game ends, if 1 decides to enter, 1 and 2 play a given a simultaneous move subgame. Following van Damme (1989), I consider outside option games in which the subgame has a finite number of equilibria and exactly one equilibrium gives more payoff to 1 than her outside option. *Forward induction* then captures the idea that, if 1 decides to enter the subgame she signals that she plans to play that equilibrium. It has been argued that if there is mutual knowledge of rationality of the players, they would only play this so-called *forward induction equilibrium*.

Hauk and Hurkens' (2002) analysis of evolutionary stability in outside option games shows that whenever an EES set exists, it uniquely selects the forward induction equilibrium, which then constitutes a REE. Similarly, I show that every outside option game has a unique minimal coarsely tenable block and that this block contains the forward induction equilibrium. As a consequence, whenever an EES set exists, the forward induction equilibrium is the unique coarsely settled equilibrium. Hence, this result implies the result of Hauk and Hurkens (2002). I show by way of example that there are games in which no EES set exists but the forward induction equilibrium is the unique coarsely settled equilibrium. Thus, coarse tenability captures forward induction, arguably, better than does EES.

I provide characterizations of both coarse and fine tenability that do not rely on the meta-game framework that defines the tenability concepts (Myerson and Weibull, 2015). In addition, I illustrate how these characterizations can be motivated by evolutionary considerations.

There are two main features of these concepts that differ from established concepts of evolutionary stability. First, such strategy blocks are robust against multiple entrants in the sense that, when already facing a small population share of entrants, the incumbents do weakly better than any potential entrant that might not be present in the population. Second, the incumbent strategy profile depends on the share of entrants in the sense that it is a best reply, among the strategies with support in the block, to the resulting population mix consisting mostly of the incumbents and a small share of entrants.

Utilizing these characterizations, I adapt coarse tenability to single population symmetric two-player games and explore connections between *symmetric* coarse tenability and evolutionary stability. I show that the best replies to any *evolutionarily stable set* (Thomas, 1985)—a set-valued generalization of evolutionarily stable strategy—is a symmetric coarsely tenable block. Furthermore, every symmetric singleton coarsely tenable block is a *neutrally stable strategy* (a weaker version of evolutionary stability (Maynard Smith, 1982)). Thus, symmetric coarsely tenability falls between the two evolutionary stability concepts, but relates to strategy blocks rather than individual strategies.

The rest of the paper is organized as follows. The upcoming section introduces the notation and some of the definitions used throughout the paper. Section 2.3 provides the main result consisting of a formal connection between tenability and EES sets. Section 2.4 contains an application of this result to outside options games and forward induction. In Section 2.5, I present the consideration-set framework, the tenable block concepts and formally establish their evolutionarily characterizations. In Section 2.6, I explore tenability's connections with evolutionary stability in symmetric two-player games. Finally Section 3.8 concludes.

2.2 Preliminaries

Let $G = \langle N, S, u \rangle$ be a finite normal-form game, where $N = \{1, 2, \ldots, n\}$ is the set of players, $S = \times_{i \in N} S_i$ is the set of pure-strategy profiles, and $u : S \to \mathbb{R}^n$ is the combined payoff function. The set of mixed-strategy profiles, $\square[S]$, is the Cartesian product of $\Delta(S_i)$, the unit simplex in $\mathbb{R}^{m_i}$, where m_i is the number of elements in S_i. Every payoff function u_i is extended to $\square[S]$ the usual way, and $u_i(s_i, x_{-i})$ denotes the payoff that player i gets from playing the pure strategy $s_i \in S_i$ when all other players play according to $x_{-i} \in \times_{j \neq i} \Delta(S_j)$. For player i and any mixed-strategy profile $x \in \square[S]$, let $\beta_i(x)$ and $C_i(x_i)$ be $i's$ set of pure best replies and strategies in the support of x_i, respectively. Let $\beta(x) = \times_{i \in N} \beta_i(x)$ and $C(x) = \times_{i \in N} C_i(x_i)$.

A strategy profile x is a Nash equilibrium if $x_i(s_i) > 0$ implies $s_i \in \beta_i(x)$ and the set of Nash equilibria of G is denoted E. This set is semi-algebraic and consists of finitely many

disjoint, closed and connected sets, so-called *equilibrium components* (Jiang, 1963). A strategy profile x is *completely mixed* if $C(x) = S$. For any $\varepsilon > 0$, a completely mixed strategy profile x is *ε-proper* if

$$u_i(s_i, x_{-i}) < u_i(s_i', x_{-i}) \quad \Longrightarrow \quad x_i(s_i) \leq \varepsilon \cdot x_i(s_i').$$

A *proper equilibrium* (Myerson, 1978) is any limit of some sequence of ε-proper strategy profiles as $\varepsilon \to 0$. The set of such Nash equilibria is nonempty in every game.

Given a game G, a *strategy block* is any set $T = \times_{i \in N} T_i$ such that $\emptyset \neq T_i \subseteq S_i \; \forall i \in N$. Note that both $\beta(x)$ and $C(x)$ are blocks for any $x \in \square[S]$. The associated *block game* is defined as $G^T = \langle N, T, u_{|T} \rangle$ where $u_{|T}$ is the payoff function of G restricted to strategy profiles in T. The set of mixed-strategy profiles of a block game can be embedded in $\square[S]$. Hence, the set of strategy profiles with support in T, and the set of mixed-strategy profiles of the block game G^T, are both denoted by $\square[T]$ whenever this does not cause confusion. The set of Nash equilibria of G^T is denoted by E^T, with $E^S = E$.

Given a game G, for any $i \in N$ and $x \in \square[S]$, denote i's set of best replies *among the strategies in T_i* by

$$\beta_{i|T_i}(x) = \arg\max_{s_i \in T_i} u_i(s_i, x_{-i})$$

with $\beta_{|T}(x) = \times_{i \in N} \beta_{i|T_i}(x)$.

The main evolutionary stability concept considered in this paper is called Equilibrium Evolutionary Stability, or EES, and was introduced by Swinkels (1992b). The concept is set-valued and is defined for any finite normal-form game. It provides sets that are robust against "equilibrium entrants" in the sense that a small share of entrants can only enter and survive if the (mixed) strategy they play is a best replay to the resulting populations' strategy profile.

Definition 4 (Swinkels, 1992b). A nonempty and closed set $X \subseteq \square[S]$ of Nash equilibria is an *equilibrium evolutionarily stable set*, or EES set, if it is a minimal set with property

($\boldsymbol{P}$) there exists an $\bar{\varepsilon} > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$, $x \in X$ and $y \in \square[S]$,

$$C(y) \subseteq \beta((1-\varepsilon)x + \varepsilon y) \Longrightarrow (1-\varepsilon)x + \varepsilon y \in X.$$

The strategy profile that constitutes a singleton EES set is called *robust against equilibrium entrants*, or REE. Swinkels (1992b) shows that every EES set is a Nash equilibrium component. Moreover, if an EES set constitutes a REE, or if the normal form represents a generic two-player extensive-form game, then it contains a proper equilibrium and a strategically stable

subset (Swinkels, 1992a,b).[1]

2.3 Analysis

In this section, I provide the main result that establishes a formal connection between tenable strategy blocks and equilibrium evolutionary stability. The result obtained here shows that oftentimes coarse and fine tenability can be motivated by notions of evolutionarily stability as formalized by the REE notion.

The next two concepts are new. It will be shown later that in all finite games they characterize coarse and fine tenability, respectively (Myerson and Weibull, 2015). For expositional purposes, the formal characterization results are postponed until Section 2.5.

Definition 5. A block T is *coarsely evolutionarily stable*, or CES, if there exists an $\bar{\varepsilon} > 0$ such that for any $x \in \square[T]$, $y \in \square[S]$ and $\varepsilon \in (0, \bar{\varepsilon})$,

$$C(x) \subseteq \beta_{|T}((1-\varepsilon)x + \varepsilon y) \Longrightarrow C(x) \subseteq \beta((1-\varepsilon)x + \varepsilon y).$$

In words, a CES block satisfies the following property: Whenever the population consists mostly of the incumbents and a small share of entrants, if the incumbent strategy profile is a (mixed) best reply among the strategies with support in the block, then there exists no (strictly) better replies outside of the block.

There are two features of this formulation that differ from established concepts of evolutionary stability. The first difference is that CES blocks are robust against multiple entrants in the sense that the type represented by y may consist of different types of entrants, each associated with a different mixed-strategy profile. The incumbents do weakly better than any entrant, even if the associated share of this entrant in y is infinitesimal. The second difference is that the incumbent strategy profile depends on the entrant strategy profiles making up y. That is, the incumbent strategy profile is a (mixed) best reply, among the strategies in the face spanned by the block, to the resulting population mix when a small share of entrants has entered the population.

Thus, the concept weakens evolutionary stability by allowing for evolutionary *instability* within the support of a limited number of strategies, i.e. the strategy block, while requiring robustness against any entrant using strategies outside the block. For any CES block, when the share of entrants tends to zero, the only incumbent strategy profiles that are best replies to the resulting population mix are those that approximate Nash equilibria of the full game.

[1] See Kohlberg and Mertens (1986) for the original definitions of strategic stability.

Therefore, one can view the set of Nash equilibria with support in the same CES block as evolutionarily stable in the just described sense.

To present the next concept, I first have to introduce a new property of strategy profiles that is closely related to ε-properness.

Definition 6. For any $x \in \square[S]$ and $\varepsilon > 0$, a completely mixed strategy profile y is *ε-proper relative to x* if, for all $i \in N$, there is an $\epsilon_i \in (0, \varepsilon)$ such that

$$u_i(s_i, z_{-i}) < u_i(s_i', z_{-i}) \quad \Rightarrow \quad y_i(s_i) \leq \varepsilon \cdot y_i(s_i') \text{ for all } s_i, s_i' \in S_i,$$

where $z = ((1 - \epsilon_i)x_i + \epsilon_i y_i)_{i \in N}$.

Definition 7. A block T is *finely evolutionarily stable*, or FES, if there exists an $\bar{\varepsilon} \in (0, 1)$ such that for any y that is *ε-proper relative to $x \in \square[T]$* for an $\varepsilon \in (0, \bar{\varepsilon})$,

$$C(x) \subseteq \beta_{|T}(z) \Longrightarrow C(x) \subseteq \beta(z).$$

This evolutionary stability property is evidently weaker than CES as the set of potential entrants is a subset of those in the latter formulation. Thus, every CES bock is FES. The difference between the concepts is that, for FES blocks, all pure strategies are used by the entrants. Moreover, entrants using strategies giving higher payoffs are more prevalent.

A set-valued solution concept based on this notion can be formalized as the set of possible incumbent strategy profiles with support in a FES block, when the share of "ε-proper entrants" tends to zero. It can be shown that such incumbent strategy profiles constitutes proper equilibria.

Every game has at least one CES (a fortiori FES) block as the whole strategy set, S, trvially constitutes a CES block. This implies that the concepts only achieve cutting power after restricting attention to a subclass of such blocks. For example, minimal (in terms of set inclusion) CES and FES blocks exist in every game.[2]

The first observation linking CES and EES together is that, in two-player games, EES sets put restrictions on the block game consisting of the best replies to an element in the set. This property was first suggested by Swinkels (1992b).

Say that a set $X \subseteq \square[S]$ has property $(\boldsymbol{P'})$ if there exists an $\bar{\varepsilon} > 0$ such that for all

[2]Swinkels (1992b, p.320-321) acknowledges that one could define a notion of EES sets without insisting on the elements being Nash equilibria. Versions of such sets are explored in Matsui (1992) and Swinkels (1992a). These sets have the benefit of existing in every game. By contrast, the evolutionary interpretations of the set of Nash equilibria with support in the same CES or FES block developed here do not rely on viewing non-equilibrium strategy profiles as "stable" while achieving existence in all finite games.

$\varepsilon \in (0, \bar{\varepsilon})$, $x \in X$ and $y \in \square[S]$,

$$y \in E^T \text{ for } T = \beta(x) \Longrightarrow (1-\varepsilon)x + \varepsilon y \in X.$$

In words, this property states that for every element x in the set, if any strategy profile y is a Nash equilibrium of the block game generated by the best replies to x, then the strategy profile given by mostly x and a small share of y is in the set.[3]

Lemma 1. *Let G be a two-player game. Then every EES set has property ($\boldsymbol{P'}$).*

Proof. Let X be an EES set and let $\bar{\varepsilon} > 0$ be such that if $\varepsilon \in (0, \bar{\varepsilon})$, $x \in X$, and $C(y) \subseteq \beta((1-\varepsilon)x + \varepsilon y)$, then $(1-\varepsilon)x + \varepsilon y \in X$. In two-player games, the utility functions are linear in probabilities of the other player's strategy profile. Thus,

$$u_i(y_i, (1-\varepsilon)x_{-i} + \varepsilon y_{-i}) = (1-\varepsilon)u_i(y_i, x_{-i}) + \varepsilon u_i(y_i, y_{-i}).$$

Note that, if $y \in E^T$ for $x \in X$ and $T = \beta(x)$, then $C(y) \subseteq T$ and $C(y) \subseteq \beta_{|T}(y)$. Since $s_i \notin \beta_i(x)$ implies that $u_i(y_i, y_{-i}) > u_i(s_i, x_{-i})$ for all $i \in \{1, 2\}$ and the game is finite, there exists an $\bar{\varepsilon}' > 0$ such that

$$(1-\varepsilon)u_i(y_i, x_{-i}) + \varepsilon u_i(y_i, y_{-i}) \geq (1-\varepsilon)u_i(s_i, x_{-i}) + \varepsilon u_i(s_i, y_{-i})$$

for all $s_i \in S_i \backslash \beta_i(x)$, all $i \in \{1, 2\}$, and all $\varepsilon \in (0, \bar{\varepsilon}')$. This implies that $C(y) \subseteq \beta((1-\varepsilon)x + \varepsilon y)$ for all $\varepsilon \in (0, min\{\bar{\varepsilon}, \bar{\varepsilon}'\})$, hence $(1-\varepsilon)x + \varepsilon y \in X$.

□

An important fact that follows from Lemma 1 is that, in two-player games, a REE's support contains *all* its best replies. I here show that this block of best replies is a minimal CES block. This result allows equilibria in minimal CES and FES block notions to be interpreted as set-valued extensions of the REE concept in two-player games.

Proposition 6. *Let G be a two-player game. If x is a REE, then $\beta(x) = C(x)$ and $T = C(x)$ is a minimal CES block.*

Proof. Assume that x is a singleton EES set, then by Lemma 1, x is the only Nash equilibrium of the block game G^T with $T = \beta(x)$. Since T contains all best replies to x, it is a CES block.

[3] As observed by Matsui (1992), Swinkels (1992b) claimed erroneously that every EES set has property ($\boldsymbol{P'}$). Matsui (1992) maintained that it is true for two-player games but provided a counterexample showing that it is false for games with more than two players.

Moreover, as all finite normal-form games contain an essential equilibrium component (Jiang, 1963), x is an essential equilibrium of G^T. As all best replies to x are in the block game, it must be essential in the full game too. An isolated essential equilibrium in a two-player game is regular (van Damme (1991) Theorem 3.4.4) and such an equilibrium has the property that $\beta(x) = C(x)$.

□

In the upcoming example, I show that this relationship does not extend beyond two-player games as there exist singleton EES sets that are not contained in the support of a minimal CES, nor a minimal FES, block. The example also serves as a counterexample to the claim in Lemma 4 in Swinkels (1992b) that every pure-strategy profile that is REE constitutes a strict Nash equilibrium.

Example 6. Fix any $(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$ and consider the slight modification of a game due to van Damme (1991, Fig 3.4.1.)

Game 1 :		C	D		C	D
	A	$0,0,0$	$2,0,0$	A	$0,1,0$	$0,0,1$
	B	$0,0,2$	$0,2,0$	B	$1,0,0$	$\lambda_1, \lambda_2, \lambda_3$
		E			F	

This game shows that Lemma 1 do no extend to games with more than two players. As will be shown below, in Game 1 there exists a pure Nash equilibrium, $x = (A, C, E)$, that is a REE independent of the payoffs from the outcome (B, D, F). Note that $\beta(x) = S$.

The set of mixed-strategy profiles of Game 1 can be represented by the unit cube and I identify the strategy profile (A, C, E) with $x = (1, 1, 1)$. For a strategy profile $y \in [0, 1]^3$ to be an equilibrium entrant to x for any $\bar{\varepsilon} \in (0, 1)$, y has to be a best reply to $z_\varepsilon = (1 - \varepsilon)x + \varepsilon y$ for all players and some $\varepsilon \in (0, \bar{\varepsilon})$. The probability that i plays x_i given z_ε is equal to $v_i = 1 - \varepsilon(1 - y_i)$.

Using this notation:

1. For player 1, B is a best reply to z_ε if $2v_3(1 - v_2) \leq v_2(1 - v_3) + \lambda_1(1 - v_2)(1 - v_3)$.
2. For player 2, D is a best reply to z_ε if $v_1(1 - v_3) \leq 2v_3(1 - v_1) + \lambda_2(1 - v_1)(1 - v_3)$.
3. For player 3, F is a best reply to z_ε if $2v_2(1 - v_1) \leq v_1(1 - v_2) + \lambda_3(1 - v_1)(1 - v_2)$.

Hence, for any mixed-strategy profile $y \in [0, 1)^3$ to be a best reply to z_ε, the following

inequality has to hold

$$4v_1v_2v_3 \leq [v_2 + \lambda_1(1 - v_2)][2v_3 + \lambda_2(1 - v_3)][v_1 + \lambda_3(1 - v_1)]. \tag{2.1}$$

However, when ε tends to zero, the left-hand side of (2.1) approaches 4 whereas the right-hand side approaches 2. Therefore, no such entrant exists. The remaining cases involve one or more players using x_i with probability 1 (hence $v_i = 1$). It is straightforward to check that no such equilibrium entrant exists.

In this game, if $\lambda_i > 0$ for all $i \in N$, then $T^2 = \{B\} \times \{D\} \times \{F\}$ is the unique minimal CES and FES block. If $\lambda_i \leq 0$ for all i, then the whole strategy set is the unique CES and FES block.

■

As is shown in the next example, there exist two-player games containing an EES set with more than one element that does not have support in any minimal CES or FES block.

Example 7. Consider the following symmetric two-player game

		Lh'	Lt'	Ch'	Ct'	Rh'	Rt'
	Lh	$1,1$	$1,1$	$-2,2$	$2,-2$	$0,3$	$-4,-8$
	Lt	$1,1$	$1,1$	$2,-2$	$-2,2$	$-4,-8$	$0,3$
Game 2 :	Ch	$2,-2$	$-2,2$	$1,1$	$1,1$	$5,5$	$5,5$
	Ct	$-2,2$	$2,-2$	$1,1$	$1,1$	$5,5$	$5,5$
	Rh	$3,0$	$-8,-4$	$5,5$	$5,5$	$8,8$	$8,8$
	Rt	$-8,-4$	$3,0$	$5,5$	$5,5$	$8,8$	$8,8$

In Game 2, there exists an EES set that does *not* have support in any minimal CES, nor minimal FES, block. This EES set is given by

$$\Theta = \{(pLh + (1 - p)Lt, qLh' + (1 - q)Lt') : p, q \in [1/4, 3/4]\}.$$

The unique minimal CES block is $T^{ces} = \{Rh, Rt\} \times \{Rh', Rt'\}$ and the minimal FES blocks are of the form $\{t\}$ for $t \in T^{ces}$.

In this game, property $(\boldsymbol{P})$ is equivalent to property $(\boldsymbol{P'})$, so to determine why Θ is an EES set, the only relevant strategies are best replies to Θ. For example, consider the element $\alpha = (\frac{1}{4}Lh + \frac{3}{4}Lt, \frac{1}{4}Lh' + \frac{3}{4}Lt')$ on the boundary of Θ (these are the only elements in Θ with best replies outside of its support). The block of best replies to α is given by $\beta(\alpha) = \{Lh, Lt, Ct\} \times \{Lh', Lt', Ct'\}$. It is easy to check that all equilibrium entrants, that

is, all convex combinations of α and elements in $T^{\beta(\alpha)}$ sufficiently close to α, are in Θ.[4] In fact, this holds for all elements on the boundary of Θ, hence it is an EES set.[5]

As the number of potential FES blocks containing the support of Θ is large, I will briefly sketch the reason why a pair of noticeable such blocks are not FES. The remaining candidates are left to the reader.

For example, consider the block $T^{\Theta} = C(\Theta)$. It is not FES as the strategy profile $z_\varepsilon = (1-\varepsilon)(Lh, Lh') + \varepsilon y$, with y assigning most weight on Rh and Rh' for both players and $\varepsilon > 0$ small, has no best replies in T^{Θ} and $\{Lh\} \times \{Lh'\} = \beta_{|T^{\Theta}}(z_\varepsilon)$. In fact, all blocks that do not contain the first four strategies for each player, have a strategy profile $x \in \square[T^{\Theta}] \setminus \Theta$ with the just described property. Thus, they are not FES. Moreover, $T' = \{Lh, Lt, Ch, Ct\} \times \{Lh', Lt', Ch', Ct'\}$ is not FES as the strategy profile

$$z_\varepsilon = (1-\varepsilon)\left(\frac{1}{2}Ch + \frac{1}{2}Ct, \frac{1}{2}Ch' + \frac{1}{2}Ct'\right) + \varepsilon y,$$

with y assigning most weight to Rh and Rt (Rh' and Rt') for both players and $\varepsilon > 0$ small, has no best replies in T' and $\{Ch, Ct\} = \beta_{i|T'_i}(z_\varepsilon)$. Continuing this exercise for all blocks T with $C(\Theta) \subseteq T$ completes the argument.

■

There exist other multipopulation evolutionary stability concepts defined for arbitrary finite games. For example, there are extensions of evolutionarily stable strategies (Maynard Smith and Price, 1973, see Section 2.6) originally defined for symmetric games using a single population framework. However, these concepts are known to be very demanding. For example, the point-valued extension of Maynard Smith and Price's evolutionary stability to asymmetric games is equivalent to strict Nash equilibrium (Selten, 1980). The analysis of symmetric versions of CES and evolutionary stability is postponed until Section 2.6.

2.4 Forward induction

As an application of the above result, I consider minimal CES and FES blocks' ability to capture the notion of *forward induction*, introduced by Kohlberg and Mertens (1986). Forward induction reasoning starts from the observation that what a player does in an early

[4] The set of Nash equilibria of the block game $G^{\beta(\alpha)}$ is given by $E^{T^\alpha} = \{(pLh + (1-p)Lt, qLh + (1-q)Lt) : p, q \in [1/4, 1]\}$.

[5] Notice that this shows that property ($\boldsymbol{P'}$) is *not* equivalent to the property: if y is Nash equilibrium of the block game G^T for $T = \cup_{x \in X} \beta(x)$, then the strategy profile given by mostly x and a small share of y, is in X.

stage of a multi-stage interaction signals what this player will do later in the game (see, e.g., Govindan and Wilson (2009) for a discussion of the literature and Evdokimov and Rustichini (2016) for experimental evidence). In this context, van Damme (1989) focuses on a simple case of such an interaction. In the two-player outside option games he considers, player 1 first chooses between 'in' or 'out.' If 'out' is chosen, the game ends, and if 'in' is chosen, 1 and 2 play a (generic) normal-form game. This subgame has a single equilibrium that gives more payoff to 1 than her outside option. In such games, forward induction captures the idea that, if 1 chooses 'in' she signals that she plans to play the equilibrium which gives her a higher payoff than what she gets from choosing 'out.'

Consider a generic two-player finite normal-form game $G^* = \langle \{1,2\}, S_1 \times S_2, u \rangle$ representing the simultaneous move subgame that takes place if 1 chooses 'in.' Here, generic means that every Nash equilibrium is *regular* in the sense of van Damme (1991).[6] Among other things, this implies that every Nash equilibrium is quasi-strict, i.e. $C(x) = \beta(x)$ for all $x \in E$, and all Nash equilibrium components are singletons. In addition, let G^* be such that there is a Nash equilibrium x^* of G^* with $u_1(x^*) > u_1(x)$ for any other Nash equilibrium x. Following van Damme (1989), an outside option game is defined as $G^{out} = \langle \{1,2\}, (S_1 \cup \{O\}) \times S_2, u \rangle$ where O is 1's outside option strategy, and $u_1(x^*) > u_1(O, s_2) = u_1(O, s_2') > u_1(x)$ for any Nash equilibrium $x \neq x^*$ of G^* and any $s_2, s_2' \in S_2$. Moreover, $u_2(s_2, O) = u_2(s_2', O)$ for all $s_2, s_2' \in S_2$. Note that the forward induction equilibrium x^* is such that $x_1^*(O) = 0$. I denote this class of outside option games by Γ.

van Damme (1989) argues that any solution concept that is consistent with forward induction should satisfy the following property: for any outside option game $G \in \Gamma$, only the equilibrium x^* is plausible. In their analysis of such games, Hauk and Hurkens (2002) show that whenever an EES set exists, it constitutes the forward induction equilibrium x^*.[7] They also show that no concept based on strategic stability (Kohlberg and Mertens, 1986) satisfies this property.

I here show that in any outside option game $G \in \Gamma$, there exists a unique minimal CES block and that this block includes the support of the forward induction equilibrium x^*. It then follows from Proposition 6 that x^* is the unique equilibrium with support in a minimal CES block whenever an EES set exists.

Proposition 7. *Let $G \in \Gamma$. Then, there exists a unique minimal CES block T and $C(x^*) \subseteq T$.*

[6] This notion is a slight modification of the notion introduced by Harsanyi (1973a).

[7] In the class of games Hauk and Hurkens (2002) considered, all the strategies in the support of x^* gives a higher payoff than the outside option strategy for 1. Thus, I consider a strictly larger class of games. However, note that the class of games illustrated by Game 3 below is a subset of the class of outside option games considered by Hauk and Hurkens (2002).

Proof. As noted by Hauk and Hurkens (2002), since the forward induction equilibrium either has index $+1$ or -1, the outside option component has index 0 or $+2$ (see Ritzberger (1994) for index theory applied to Nash equilibrium components). Now, consider a block T with the property that $\beta(x) \subseteq T$ for all $x \in E^T$, a Nash block in the vocabulary of Wikman (2020). Wikman (2020) shows that the index sum across all Nash equilibrium components in a Nash block has index $+1$. Thus, any Nash block either contains only x^* or both components. (Wikman (2020) shows that any equilibrium component is either contained in or disjoint from the face spanned by a Nash block.) The remainder of the proof establishes that Nash and CES blocks only differ in inessential ways (as made precise below) in the class of games Γ. Thus, it must be the case that any CES block also only contains x^* or Nash equilibria from both components.

Consider any minimal CES block T' that contains O for 1 but not the support of x^*. It must contain the support of at least one Nash equilibrium of the outside option component since otherwise it contains no Nash equilibria, which is impossible. It is straightforward to show that all strategies that are not weakly dominated for 2 are included in this block. To see this, assume that s_2 is not weakly dominated and not included in the block T for 2. Then, for any $\bar{\varepsilon} > 0$ there exists an $\varepsilon \in (0, \bar{\varepsilon})$ with $x_1 = (1-\varepsilon)O + \varepsilon y_1$ where $y_1 \in \Delta(S_1)$ is such that $u_2(s_2, y_1) > u_2(s_2', y_1)$ for all $s_2' \neq s_2 \in \Delta(S_2)$. Such a strategy exists for s_2 since it is not weakly dominated. Consider now the strategy profile $x = (x_1, x_2)$ for $x_2 = (1-\varepsilon)z_2 + \varepsilon y_2$ where $(O, z_2) \in E^T$ such that $u_1(O, z_2) > u_1(s_1'', z_2)$ for any other $s_1'' \in T_1$ (such a strategy z_2 must exist unless O is weakly dominated which is impossible since O is the only strategy for 1 used in any of the equilibria with support in the CES block). By bilinearity of the payoff function, if $\varepsilon > 0$ is small enough $C(O, z_2) \subseteq \beta_{|T}(x)$ and $u_2(s_2, x_1) > u_2(s_2', x_1)$ for all $s_2' \neq s_2 \in \Delta(S_2)$, showing that T is not a CES block.

Consider now the outside option game, G^w which is obtained after removing all pure strategies in G^{out} that are weakly dominated for 2 . Clearly, T' is a CES block in this game too since the block includes all the dominating strategies: only weakly dominated strategies were removed which, by the genericity of the subgame, could not have been used in any equilibrium of the block game. Thus, the resulting new subgame G^{*w} for G^w is still generic and cannot have new equilibria. Since CES and Nash blocks can be shown to agree on generic games (Wikman, 2020), it must be the case that T' is a Nash block in this game too since all the best replies to the outside option component are included in the block in the new game, so all that matters is the properties of the subgame. But this implies that there exists an outside option game in which there is a Nash block not including x^*, a contradiction.

□

There exist outside option games in which the unique equilibrium with support in a

minimal CES block is x^* when even no EES set exists, and, moreover, there exist outside option games in which the whole strategy space is the unique CES block. Thus, although the CES notion does not perfectly adhere to van Damme's (1989) definition of forward induction, it comes, arguably, closer than does EES.

Example 8. Fix $\alpha \in (-1,0) \cup (0,1)$ and consider the game

		Lh	Ch	Rh	Lt	Ct	Rt
	O	$2,2$	$2,2$	$2,2$	$2,2$	$2,2$	$2,2$
Game 3 :	U	$7,7$	$4,0$	$3,0$	$\alpha,9$	$-4,-2$	$0,-1$
	M	$3,0$	$7,7$	$4,0$	$0,-1$	$\alpha,9$	$-4,-2$
	D	$4,0$	$3,0$	$7,7$	$-4,-2$	$0,-1$	$\alpha,9$

This game has two Nash equilibrium components regardless of $\alpha \in (-1,0) \cup (0,1)$: the outside option component where 1 plays O and 2 plays any strategy combination with sufficiently small weight on the strategies Lh, Ch, and Rh; and the strategy profile $x^* = (\frac{1}{3}U + \frac{1}{3}M + \frac{1}{3}D, \frac{1}{3}Lh + \frac{1}{3}Ch + \frac{1}{3}Rh)$. This game does not have an EES set, and for any $\alpha \in (-1,0) \cup (0,1)$ there is a unique minimal CES block, which differs depending on the sign of α. If $\alpha \in (-1,0)$, then the unique Nash equilibrium of the block game G^T for $T = (S_1 \setminus \{O\})$ is x^* and the unique minimal CES block is T. If $\alpha \in (0,1)$, then G^T has 7 Nash equilibria: x^*, (U, Lt), (M, Ct), (D, Rt), and the three equilibria which involves mixing the strategies from any pair of the pure equilibria. Therefore, the unique minimal CES block is S.

■

There exist outside option games in which there are more than one minimal FES block, where one of them contains the forward induction equilibrium and another solely contains equilibria from the outside option component (see Myerson and Weibull (2015, Example 3)). Moreover, there also exist outside option games where the unique minimal FES block only contains outside-option equilibria (see Hauk and Hurkens (2002, Fig. 7)).

2.5 The consideration-set framework

In this section, I first present the consideration-set framework developed by Myerson and Weibull (2015) and then show that the CES and FES concepts are equivalent to coarse and fine tenability, respectively.

2.5.1 Preliminaries

Fix any game $G = \langle N, S, u\rangle$—which in this context will be referred to as the *full game*. A *consideration-set game* is a meta game in which there exists a large population of individuals for every player role $i \in N$. Recurrently, an individual from every player role is randomly drawn to play the full game against each other. Every individual is boundedly rational in the sense that she only considers a subset of the set of strategies, S_i, available to her. Such a set of strategies is called the individual's *consideration set* and is identified with her *type* $\theta_i \in \Theta_i = \mathcal{C}(S_i)$, where $\mathcal{C}(S_i)$ is the set of nonempty subsets A_i of S_i. Let μ_i denote any probability distribution over $\mathcal{C}(S_i)$ where the random draws of types are statistically independent. A *type distribution* $\mu = (\mu_1, \ldots, \mu_n) \in \times_{i\in N}\Delta(\mathcal{C}(S_i))$ is a vector of such probability distributions.

Given a type distribution μ, the consideration-set game $G^\mu = \langle N, \times_{i\in N} F_i, u^\mu\rangle$ is a game of incomplete information. Here, a pure strategy is a function $f_i : \mathcal{C}(S_i) \to S_i$ assigning each type $\theta_i \in \Theta_i$ a strategy such that $f_i(A_i) \in A_i$ for all $A_i \in \mathcal{C}(S_i)$. Each player role's set of pure strategies, F_i, is embedded in the unit simplex $\Delta(F_i)$. A consideration-set game G^μ is connected to the full game as every strategy profile $\tau \in \times_{i\in N}\Delta(F_i)$ in G^μ induces a strategy profile $\tau^\mu \in \square[S]$ in G. The conditional probability distribution over the strategies in S_i given a type $\theta_i = A_i$, is denoted $\tau_{i|A_i}$. Hence, the probability that player i will use the pure strategy s_i in the strategy induced by τ_i is

$$\tau_i^\mu(s_i) = \sum_{A_i\in\mathcal{C}(S_i)} \mu_i(A_i)\cdot\tau_{i|A_i}(s_i).$$

The payoff functions in G^μ are defined by $u_i^\mu(\tau) = u_i(\tau^\mu)$ for all $i \in N$. A consideration-set game is finite and has at least one Nash equilibrium. In addition, τ is a Nash equilibrium in a consideration-set game if and only if, for all $A_i \in \mathcal{C}(S_i)$ and $i \in N$,

$$\mu_i(A_i) > 0 \implies C(\tau_{i|A_i}) \subseteq \beta_{i|A_i}(\tau^\mu). \tag{2.2}$$

The first block concept defined within this framework is called coarse tenability. It defines a set of conventional strategies (a block) such that, if almost all individuals are conventional in the sense that they only consider the set of conventional strategies, then in any Nash equilibrium of such a consideration-set game, a conventional individual is at least as well off as any other type of individual.

Definition 8 (**Myerson and Weibull, 2015**)**.** A block T is *coarsely tenable* if there exists an $\bar\varepsilon \in (0, 1)$ such that $T \cap \beta(\tau^\mu) \neq \emptyset$ for every type distribution μ with $\mu_i(T_i) > 1 - \bar\varepsilon\ \forall i \in N$

and every Nash equilibrium τ of G^μ.

In addition to the external stability requirement, internal stability is obtained by restricting attention to blocks that are minimal with respect to the above property. Such a block exists in every game. A point-valued solution concept is defined by focusing on Nash equilibria with support in such blocks.

Definition 9 (**Myerson and Weibull, 2015**)**.** A *coarsely settled equilibrium* is any Nash equilibrium that has support in some minimal coarsely tenable block.

Myerson and Weibull (2015) also formalize a less demanding notion of a convention by defining a class of type distributions that are "biased" towards more "rational" types.

Definition 10 (**Myerson and Weibull, 2015**)**.** For any block T and any $\varepsilon \in (0,1)$, a type distribution μ is $\varepsilon - proper$ on T if for every player $i \in N$

$$\begin{cases} (a) \;\; \mu_i(T_i) > 1-\varepsilon, \\ (b) \;\; \mu_i(A_i) > 0 \quad \forall A_i \in \mathcal{C}(S_i), \\ (c) \;\; T_i \neq A_i \subset B_i \in \mathcal{C}(S_i) \quad \Rightarrow \quad \mu_i(A_i) \leq \varepsilon \cdot \mu_i(B_i). \end{cases}$$

Definition 11 (**Myerson and Weibull, 2015**)**.** A block T is *finely tenable* if there exists an $\varepsilon \in (0,1)$ such that $T \cap \beta(\tau^\mu) \neq \emptyset$ for every type distribution μ that is $\varepsilon - proper$ on T and every Nash equilibrium τ of G^μ.

Clearly, every coarsely tenable block is finely tenable. Moreover, Myerson and Weibull (2015) show that every Nash equilibrium in a consideration-set game, with an $\varepsilon - proper$ type distribution on a finely tenable block, induces an ε-proper strategy profile in the full game. Therefore, any limit of a sequence of such strategy profiles when ε tends to zero is a proper equilibrium.

Definition 12 (**Myerson and Weibull, 2015**)**.** A *finely settled equilibrium* is any proper equilibrium that has support in some finely tenable block.

An equilibrium that is both finely and coarsely settled is called *fully settled*.

2.5.2 Results

The characterization results provided here consist of showing that the CES (FES) block property is equivalent to the coarse (fine) tenability block property. The first result follows from the observation that to verify whether a block is coarsely tenable, it suffices to analyze consideration-set games in which every unconventional individual considers a single pure strategy.

Proposition 8. *A block T is coarsely tenable if and only if it is CES.*

Proof. For any Nash equilibrium τ in a consideration-set game, every type is playing a best reply according to her consideration set A_i. Thus, from equation (2.2) the induced strategy profile $\tau^{\mu}_{i|A_i}$ by $\tau_{i|A_i}$ is such that $C(\tau^{\mu}_{i|A_i}) \subseteq \beta_{i|A_i}(\tau^{\mu})$. I now claim that it is without loss of generality to assume that all unconventional individuals have a singleton consideration set. To see this, assume that $\tau^{\mu}_{i,uc}$ is the induced strategy profile by the unconventional individuals in player role i. Then, the same induced strategy profile can be achieved by defining a type distribution for which the consideration sets of the unconventional individual be such that $\tau^{\mu}_{i,uc}(s_i) = \mu'_i(\{s_i\})$ for all $s_i \in S_i$ and $i \in N$ with $\mu'_i(A_i) = 0$ for all $A_i \neq T_i$ that are not a singleton. Moreover, let $\mu_i(T_i) = \mu'_i(T_i)$. It is easy to see that $G^{\mu'}$ has a Nash equilibrium τ' such that $\tau'^{\mu'} = \tau^{\mu}$.

From the above observations, every strategy profile $z = (1-\varepsilon)x + \varepsilon y$ where $C(x) \subseteq \beta_{|T}(z)$, $y \in \square[S]$ and $\varepsilon \in (0,1)$, can be induced by a Nash equilibrium τ in a consideration-set game with $\mu_i(T) > 1-\varepsilon$ for all $i \in N$. The condition for a block to be coarsely tenable, $T \cap \beta(\tau^{\mu}) \neq \emptyset$, is equivalent to $C(x) \subseteq \beta(\tau^{\mu})$ since $C(x) \subseteq T$.

□

The characterization result for fine tenability is more involved. It builds on the fact that every Nash equilibrium of a consideration-set game with an ε-proper type distribution is an ε-proper strategy profile in the full game.

Proposition 9. *A block T is finely tenable if and only if it is FES.*

Proof: See the Appendix.

2.6 A single-population approach

To complete the formal connections between tenability and evolutionary stability, I here consider situations when individuals from a *single population* are uniformly randomly matched with each other to play a *symmetric* game. For such situations, the notion of evolutionarily stable strategies, or ESS, was introduced by Maynard Smith and Price (1973). Although this notion can be defined for N-person games, I restrict for simplicity attention to two-player games. In addition to ESS, I will also consider the weaker notion of neutral evolutionary stability, or NSS, by Maynard Smith (1982) and the set-valued notion of evolutionarily stable, or ES, sets by Thomas (1985).[8]

A symmetric and finite two-player game is defined by $G = \langle\{1,2\}, S, u\rangle$ with $S = K^2$ and $u_2(s_2, s_1) = u_1(s_2, s_1)$ for all $(s_1, s_2) \in K^2$. Write $\pi(x, y)$ for the payoff to a player from playing $x \in \Delta(K)$ when the other player plays $y \in \Delta(K)$. A *symmetric* block T is such that $\emptyset \neq T \subseteq K$. An example of such a block is the set of best replies to any $x \in \Delta(K)$ defined by $\beta(x) = \{s \in K : \pi(s, x) \geq \pi(s', x) \text{ for all } s' \in K\}$.

Definition 13 (**Maynard Smith and Price, 1973; Maynard Smith, 1982**)**.** $x \in \Delta(K)$ is an *evolutionarily stable strategy* if there exists an $\bar{\varepsilon} \in (0, 1)$ such that for any $y \in \Delta(K)$ and $\varepsilon \in (0, \bar{\varepsilon})$,

$$\pi[x, (1-\varepsilon)x + \varepsilon y] > \pi[y, (1-\varepsilon)x + \varepsilon y].$$

It is a *neutrally stable strategy* if the strict inequality is replaced with a weak one.[9]

Definition 14 (**Thomas, 1985**)**.** A nonempty and closed set $X \subseteq \Delta(K)$ is an *evolutionarily stable set* if each $x \in X$ has some neighborhood $U \subseteq \mathbb{R}^{\Pi_{i \in N} m_i}$ such that for any $y \in U \cap \Delta(K)$, $\pi(x, y) \geq \pi(y, y)$ with strict inequality if $y \notin X$.[10]

Here, I consider a symmetric version of coarse tenability. To highlight symmetric coarse tenability's connection to evolutionary stability, the concept is formalized similarly to the CES characterization.

Definition 15. A symmetric block T is *symmetric coarsely tenable* if there exists an $\bar{\varepsilon} > 0$

[8] Balkenborg (1994) proposes an asymmetric version of evolutionarily stable sets which is quite demanding. It can be shown that in two-player games, the support of such a set constitutes a CES block.

[9] The present definition of evolutionary stability is different but equivalent to Maynard Smith and Price's (1973) definition for symmetric two-player finite normal-form games (see Hofbauer et al. (1979)).

[10] Again, this definition of evolutionarily stable sets is different but equivalent to Thomas' (1985) definition (see Weibull (1995)).

such that for any $\varepsilon \in (0, \bar{\varepsilon})$, $y, \hat{y} \in \Delta(K)$ and $x, z \in \Delta(T)$

$$\pi[x, (1-\varepsilon)x + \varepsilon y] \geq \pi[z, (1-\varepsilon)x + \varepsilon y] \Longrightarrow \pi[x, (1-\varepsilon)x + \varepsilon y] \geq \pi[\hat{y}, (1-\varepsilon)x + \varepsilon y].$$

In words, given that the population consists mostly of the incumbents and a small share of entrants, if the incumbent strategy profile is a (mixed) best reply among those with support in T, then it is also a best reply in the entire game.

The results in this section states that the set of best replies to any ES set constitutes a symmetric coarsely tenable block, and every singleton symmetric coarsely tenable block is a pure NSS. Since every singleton ES set is an ESS, this implies that the best replies to any ESS is a symmetric coarsely tenable block. It shows that the latter's robustness properties lie between those of the two evolutionary stability concepts albeit related to different objects, strategy blocks versus (sets of) mixed-strategy profiles.

Proposition 10. *Let G be a symmetric two-player game.*

1. *If $X \subseteq \Delta(K)$ is an ES set, then $T = \cup_{x \in X} \beta(x)$ is symmetric coarsely tenable.*
2. *If $\{t\}$ is a symmetric coarsely tenable block, then $t \in \Delta(K)$ is a NSS.*

Proof. (i) First, from Theorem 3 in Balkenborg and Schlag (2001), there exists an $\bar{\varepsilon} \in (0, 1)$ such that for all $x \in X$, $\pi(x, (1-\varepsilon)x + \varepsilon y) > \pi(y, (1-\varepsilon)x + \varepsilon y)$ for all $\varepsilon \in (0, \bar{\varepsilon})$ and $y \notin X$. Now, if $y \in E^T$ for $T = \beta(x)$ and $x \in X$, then $y \in X$. To see this, note that, among the strategies in T, y is a (mixed) best reply to both y and x, implying that $\pi(x, (1-\varepsilon)x + \varepsilon y) \leq \pi(y, (1-\varepsilon)x + \varepsilon y)$ for all $\varepsilon \in [0, 1]$. Thus, $(1-\varepsilon)x + \varepsilon y \in X$ for $\varepsilon \in (0, \bar{\varepsilon})$. Since y is a weakly better reply than x to both x and y, by the bilinearity of the payoff function in two-player games, again $(1-\epsilon)((1-\varepsilon)x + \varepsilon y) + \epsilon y \in X$ for any $\epsilon \in (0, \bar{\varepsilon})$. Using the same argument repeatedly gives the result as X is closed. This implies that $X = E^T$ for $T = \cup_{x \in X} \beta(x)$, which implies that T is symmetric coarsely tenable (see Wikman (2020)).

(ii) Let $\{t\} \subset K$ be a symmetric coarsely tenable singleton block. Then there exists an $\bar{\varepsilon} \in (0, 1)$ such that $\pi(t, z) \geq \pi(y', z)$ for all $y' \in \Delta(K)$ and all strategies $z = (1-\varepsilon)t + \varepsilon y \in \Delta(K)$ with $y \in \Delta(K)$ and $\varepsilon \in (0, \bar{\varepsilon})$. Since the inequality has to hold for all $y' \in \Delta(K)$ not only for y, it is immediate that t is a neutrally stable strategy.

□

Since coarse tenability is a strictly more demanding robustness requirement than its symmetric counterpart, the claim made by Myerson and Weibull (2015)—that every symmetric singleton block that is coarsely tenable constitutes a NSS—follows as a corollary to

Proposition 10.[11]

Swinkels (1992b) also developed a symmetric version of his evolutionary stability notion. Analogous to the result in Section 2.3, in symmetric two-player games the set of best replies to a symmetric REE is a symmetric CES block. In contrast to the asymmetric definition, this block does not necessarily consist of the support of the REE. This is shown in Game 4 below. Moreover, there exist games with symmetric EES sets without support in any minimal symmetric coarsely tenable block. Game 3 above is such an example.

Example 9. Consider the following symmetric two-player game due to Xu (2019)

		L	C	R
	L	$3,3$	$3,3$	$0,0$
Game 4 :	C	$3,3$	$2,2$	$1,0$
	R	$0,0$	$0,1$	$0,0$

Game 4 has a unique symmetric Nash equilibrium L which is also the unique ESS. Note that L does not constitute a symmetric coarsely tenable block since C is the unique best reply to $(\gamma L + (1-\gamma)R)$ for all $\gamma \in (0,1)$. Instead, the unique minimal (symmetric) coarsely tenable block is $\beta(L) = \{L, C\}$. The unique minimal finely tenable block is $L \times L$.[12]

■

2.7 Conclusion

The notion of evolutionarily stable strategies is a cornerstone in evolutionary game theory. However, as is well known, the concept is very demanding and such strategies do not exist in important classes of games. The need for a set-valued weakening of the concept has been recognized by many, including Thomas (1985) and Swinkels (1992b). Although there are results showing that such sets are consistent with strong notions of rationality (see, e.g., Swinkels (1992a)), these sets still fail to exist in many games. A case in point is the class of outside option games studied in this paper. Even though EES sets are consistent with forward induction, there are outside option games in which such sets do no exist.

In the present paper, the analysis and characterizations of Myerson and Weibull's (2015) tenability concepts indicate that these concepts might help fill this void. Not only can

[11] See Xu (2019) for an in-depth study of tenable strategy blocks, evolutionary stability, and strategic stability in finite symmetric two-player games.

[12] This observation highlights the fact that Proposition 10 (i) can be strengthen to read that (in symmetric two-player games) the support of an ES set is *symmetric* finely tenable (with the latter definition being analogous to the definition of symmetric coarse tenability).

they credibly be interpreted as evolutionary stability properties but also they exist in every finite normal-form game. The former is highlighted by the fact that, in two-player games, coarse and fine tenability are shown to be generalizations of both REE and ES sets both for symmetric and asymmetric versions of the concepts. While EES set weakens Maynard Smith and Price's (1973) notion of evolutionary stability by requiring the entrants to be "rational," coarse and fine evolutionarily stability instead weaken it by requiring the incumbents to be "boundedly rational" in the sense that they only use best replies among those with support in the evolutionarily stable block in question. This allows the CES concept to capture forward induction much in the same way as EES does while offering cutting power in all outside option games.

2.8 Appendix: Proof of Proposition 9

The proof strategy is to show that if a block does not satisfy one of the properties it does not satisfy the other.

$\Leftarrow$: I establish that if a block is not finely tenable then it is not FES. Then, for every $\bar{\varepsilon} \in (0,1)$ there exists an ε-proper type distribution for $\varepsilon \in (0, \bar{\varepsilon})$ on T with a Nash equilibrium τ in the consideration-set game inducing a strategy profile $\tau^\mu = (1-\epsilon_i)x + \epsilon_i y$ with $\epsilon_i \in (0, \varepsilon)$ and $C(x) \cap \beta(\tau^\mu) = \emptyset$. As $C(x) \subseteq \beta_{|T}(\tau^\mu)$, the proof reduces to showing that y is ε-proper relative to x.

My approach mirrors the proof of Proposition 2 in Myerson and Weibull (2015). However, care has to be taken as the block that is studied is not finely tenable. Let $s_i, s_i' \in S_i$ be such that $u_i(\tau^\mu_{-i}, s_i) > u_i(\tau^\mu_{-i}, s_i')$. Notice first that $\mu_i(T_i') > 0$ for any $T_i' \subseteq S_i$. Therefore, there exists at least a $T_i' \subseteq S_i$ such that $s_i' \in \beta_{i|T_i'}(\tau^\mu)$. Denote

$$\mathcal{T}(T_i) = \{T_i' \subseteq S_i : s_i' \in \beta_{i|T_i'}(\tau^\mu) \wedge T_i' \neq T_i\}.$$

This set is nonempty if $T_i \neq \{s_i'\}$. For each $T_i' \in \mathcal{T}(T_i)$, it follows that:

$$\begin{aligned} &(i) \quad \{s_i\} = \beta_{i|T_i' \cup \{s_i\}}(\tau^\mu), \\ &(ii) \quad \mu_i(T_i') \leq \varepsilon \cdot \mu_i(T_i' \cup \{s_i\}), \text{ and} \\ &(iii) \quad \sum_{T_i' \in \mathcal{T}(T_i)} \mu_i(T_i' \cup \{s_i\}) \leq y_i(s_i). \end{aligned}$$

This implies that $y_i(s_i') \leq \sum_{T_i' \in \mathcal{T}(T_i)} \mu_i(T_i') \leq \varepsilon \cdot y_i(s_i)$. Moreover, it must be the case that $\mu_i(T_i) > 1 - \varepsilon$ so τ^μ can be written as $\tau_i^\mu = (1-\epsilon_i)x_i + \epsilon_i y_i$ where $x_i \in \beta_{i|T_i}(\tau^\mu)$ with $\epsilon_i = \mu_i(T_i) - 1$. The last issue that needs to be addressed is when $T_i = \{s_i'\}$ and s_i' is the worst possible strategy against τ^μ, as in this case $y_i(s_i') = 0$. When this occurs, notice that τ_i^μ can be rewritten as $(1-\delta_i)x_i + \delta_i y_i'$ with $y_i' = \frac{\epsilon_i}{\delta_i} y_i + \frac{\delta_i - \epsilon_i}{\delta_i} x_i$ for $\delta_i \in (\epsilon_i, \varepsilon)$. Hence, it is clear that the induced y (y') is ε-proper relative to x.

$\Rightarrow$:

Assume that for every $\bar{\varepsilon}$, there exists an $\epsilon \in (0, \bar{\varepsilon})$ such that y is *ϵ-proper relative to x* with $C(x) \subseteq \beta_{|T}(z)$ but $C(x) \nsubseteq \beta(z)$ with $z_i = (1-\delta_i)x_i + \delta_i y_i$ for $\delta_i \in (0, \epsilon)$. Fix $\varepsilon \in (0,1)$ and note that the task is now to construct an ε-proper type distribution such that $\tau^\mu = z$ with the just-described properties. Let $\epsilon \cdot 2^{\max_{i \in N} |S_i|} = \varepsilon$ for the associated z. Assume that $|T_i| > 1$ for all $i \in N$ (the case when $|T_i| = 1$ for some i can be dealt with in the same way as in the proof of the other direction). Let $\mu_i(T_i) = 1 - \delta_i$.

For all i define $[s_i]^k \subset S_i$ for $k = 1, \ldots, n$ by $s_i', s_i'' \in [s_i]^i$ implies $u_i(s_i', z_{-i}) = u_i(s_i'', z_{-i})$ and $s_i' \in [s_i]^k, s_i'' \in [s_i]^l$ for $k < l$ implies $u_i(s_i', z_{-i}) > u_i(s_i'', z_{-i})$. Since $\mu_i(T_i') \cdot \varepsilon \geq \mu_i(T_i'')$

for any $T_i', T_i'' \neq T_i$ where $T_i'' \cap [s_i]^1 = \emptyset$ and $T_i' \cap [s_i]^1 \neq \emptyset$. Thus, $y_i(s_i'')\varepsilon \cdot 2^{|S_i|} \geq y_i(s_i')$ for $s_i' \notin [s_i]^1$ and $s_i'' \in [s_i]^1$ is consistent with such a $\varepsilon - proper$ type distribution as there are less than $2^{|S_i|}$ of sets of strategies for i that includes s_i' where it is optimal against z. Moreover, the relative usage of each strategy in $[s_i]^1$ can be made arbitrary large since the individuals with consideration set S_i are indifferent between all strategies in $[s_i]^1$ and all sets of the form $S_i \setminus \{s_i'\}$ for $s_i' \in [s_i]^1$ are non-nested so they can be given arbitrary unequal weight relative to each other. Thus, μ can be chosen such that for all $s_i' \in [s_i]^1$, $\tau_i^\mu(s_i') = y_i(s_i')$.

The argument is identical for each class $[s_i]^k$ for $k > 1$ excluding the blocks including strategies in $[s_i]^1$. Conclude that T is not finely tenable as for any $\bar{\varepsilon}$ there is a $\varepsilon-$proper type distribution with $\varepsilon \in (0, \bar{\varepsilon})$ such that $\tau^\mu = z$ which implies that $\beta(\tau^\mu) \cap T = \emptyset$. □

Chapter 3: Anticipation-dependent preferences

Abstract

This paper develops a model of a decision-maker who dynamically evaluates outcomes as gains and losses relative to an endogenously determined reference point. The model can be interpreted as if the decision-maker optimally chooses each reference point given a trade-off between anticipatory utility and loss aversion: anticipating a better outcome jointly increases current utility and the reference point for tomorrow's outcome. The main result is an axiomatic characterization of such preferences over infinite-horizon temporal lotteries. The obtained utility representation has a recursive form that is amenable to dynamic programming, and it provides an operational welfare criterion. I show that it is possible to uniquely identify the decision-maker's utility representation without observing her current reference point. Finally, the model is applied to asset pricing and life-cycle consumption. In the first application, the model can account for a sizable equity premium together with low risk aversion and low aversion to temporal resolution of uncertainty. In the second application, it can account for excess sensitivity and smoothness of consumption responses to permanent income shocks, which vanish for large shocks.

3.1 Introduction

Experimental evidence and introspection suggest that people do not experience outcomes on an absolute scale but instead relative to a point of reference determined by past experiences.[1] Reference dependence is not only a core topic in behavioral economics but also widely influential in the literature on dynamic choice under uncertainty. The latter serves as the foundation for a variety of fields, such as macroeconomics, finance, and labor economics. In particular, these fields have been heavily influenced by models in which the reference point is a function of past consumption—so-called habit formation models.

An issue with habit models, which is shared with most applications of reference-dependent preferences, is that reference point formation has largely been up to the discretion of the modeler.[2] Not surprisingly, this powerful degree of freedom has led to substantial disagreement in predictions even within narrowly defined classes of models.[3] In my paper, I address this issue by proposing a novel dynamic model describing decision-makers whose preferences depend on past experiences.

My analysis starts from preferences that are conditional on *recent anticipations*. The main result is an axiomatic characterization of choices over infinite-horizon temporal lotteries. This characterization provides necessary and sufficient conditions for a decision-maker to act *as if* dynamically evaluating each period's outcome as a gain or a loss relative to a reference point, which is endogenously formed by her recent anticipations.[4] The axioms provide testable implications for choice behavior in the presence of unobservable reference points. Furthermore, I show that it is possible to uniquely recover the decision-maker's utility representation and thus how her reference points are formed without observing the latter.

Deriving my model from restrictions on observable choice behavior not only gives it a strong theoretical foundation but also allows it to benefit from the entire neoclassical toolbox

[1] This notion was first suggested in economics by Markowitz (1952).

[2] See ODonoghue and Sprenger (2018) for a recent review.

[3] Habit models are typically divided into two types: models in which the habit is *intrinsic*, i.e., the habit stock depends on past consumption (Ryder and Heal, 1973), and models in which it is *extrinsic*, i.e., it depends on society's average consumption (Abel, 1990). In the latter class of preferences, Ljungqvist and Uhlig (2015) show that whether the habit is intrinsic or extrinsic plays a large role in optimal governmental policies in the canonical asset pricing setting analyzed by Campbell and Cochrane (1999).

[4] Kőszegi and Rabin (2006).

(Backus et al., 2004). Conditional on date and anticipations, the decision-maker's preferences are identical across time, and the utility representation has a recursive form that is amenable to dynamic programming. As a result, the model is tractable and portable across many different contexts. This makes it possible to evaluate the economic significance of notions related to reference dependence that have been very successful in describing the behavior of people in the laboratory.

Moreover, and perhaps more important, the model suggests *an operational welfare criterion* to evaluate the effects of changing policies or circumstances—a topic that has often been neglected in the literature on reference-dependent preferences (ODonoghue and Sprenger, 2018). Since the conditional preferences agree with each other, this welfare measure is unambiguous. I show by way of example and applications of the model that gain-loss utility can be given normative weight without counterintuitive implications.

The choice domain of infinite-horizon temporal lotteries is rich and also used to describe the decision-maker's anticipations about future outcomes. The novelty of the axiomatization is that I impose my behavioral axioms on the decision-maker's preferences, conditional on her anticipation from the previous period about the same choice. This approach stands in stark contrast to the axiomatizations of habit formation models by Rozen (2010) and Tserenjigmid (2019) in which preferences are conditional on the history of *past consumption*. My forward-looking preferences introduce novel difficulties that I address by modifying existing axioms. I then provide a novel axiom that captures the idea that outcomes are evaluated relative to anticipated consumption levels.

The main axiom, called anticipation, consists of two parts. The first part specifies how conditional preferences relate to one another. It suggests a novel way to compensate a decision-maker for anticipating a better outcome by increasing the current consumption level to make the utility from the outcome comparable to the utility from consumption when anticipating a worse outcome.[5] A measure of the level of anticipations is provided by the notion of *intrinsic* consumption preferences that are independent of anticipations. This notion has frequently been used in the behavioral literature (see, e.g., Bell (1985)) but has never been given axiomatic foundations. I define intrinsic preferences as the preferences over

[5]See Schmidt (2003), Rozen (2010) and Neilson (2006) for related axioms.

consumption outcomes when the level of consumption is known well in advance.

The second part of reference dependence implies that every anticipation is mapped to a consumption level and that conditional preferences does not play a role in choices that induce the same anticipation level. This axiom is motivated by a substantial behavioral literature on reference dependence (see, e.g., Kahneman and Tversky (1979)). The rest of the axioms are well known in the literature on dynamic choice initiated by Kreps and Porteus (1978). I impose them to bring the model as close as possible to the standard time- and state-separable expected utility model. This is done to highlight the implications of reference dependence stemming solely from the introduction of endogenous reference points formed by anticipations.

My utility representation reveals several new insights. First, utility from anticipating an outcome is a crucial component in the formation of the reference point. Specifically, the reference point for tomorrow's outcome is formed today by a trade-off between two conflicting effects. The first effect can be interpreted as if the decision-maker derives felicity from anticipatory feelings about tomorrow's outcome. The second effect stems from the expected utility from tomorrow's outcome being measured relative to the anticipated outcome. Thus, the representation can be interpreted as if the reference point is chosen by the decision-maker to maximize the utility from tomorrow's outcome by trading off anticipatory utility against the risk of being disappointed by the realization.

Second, my preferences are loss averse.[6] Thus, the verification of my axiomatization can also be taken as evidence of loss aversion. This is important because recent reviews of experimental evidence on loss aversion suggest that earlier evidence in support of the phenomenon might have been over-interpreted (see Gal and Rucker (2018) and Yechiam (2018)). Third, my model formalizes an intuitive connection between status quo bias and relative risk aversion, which is separate from intertemporal substitution preferences.[7] It is shown that in my setting, only two axioms, completeness (preferences can rank every pair of elements) and dynamic consistency (conditional preferences are identical), are enough two imply that the decision-maker is status quo biased.

[6]Loss aversion refers to people's tendency to prefer avoiding losses to acquiring same-sized gains (Kahneman and Tversky, 1979).

[7]Status quo bias refers to a preference for the current state of affairs (Samuelson and Zeckhauser, 1988).

A fourth insight is that reference point formation is heterogeneous and inherently linked to consumption preferences: (i) any two utility representations that form the same reference points given the same anticipations must represent the same preferences; (ii) the decision-maker's preferences over consumption, conditional on the reference point, generically determine how future reference points are formed. The latter feature eliminates the degree of freedom associated with having the reference point formed by a mechanism that is external to consumption preferences. Finally, the behavioral literature has noted that small changes in experimental procedures may have significant effects on perceptions of gains and losses (ODonoghue and Sprenger, 2018). My model can account for this phenomenon without counterintuitive welfare implications. In particular, my preferences can exhibit discontinuously changing reference points while preferences over outcomes remain continuous. This implies that such small procedural changes do not significantly affect welfare.

To highlight the tractability and portability of my model, I consider two applications of a special case of my utility representation. This version of my model is one parameter, the coefficient of loss aversion, richer than the standard time- and state-separable model. The first application is an asset pricing model with a representative agent. I show that my preferences can account for a sizable equity premium together with low risk aversion and low aversion to temporal resolution of uncertainty. The latter is noteworthy as a popular class of models in the asset pricing literature, which are capable of explaining a high equity premium, necessarily generates high aversion to temporal resolution of uncertainty (Bansal and Yaron (2004), (Epstein et al., 2014)). This feature of my model is a result of its ability to generate high aversion to small but moderate aversion to medium-stakes risks while maintaining time-separability under uncertainty (Rabin, 2000).

The second application is a simple infinite-horizon life-cycle model. I focus on behavior associated with permanent income shocks. In particular, I derive a closed-form consumption function consisting of the sum of a permanent income term and a biased precautionary savings term. The properties of this function imply that the model can explain excess sensitivity and smoothness of consumption responses to permanent income shocks. Moreover, this feature vanishes for large shocks—a prediction that is consistent with the so-called *magnitude hypothesis* (Jappelli and Pistaferri, 2010). The reason that my preferences can explain this

behavior is because they are status quo biased and, therefore, can account for an endowment effect for consumption.

The remainder of the paper is organized as follows. In the upcoming subsection, I review the related literature on dynamic choice under uncertainty. In subsection 3.1.2, I provide two simplified examples of the model to facilitate intuition. The framework, representation and axioms are introduced in Section 3.2. In Section 3.3, I state the representation theorem, provide a uniqueness result and provide results related to risk aversion. Section 3.4 analyzes behavior associated with status quo bias that is compatible with my preferences. In Section 3.5, I discuss an interpretation of the model and illustrate some salient properties by way of examples. Section 3.6 provides results showing that the model is compatible with behavior associated with reference-dependent preferences such as first-order risk aversion. Section 3.7 applies the model to asset pricing and life-cycle consumption. Finally, Section 3.8 concludes the paper with a discussion of my findings.

3.1.1 Related literature

My paper contributes to the behavioral literature on models of reference-dependent preferences. In Kőszegi and Rabin (2006, 2007, 2009), the authors take the essential intuitions, in terms of the functional form of the 'value-function,' from Kahneman and Tversky's (1979) prospect theory as primitive and postulates that "a person's reference point is the probabilistic beliefs she held in the recent past about outcomes" (Kőszegi and Rabin, 2006, p.1134).[8] A key difference between their model and my model is that, in Kőszegi and Rabin (2006), the decision-maker has time-inconsistent preferences in the sense that her preferences ex ante and ex post the formation of the reference point may be misaligned. As a consequence, the reference point is fully endogenized only after the introduction of a solution concept that specifies how the decision-maker predicts her behavior. However, this allows their model to account for a demand for commitment that is absent in my model.[9]

[8] See also Bell (1985), Loomes and Sugden (1986) and Gul (1991) for models in this vein. Relatedly, Ok et al. (2015), Tserenjigmid (2018) and Kıbrıs et al. (2018) axiomatize models that endogenize the reference point by equating it with a 'salient,' possibly unchosen, alternative.

[9] As Kőszegi and Rabin (2006) note, the reliance on solution concepts is in tension with the discipline of the model.

My paper is also related to the large literature on habit formation models (See, e.g., Ryder and Heal (1973), Becker and Murphy (1988), Constantinides (1990), Campbell and Cochrane (1999) Bowman et al. (1999), and Yogo (2008)).[10] Since the reference point is a function of past consumption, these models differ from mine in several ways. First, the timing is different; in habit models, the reference point is fixed when decisions are made, whereas the reference point typically depends on these decisions in my model. Second, habit models are typically not framed in terms of gains and losses.[11] Finally, habit models are expected utility models and, thus, are not able to account for behavior associated with violations of the independence axiom. By contrast, my preferences are compatible with violations of expected utility à la Allais.

My model also relates to the literature that studies decision-making under non-expected utility. This literature was initiated by Kreps and Porteus (1978) and further developed by Epstein and Zin (1989) and Weil (1990), among others. One difference between my model and recursive models is that in the latter, preferences are independent of the history leading up to each choice situation. Thus, these models cannot account for status quo bias. Another difference is that recursive models are not separable across states. As shown below, this is not the case for my preferences. In a recent paper, Sarver (2018) analyzes preferences in the recursive class that are conceptually closely related to my preferences. In his model, it is as if the decision-maker optimally selects her risk attitude from a feasible set that includes, as a special case, reference-dependent preferences. However, in the recursive class of models, risk attitudes refer to uncertainty regarding the entire future consumption stream. Thus, gains and losses are framed in terms of expected *lifetime utility*.

There are also other models in which the decision-maker derives utility from thinking about, or anticipating, future consumption (see, e.g., Loewenstein (1987), Caplin and Leahy (2001), and Kőszegi (2010)). The two papers most closely related to mine are Brunnermeier and Parker (2005) and Gollier and Muermann (2010). In the former paper, the agents derive utility from thinking about expected future utility flows. These agents optimally choose their

[10] Rozen (2010) and Tserenjigmid (2019) provides axiomatic foundations for linear habit models. Other related models include models of consumption commitments and adjustment costs, see, e.g., Chetty and Szeidl (2016).

[11] For example, the most prominent applications of habit models do not allow for consumption below the reference point (see, e.g., Constantinides (1990) and Campbell and Cochrane (1999)).

beliefs in the initial period to maximize average utility. Since the agents take their beliefs as given when making future investment decisions, they have to balance the utility from more optimistic beliefs against the cost of worse decision-making. Thus, in their model disutility from distorting expectations is instrumental, whereas it is intrinsic in my model.

Finally, Gollier and Muermann (2010) consider a decision-maker who faces the same trade-off between anticipatory feelings and the risk of being disappointed as in my model. They consider a two-stage setting in which the decision-maker has reference-dependent preferences regarding the outcome resolved in the second stage. Similar to my model, the decision-maker chooses her degree of optimism in the ex ante stage by trading off optimistic expected ex ante utility against objective expected ex post utility. The decision-maker is punished for being overly optimistic as the reference point, affecting both ex ante and ex post utility negatively, is increasing in the degree of optimism. The main difference between my and Gollier and Muermann's (2010) model is that the latter only consider ex ante choices.

3.1.2 Two illustrative examples

Example 10. The first example highlights the intuition underlying the model in the context of risky choice in a simple two-stage model. Here, the two stages are ex ante and ex post the formation of the reference point. Consider a student who is about to receive his grade on an exam. Since he has time to think about the potential outcomes of the exam before he actually receives the result, he can mentally prepare himself for the outcome. The student's preference is in the class of preferences axiomatized below, and his beliefs about the outcome can be modeled as a probability measure m^1 on a bounded interval $C \subset \mathbb{R}$, representing potential outcomes of the exam. In a special case of the model, his preference can be described by a utility function

$$V(m^1) = \max_{r \in [\min u, \max u]} \left\{ \underbrace{r}_{\text{ex ante anticipatory utility}} + \underbrace{\int \phi(u(c) - r) d\mu(c)}_{\text{expected ex post utility}} \right\}, \tag{3.1}$$

where, in this example, u is strictly increasing, and ϕ is piecewise linear with slope $(1 + \kappa)$ below the origin and $(1 - \kappa)$ above it with $\kappa \in (0, 1)$ and $\phi(0) = 0$.

In equation (3.1), the reference point, $r \in [\min u, \max u]$, is formed as if the decision-maker is trading off anticipatory utility with the risk of being disappointed: if he imagines himself receiving a higher score on the exam his ex ante utility, r, increases (he is content with the grade he anticipates to receive), however, this also increases the likelihood that he will be disappointed by any outcome c such that $u(c) < r$, thus decreasing his expected ex post utility. Note that ϕ is such that when there is no uncertainty, the optimal reference point, r, equals the utility level $u(c)$. Thus, the student is not elated or disappointed by a grade that he knew he would receive. Moreover, the larger κ, the more loss averse the student. It is easy to verify that the optimal reference point is weakly decreasing in κ. This implies that a person who is more harmed by losses tends to anticipate worse outcomes.

□

Example 11. The second example focuses on a slightly more complicated setting consisting of three periods, $0, 1, 2$, each divided into two stages. Consider a consumer who needs to decide how to allocate consumption between periods 1 and 2. Her lifetime wealth W is initially stochastic with three potential outcomes summarized by the state space $\Omega = \{w_l, w_m, w_h\}$, where $w_h > w_m > w_l > 0$. All uncertainty is resolved at the end of period 0. Thus, for any realization $w \in \Omega$, the consumer chooses non-negative consumption levels c_1 and c_2, given an intertemporal budget constraint $c_1 + c_2 = w$.

In this setting, a consumer with reference-dependent preferences (adapted to the three-period setting) forms two reference points, one for each period in which consumption takes place. Crucially for this example, the reference point for period 1 is formed *before* the wealth level is realized in period 0. Once formed, it is taken as given for the consumption decision in period 1. Similarly, the reference point for period 2 is formed in period 1 and is taken as given in period 2. The timing of the model is presented in Figure 3.1.

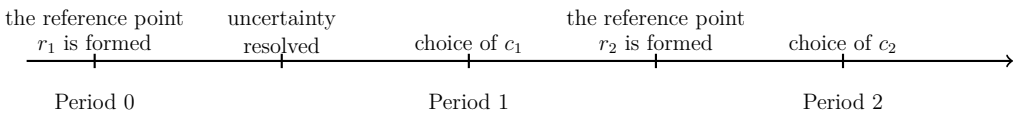

Figure 3.1: The timing of reference point formation and consumption.

The consumer's preferences are in the class of preferences axiomatized below and can be

described by the following additively separable utility function

$$\max_{r_1 \in [\min u, \max u]} \sum_{w \in \Omega} \Big[r_1 + \phi(u(c_1) - r_1) + \max_{r_2 \in [\min u, \max u]} \{ r_2 + \phi(u(c_2) - r_2) \} \Big] p(w),$$

where $p(w)$ is the probability that the wealth level $w \in \Omega$ is realized. For simplicity, there is no discounting, $u' > 0 > u''$ and ϕ is piecewise linear, as in the previous example.

Consider now the consumer's optimal consumption plan. Her consumption-choice problem can be solved by backward induction. In period 2, the consumer always anticipates to consume $c_2 = w - c_1$ and her reference point is thus formed such that $r_2 = u(c_2)$, as w and c_1 are already given. This implies that utility from consumption in period 2 is given by

$$w - c_1 = \max_{r_2 \in [\min u, \max u]} \{ r_2 + \phi(u(w - c_1) - r_2) \}.$$

In contrast, the reference point for period 1 is formed *before the resolution of uncertainty*. What complicates matters is that the reference point itself will affect the level of consumption in period 1 and, therefore, change the nature of the uncertainty. Thus, the reference point in period 1 is given by

$$r_1 \in \arg \max_{r_1 \in [\min u, \max u]} \left\{ r_1 + \sum_{w \in \Omega} \phi(u(c_1(r_1, w)) - r_1) p(w) \right\},$$

where $c_1(r_1, w)$ is such that the first-order condition satisfies

$$\begin{aligned}
&u'(w - c_1) = (1 + \kappa) u'(c_1) && \text{if} \quad u(c_1) < r_1, \\
&u'(w - c_1) = (1 - \kappa) u'(c_1) && \text{if} \quad u(c_1) > r_1, \\
&(1 - \kappa) u'(c_1) < u'(w - c_1) < (1 + \kappa) u'(c_1) && \text{if} \quad u(c_1) = r_1.
\end{aligned}$$

When $p(w_m)$ is large enough relative to $p(w_l)$ and $p(w_h)$, the optimal reference point in period 1 is $r_1 = u(w_m/)$. When the range of wealth outcomes is large, the model produces a novel prediction as a result of anticipatory utility. When the realized wealth is low ($w = w_l$) the decision-maker consumes more in period 1 than in period 2. The *opposite* is true when the realized wealth is high ($w = w_h$). This follows from the optimal consumption level being given

by $(1+\kappa)u'(c_1) = u'(w_l - c_1)$ and $(1-\kappa)u'(c_1) = u'(w_h - c_1)$ for low and high realizations of w, respectively. This behavior is supported by evidence from the literature on consumption (see Jappelli and Pistaferri (2010) for a recent survey and Section 3.7 for a more general analysis). Thus, even though the decision-maker's preferences for consumption are ex ante identical in both periods, the timing of the resolution of uncertainty makes it optimal to not smooth consumption.

When $p(w_m)$ is close to 1, Table 3.1 provides predicted consumption levels given each realization of W for my preferences and related models discussed in the literature review. The model developed by Kőszegi and Rabin (2009) always predicts (weak) overconsumption in period 1 compared to period 2. A consumer with their preferences never benefits, in terms of 'gain-loss utility' relative to the reference point, from allocating more consumption in period 2 after she has learned her wealth level. The reason is that, in any consistent plan, the reference point will equal the planned consumption outcome. The standard time- and state-separable model and recursive models (such as the Epstein-Zin model and Sarver's (2018) model) coincide when there is no uncertainty. Since all decisions are made after the resolution of any uncertainty, the predictions of these models coincide in this setting. This is also the case for the model developed by Brunnermeier and Parker (2005), or BP. In contrast, habit formation models depend on the initial reference point, or 'habit stock,' in period 1 denoted h_0. Thus, fixing the initial habit stock, these models will predict different consumption levels depending on the potential wealth outcomes. The last column refers to a consumer with the same utility function representing my preferences except that the reference points, r_1 and r_2, are exogenously fixed.

Model	$w = w_m$	$w = w_l$	$w = w_h$
This paper	$c_1 = c_2$	$c_1 > c_2$	$c_1 < c_2$
Kőszegi and Rabin	$c_1 \geq c_2$	$c_1 \geq c_2$	$c_1 \geq c_2$
Standard/Recursive/BP	$c_1 = c_2$	$c_1 = c_2$	$c_1 = c_2$
Habit Formation	$c_1 \gtrless c_2$ if $h_0 \gtrless h_0^*$	$c_1 \gtrless c_2$ if $h_0 \gtrless h_0^*$	$c_1 \gtrless c_2$ if $h_0 \gtrless h_0^*$
This paper, fixing $r_1 \geq (\leq) r_2$	$c_1 \geq (\leq) c_2$	$c_1 \geq (\leq) c_2$	$c_1 \geq (\leq) c_2$

Table 3.1: Comparisons of Models

□

3.2 Anticipation-dependent preferences

3.2.1 Framework

An important part of the axiomatic analysis conducted in this paper is the way information regarding consumption is resolved over time. This consideration is important because the decision-maker's preferences may depend on her prior objective anticipations. To distinguish among infinite stochastic consumption streams that only differ in the way risk is resolved, the consumption space requires a complicated mathematical construction. The domain considered in this paper was first analyzed in the theoretical literature on intertemporal utility by Kreps and Porteus (1978) in a finite horizon setting and by Epstein and Zin (1989) in an infinite horizon setting in which consumption in the current period is deterministic.[12]

For any separable metric space X, let $\Delta(X)$ denote the set of all Borel probability measures on X. For technical reasons described in Appendix 3.9.1, I endow $\Delta(X)$ with a metric that is equivalent to the Kantorovich-Rubinstein metric which metricizes the topology of weak convergence. For any $x \in X$, let $\delta_x \in \Delta(X)$ denote the Dirac probability measure associated with x. The Cartesian product of metric spaces is endowed with the product metric.

The environment is presented recursively. Per-period consumption is assumed to lie in a compact and connected metric space (C, d_C), which infinite Cartesian product $C^{\mathbb{N}}$ represents the space of deterministic consumption streams. Following Epstein and Zin (1989), in Appendix 3.9.1 I construct a compact and connected metric space $D \subset \prod_{t=0}^{\infty} D_t$ where $D_0 = C^{\mathbb{N}}$ and $D_t = \Delta(C \times D_{t-1})$ for $t \geq 1$. The metric d_D on D is inherited from $\prod_{t=0}^{\infty} D_t$. It is well-known that D can be identified with $\Delta(C \times D)$ by a linear homeomorphism g. Given this identification, one can think of elements in D as the joint distribution over current consumption C and lotteries over $C \times D$ beginning in the next period. For a more thorough account of the construction, see Epstein and Zin (1989) and Chew and Epstein (1991).

[12] The construction considered in this paper was first studied by Chew et al. (1991). See also Gul and Pesendorfer (2004) for the development of infinite-horizon decision problems.

A generic element $m \in D$ is called an *(infinite-horizon) temporal lottery*. For ease of presentation, let (c, m) be identified with $\delta_{(c,m)} \in D$ and call m in this context a *continuation lottery*. Note that by construction, $C^{\mathbb{N}}$ can be embedded as a subset of D. Denote any such element by $\boldsymbol{c} = (c_1, c_2, c_3, \ldots) \in C^{\mathbb{N}}$. For $\alpha \in (0, 1)$, let $\alpha m + (1 - \alpha)m'$ be the measure that assigns $\alpha m(B) + (1 - \alpha)m'(B)$ to each Borel measurable set B.

For any space X with metric d_X, a function $f : X \to \mathbb{R}$ is Lipschitz continuous if there is some $M > 0$ such that $|f(x) - f(\hat{x})| \leq M d_X(x, \hat{x})$ for every $x, \hat{x} \in X$. In Appendix 3.9.1, I show that g is such that if $f : D \to \mathbb{R}$ is Lipschitz continuous, then $f \circ g : \Delta(C \times D) \to \mathbb{R}$ is Lipschitz continuous.

3.2.2 Representation

I consider (binary) preference relations over infinite-horizon temporal lotteries, describing a decision-maker who knows that her tastes may change over time. The decision-maker's current preference over D depends on her anticipations of future consumption formed in the previous period. This implies that the decision-maker's preferences may undergo a potentially infinite sequence of endogenous preferences change induced from her choice of temporal lottery.

For any (c, m), the continuation lottery $m \in D$ is interpreted as capturing the decision-maker's anticipation of the distribution of consumption from the next period onward. In this interpretation, the decision-maker's preference given the one-period-lagged anticipation, $a \in D$, is described by a conditional preference relation $\succsim_a$ on D. Every such preference relation is a member of the *family of preferences* $\succsim = \{\succsim_a\}_{a \in D}$.[13] I will refer to the index of the decision-maker's current preference relation as her *anticipation*. To reiterate, I assume that the decision-maker's preferences depend on anticipations but not consumption histories or calendar time. As is standard, a function $f : D \to \mathbb{R}$ is said to *represent* $\succsim_a$ when $m \succsim_a \hat{m}$ if and only if $f(m) \geq f(\hat{m})$ for all $m, \hat{m} \in D$.

[13] An alternative way of presenting the same idea is to think of the decision-maker's preference $\succsim$ as being defined on $D \times D$, where $(a, m) \succsim (\hat{a}, \hat{m})$ denotes the preference for facing the temporal lottery m given the lagged anticipation a over the temporal $\hat{m}$ given the lagged anticipation $\hat{a}$. However, the axiomatization only requires that the modeler observes comparisons between temporal lotteries given the same lagged anticipation (i.e., the relation, $\succsim$, just defined is not complete), hence the use of the more compact notation.

The utility representation below requires a couple of properties of per-period utilities.

Definition 16. A function $\phi : [-a, a] \to \mathbb{R}$ (for $a = \max u - \min u$) is *$\beta-$compatible with* u if $\beta \in (0, 1)$, $u : C \to \mathbb{R}$ is Lipschitz continuous, and for all $y \in [\min u, \max u]$, $\phi \circ (u - y) : C \to \mathbb{R}$ is Lipschitz continuous and nonconstant, normalized such that $\phi(0) = 0$, $\beta\phi(y) - y \leq 0$ for all $y \in [-a, a]$ and $\beta\phi(y) = y$ implies $\beta\phi(x) \geq \beta\phi(y + x) - y$ for all $x \in [-a, a]$ with $-a \leq x + y \leq a$.

Given any $\beta \in (0, 1)$, a function that is $\beta-$compatible with *any* Lipschitz continuous functions on C is the *piecewise linear* function

$$\phi(y) = \begin{cases} \lambda\eta y/\beta & \text{for} \quad y \leq 0, \\ \eta y/\beta & \text{for} \quad y > 0, \end{cases} \tag{3.2}$$

where $\lambda\eta \geq 1 \geq \eta > 0$. Other examples are Lipschitz continuous and concave functions with left- and right-side derivatives $\phi'_-(0) \geq 1/\beta$ and $\phi'_+(0) \leq 1/\beta$, respectively. For example, the value function in prospect theory, which is convex for losses, is $\beta-$consistent provided that the derivatives at the origin are finite (Kahneman and Tversky, 1979). Figure 3.2 depicts an example of such a β-compatible function.

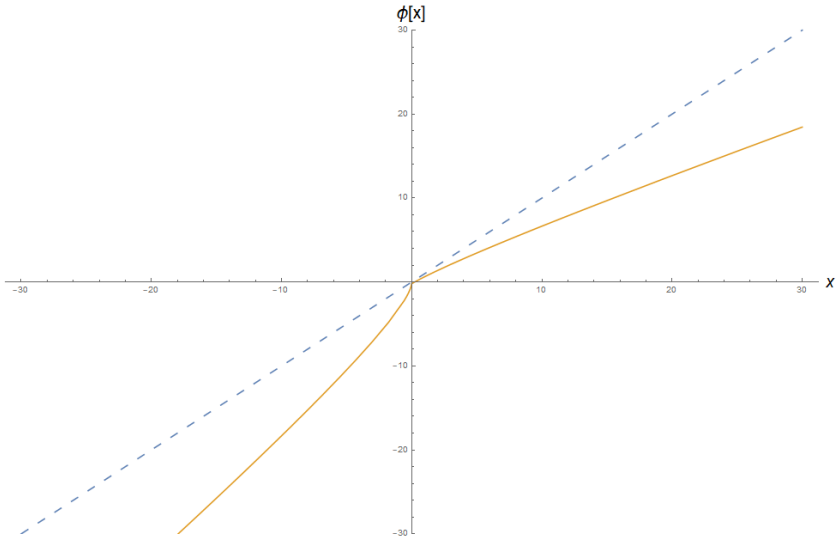

Figure 3.2: A $\beta-$compatible function $\phi(x)$ (solid) must lie below the line with slope $1/\beta$ (dashed).

For any anticipation $a \in D$, let $a^1 \in \Delta(C)$ be the marginal distribution over C. I am now ready to state the utility representation.

Definition 17. A family of preferences $\succsim$ has an *anticipation-dependent* (AD) representation if there exists a tuple (u, ϕ, r, β) where ϕ is nondecreasing and β-compatible with u, and $r : \Delta(C) \to \mathbb{R}$ is continuous at $\delta_c \in \Delta(C)$ with $r(\delta_c) = u(c)$ for all $c \in C$ and

$$r(m^1) \in \underset{r \in [\min u, \max u]}{\arg\max} \left\{ r + \beta \int \phi(u(c) - r) dm^1(c) \right\} \quad \forall m^1 \in \Delta(C) \tag{3.3}$$

such that for every $a \in D$, $\succsim_a$ can be represented by a function $V_a : D \to \mathbb{R}$ defined as

$$V_a(m) = \int \left[\phi(u(c) - r(a^1)) + r(\hat{m}^1) + \beta V_{\hat{m}}(\hat{m}) \right] dm(c, \hat{m}) \quad \forall m \in D. \tag{3.4}$$

It is instructive to 'unpack' the representation given by equation (3.4). To this end, let $\Omega =: C^{\mathbb{N}}$ represent the states of the world, and let $\{\mathcal{G}_t\}_t$ be a filtration on Ω that represents how information accumulates as time progresses in the sense that $\mathcal{G}_0 \subseteq \mathcal{G}_1 \subseteq \mathcal{G}_2 \ldots$. Specifically, let $\mathcal{B}$ be the Borel σ-algebra on C and $\mathcal{G}_0 = \{\emptyset, \Omega\}$ and, for every $t > 0$, $\mathcal{G}_t := \mathcal{B}^t \times \{\emptyset, C\}^\infty$. I show in Appendix 3.9.1 that equation (3.4) can, for all $a, m \in D$, be written in the following additively separable form (where the conditional expectations operator, $\mathbb{E}[\cdot|\cdot]$, is defined by integrating regular conditional probabilities)

$$V_a(m) = \mathbb{E}_m \left[\phi(u(c_1) - r(a^1)) + \sum_{t=1}^{\infty} \beta^{t-1} \max_{r \in [\min u, \max u]} \left(r + \beta \mathbb{E} \left[\phi(u(c_{t+1}) - r) | \mathcal{G}_t \right] \right) \right], \tag{3.5}$$

where m induces a unique probability measure on Ω. In this formulation, I treat the expected utility conditional on the optimal reference point in each period as a random variable, where the initial anticipation, a, is taken as given.

Note that, by construction, the reference point coincides with the actual consumption level when there is no uncertainty. Therefore, for any deterministic consumption stream $\boldsymbol{c} \in D$ with an initial consumption level c_1, such that $u(c_1) = r(a^1)$, V_a reduces to the standard time-separable model $V_a(\boldsymbol{c}) = \sum_{t=0}^{\infty} \beta^t u(c_{t+1})$.

Proposition 11. *Each AD representation (u, ϕ, r, β) and $a \in D$ induces a unique Lipschitz continuous function $V_a : D \to \mathbb{R}$ that satisfies equation (3.4).*

Proof: See Appendix 3.9.3.

3.2.3 Axioms

The quantifier "for all $c, \hat{c} \in C$, all $a, \hat{a}, m, \hat{m}, \bar{m} \in D$, and all $\alpha \in [0, 1]$" in the beginning of each axiom is suppressed throughout this subsection. The axioms are imposed for all members of $\succsim$.

The first three axioms are standard and have been proposed elsewhere. The only novelty is a straightforward addition to the continuity axiom that is peculiar to my setting.

Axiom 1. (Weak Order) $\succsim_a$ is complete and transitive.

Axiom 2. (Strong Continuity)

1. **(vNM Continuity)** If $m \succ_a \hat{m} \succ_a \bar{m}$, then there are $\alpha, \bar{\alpha} \in (0, 1)$ such that

$$\alpha m + (1 - \alpha)\bar{m} \succ_a \hat{m} \succ_a \bar{\alpha} m + (1 - \bar{\alpha})\bar{m}.$$

2. **(L-Continuity)** There are $m^*, m_* \in D$ and $M > 0$ such that if $d_D(m, \hat{m}) \leq \alpha / M$, then

$$\alpha m + (1 - \alpha)m^* \succsim_a \alpha \hat{m} + (1 - \alpha)m_*.$$

3. **(Degenerate Anticipations Continuity)** For any sequence $\{a_n\}_{n=1}^{\infty}$ with limit $(c, \bar{m})$, if $m \succ_{a_n} \hat{m}$ for all n, then $m \succ_{(c,\bar{m})} \hat{m}$.

Axiom 3. (Independence) $m \succ_a \hat{m} \implies \alpha m + (1 - \alpha)\bar{m} \succ_a \alpha \hat{m} + (1 - \alpha)\bar{m}$.

The representation requires a continuity axiom that is stronger than usual.[14] However, the first two parts of the strong continuity axiom are standard in the literature on preferences

[14]The need for the second part of the strong continuity axiom is related to the fact that a compact set in an infinite-dimensional metric space has an empty interior. It is a structural axiom used to solve technical issues. Note that the set of all Lipschitz continuous functions are dense in the set of all continuous functions. Thus, any continuous preference can be arbitrarily approximated by a Lipschitz continuous preference. However, strictly speaking, the axiom rules out natural utility functions such as $f(x) = \sqrt{x}$.

over menus (see Dekel et al. (2007) for a discussion that is also relevant for my setting). The third part of the strong continuity axiom implies that unconditional preferences are continuous at degenerate anticipations.

To present the next axiom, it is helpful to introduce some additional notation. For any $m \in D$, let m^1 be as before and m^2 the marginal distribution over D. For any marginal distributions m^1 and $\hat{m}^2$, denote their product distribution by $m^1 \times \hat{m}^2 \in D$.

Axiom 4. (Strong Separability) For any $c, \hat{c} \in C$ and $m, \hat{m} \in D$,

$$\frac{1}{2}(c,m)+\frac{1}{2}(\hat{c},\hat{m}) \sim_a \frac{1}{2}(c,\hat{m})+\frac{1}{2}(\hat{c},m) \quad \text{and} \quad \frac{1}{2}(c,m)+\frac{1}{2}(c,\hat{m}) \sim_a \frac{1}{2}(c,\hat{m}^1\times m^2)+\frac{1}{2}(c,m^1\times\hat{m}^2).$$

Axiom 4 allows for an additively separable representation. My version of separability is slightly stronger than the standard version of separability (see, e.g., Gul and Pesendorfer (2004)). However, if the dynamic consistency axiom below is replaced by the standard version of the same axiom, strong separability can without loss of generality be replaced by the standard version of separability.

The next axiom was introduced by Kreps and Porteus (1978). PERU is mnemonic for 'preferences for early resolution of uncertainty' and implies that the decision-maker (weakly) prefers having all uncertainty regarding the continuation lottery resolved in the current period.[15]

Axiom 5. (PERU) $\alpha(c,m) + (1-\alpha)(c,\hat{m}) \succsim_a (c, \alpha m + (1-\alpha)\hat{m})$.

The next axiom is an appropriate modification of the standard dynamic consistency axiom to my setting.

Axiom 6. (Dynamic Consistency) $(c,m) \succsim_a (c,\hat{m}) \implies m \succsim_m \hat{m}$.

The present notion of dynamic consistency is inspired by Machina (1989) and allows for violating *consequentialism* (see also McClennen (1988), McClennen et al. (1990)). The latter

[15] I show in Appendix 3.9.4 that if a preference for *late* resolution of uncertainty is maintained instead, the only family of preferences that, in addition, satisfies Axioms 1-4,6-7 is one that is singleton-valued and represents the same preferences as the standard time- and state-separable expected utility model.

condition means, in the interpretation of Machina (1989), that preferences are independent of risk forgone in the past. Relaxing this condition allows for dynamically consistent and separable non-expected utility preferences. This assumption allows for unambiguous welfare comparisons and ensures that it is possible to use dynamic programming to solve the decision-maker's choice problem.[16] See Hanany and Klibanoff (2007, 2009) for a discussion in the context of ambiguity.

The next axiom is novel and consists of two parts. The first part provides a revealed preference theory of compensation a decision-maker for her anticipations. The second part implies that the decision-maker evaluates outcomes relative to an anticipated (deterministic) consumption level.

To define the induced ranking over equiprobable consumption outcomes, $\succsim^I$, fix $c^* \in C$ and $m^*, a^* \in D$ and, for any $c, c', \hat{c}, \bar{c} \in C$, let $c\frac{1}{2}c' \succsim^I \hat{c}\frac{1}{2}\bar{c}$ if and only if

$$\frac{1}{2}(c^*, c, m^*) + \frac{1}{2}(c^*, c', m^*) \succsim_{a^*} \frac{1}{2}(c^*, \hat{c}, m^*) + \frac{1}{2}(c^*, \bar{c}, m^*).$$

The interpretation of this order is elaborated below. Under the axioms in this section, $\succsim^I$ turns out to be independent of the particular c^*, m^* and a^* chosen in the definition.

Axiom 7. (Anticipation)

1. **(Compensation)** Whenever $\bar{c}\frac{1}{2}c_a \succsim^I c\frac{1}{2}c_{\hat{a}}$ and $c'\frac{1}{2}c_{\hat{a}} \succsim^I \hat{c}\frac{1}{2}c_a$ for $c_a, c_{\hat{a}}, c', \bar{c} \in C$, then

$$(c, m) \succsim_{(c_a, a)} (c', \hat{m}) \Longrightarrow (\bar{c}, m) \succsim_{(c_{\hat{a}}, \hat{a})} (\hat{c}, \hat{m}).$$

2. **(Certainty)** For any $a \in D$, there exist $(\hat{c}, \hat{a}) \in D$ such that $\succsim_a = \succsim_{(\hat{c}, \hat{a})}$, and

$$\succsim_m = \succsim_{\hat{m}} \Longrightarrow \alpha(c, m) + (1-\alpha)(c, \hat{m}) \sim_a (c, \alpha m + (1-\alpha)\hat{m}).$$

The first part of the anticipation axiom, called *compensation*, specifies how consumption today can be altered to 'compensate' for differences in anticipated consumption levels. Compensation postulates that to compensate a decision-maker for having anticipated $(\hat{c}, \hat{a})$

[16] Consistency also avoids having to use equilibrium concepts to specify the decision-maker's choice (see, e.g., Kőszegi (2010)).

to be 'as well off' as having anticipated (c, a), the consumption outcome c' given preferences conditional on (c, a) has to be replaced by any $\hat{c}'$ that satisfies $c\frac{1}{2}\hat{c}' \sim^I \hat{c}\frac{1}{2}c'$ given preferences conditional on $(\hat{c}, \hat{a})$.

To understand the intuition behind this part of the axiom, consider the case when $C = [0, 1]$ and the decision-maker's preferences are strictly monotone. A possible notion of reference dependence implies that the preference $\succsim_{(c,a)}$ between two degenerate temporal lotteries (c', m) and $(\hat{c}', \hat{m})$ is the same as the preference $\succsim_{(\hat{c},\hat{a})}$ between $(c' + \hat{c} - c, m)$ and $(\hat{c}' + \hat{c} - c, \hat{m})$.[17]

This notion has to be modified to fit a setting in which C is an abstract consumption space. To this end, I define *intrinsic preferences*, $\succsim^I$, over two equiprobable consumption outcomes to be the preferences over the same equiprobable outcomes when *all uncertainty* is resolved *before* the period in which consumption takes place, holding consumption in all other periods fixed. The notion of intrinsic preferences (or utility) has often been used in the literature (see, e.g., Bell (1985) and Kőszegi and Rabin (2006)). In particular, the present notion of intrinsic utility can be thought of as representing preferences when the decision-maker is allowed to mentally prepare herself for the outcome she faces.

The second part, *certainty*, postulates that every anticipation has a degenerate anticipation generating the same preferences. In addition, if two temporal lotteries (anticipations) induce the same preferences, then the decision-maker is indifferent between early and late resolution of the uncertainty associated with which of these two temporal lotteries she will face tomorrow. This part of the anticipation axiom captures the central intuition that outcomes are evaluated as gains and losses relative to an anticipated consumption level.

[17] This notion of reference independence is used as an axiom by Schmidt (2003). Other related 'compensation' axioms are proposed by Rozen (2010) and Tserenjigmid (2019) in the context of linear habit formation models and Neilson (2006) in terms of risk and other-regarding preferences.

3.3 Main results

3.3.1 Representation theorem and uniqueness

The family of preferences $\succsim$ is said to be *nondegenerate* if, for any $m, a \in D$, there exists $c^*, c_* \in C$ such that $(c^*, m) \succ_a (c_*, m)$.

The representation theorem below provides a recursive, time- and state-separable, but nonstationary representation of a nondegenerate family of preferences satisfying Axioms 1-7.

Theorem 1. *A nondegenerate family of preferences $\succsim$ has an anticipation-dependent representation if and only if it satisfies Axioms 1-7. Moreover, ϕ and β are unique, and u and r are unique up to the same additive constant except when ϕ is piecewise linear.*

Proof: See Appendix 3.9.2.

When ϕ is piecewise linear but not the identity function, then u and r are unique up to a joint affine transformation. If ϕ is the identity function, u is unique up to an affine transformation and r does not affect preferences. This constitutes the special case in which the model reduces to the standard discounted expected utility model.

The above uniqueness result hinges on the observability of the decision-maker's initial anticipation. In practice, it might be difficult to determine what the decision-maker was anticipating *before* attempting to elicit her preferences. To abstract from such considerations, it is possible to hold consumption in the initial period fixed and only study preferences over continuation lotteries. In this setting, one gains nothing from explicitly specifying the reference point map. Thus, it is possible to let the preferences $\succsim_a$ on $\{c^*\} \times D$, for some $c^* \in C$ and any initial anticipation, a, be represented by

$$V_a(c^*, m) = V(m) = \sum_{t=0}^{\infty} \beta^t \max_{r \in [\min u, \max u]} (r + \mathbb{E}\left[\beta\phi(u(c_{t+1}) - r)|\mathcal{G}_t\right]) \quad \forall (c^*, m) \in \{c^*\} \times D. \tag{3.6}$$

Denote such an AD representation by (u, ϕ, β). The question is whether this representation is identified in this more limited setting. The uniqueness result below is weaker than the above result because it does not pin down continuation preferences after some anticipations.

The reason is that, as shown by way of example in Section 3.5, the optimal reference point in equation (3.6) is not always unique.

Theorem 2. *Let $\succsim$ on $\{c^*\} \times D$ have an AD representation (u_1, ϕ_1, β_1). Then, $\succsim$ has an AD representation (u_2, ϕ_2, β_2) if and only if $\beta_1 = \beta_2$ and there exist scalars $\sigma \in \mathbb{R}$, $\alpha > 0$ such that $u_2 = \alpha u_1 + \sigma$, and $\phi_2 = \alpha\phi_1(\alpha^{-1}(\cdot))$.*

Proof: Follows immediately from the proof of Theorem 3 in Appendix 3.9.3.

3.3.2 Relative risk aversion

I here show that AD preferences separates elasticity of intertemporal substitution from relative risk aversion. Specifically, I shown that shape of the ϕ function determines the relative risk aversion of the decision-maker. By contrast, when all the uncertainty regarding consumption is resolved before the period a good is consumed, the model reduces to the standard time- and state-additive model depending on the intrinsic utility function, u, only. Thus, u captures the decision-maker's elasticity of intertemporal substitution.

Since tastes change over time, it makes sense to focus on degenerate temporal lotteries with current consumption held fixed, as the initial reference point is then not affecting risk preferences.

Definition 18. A family of preferences $\succsim^1$ is *more risk averse* than $\succsim^2$ if, for all $(c^*, \boldsymbol{c}) = (c^*, c_2, c_3, \ldots) \in C^{\mathbb{N}}$ and $a, m \in D$,

$$(c^*, m) \succsim_a^1 (c^*, \boldsymbol{c}) \Longrightarrow (c^*, m) \succsim_a^2 (c^*, \boldsymbol{c}).$$

The result below essentially states that (u_1, ϕ_1, β_1) represents preferences that are more risk averse than those represented by (u_2, ϕ_2, β_2) if and only if ϕ_2 pointwise dominates ϕ_1, $\beta_1 = \beta_2$, and u_1 and u_2 are equal up to a constant.

Theorem 3. *Let $\succsim^1$ and $\succsim^2$ have AD representations (u_1, ϕ_1, β_1) and (u_2, ϕ_2, β_2), respectively. Then $\succsim^1$ is more risk averse than $\succsim^2$ if and only if $\beta_1 = \beta_2$, and there exist scalars $\sigma \in \mathbb{R}$, $\alpha > 0$*

such that $u_2 = \alpha u_1 + \sigma$ *and* $\phi_2(\alpha x) \geq \alpha \phi_1(x)$ *for all* $x \in [-a, a]$ *with* $a = \max u_1 - \min u_1$,.

Proof: See Appendix 3.9.3.

3.3.3 An alternative representation and proof sketch of theorem 1

The subsection focuses on technical details regarding Theorem 1 and highlights the role of the axioms introduced above. It can be skipped without loss of continuity by readers who are anxious to move on to the behavioral analysis.

To highlight the implications of Axiom 5 in the above representation theorem, I will here consider a slightly less general class of preferences but without imposing axiom 7.

I denote by $C(D)$ the set of all continuous functions on D.

Definition 19. A function $V : D \to \mathbb{R}$ is *Gateaux differentiable* at $m \in D$ if there exists a function $u_m \in C(I)$ such that for each $\hat{m} \in D$,

$$\lim_{\theta \downarrow 0} \frac{V((1-\theta)m + \theta\hat{m}) - V(m)}{\theta} = \int u_m(c, m')d(\hat{m} - m)(c, m'). \tag{3.7}$$

A function V is Gateaux differentiable, or smooth, if it is Gateaux differentiable at each $m \in D$.

Axiom 8. (Smooth Utility) For all $m, a \in D$,$\succsim_a$ can be represented by a continuous and Gateaux differentiable function $V_a : D \to \mathbb{R}$ where $V_a(m)$ is continuous in a.

While axiom 8 is not expressed using primitive assumptions on conditional preferences, it avoids problems with technical nature. Note that in the above theorem, axiom 4 is needed for technical reasons rather than imposing some qualitative behavioral restrictions on preferences as it ensures generic smoothness of the utility representation. Indeed, if PERU is replaced by preference for *late* resolution of uncertainty, the only preferences that, in addition, satisfies axioms 1-3, 5, and 7, are discounted expected utility preferences. Thus, if one is willing to assume that preferences are smooth in the sense of satisfying axiom 8, then axiom 5 is implied by axioms 3, 4, and 6.

It is straightforward to check the necessary conditions for $\succsim$ having an AD representation except for strong continuity, which is partly (the Lipschitz continuity part) dealt with in Dekel et al. (2007). Therefore, this section will focus on sketching the sufficiency proofs of Theorem 1.

By standard arguments, weak order, strong continuity, separability, are sufficient to obtain an additive utility function, which implies that each $\succsim_a$ can be represented by a Lipschitz continuous function

$$V_a(m) = v_a^1(m^1) + v_a^2(m^{21}) + v_a^3(m^{22}), \tag{3.8}$$

where m^1 is as before, m^{21} denotes the marginal distribution over $\Delta(C)$ and m^{22} denotes the marginal distribution over $\Delta(D)$ (remember, D can be identified with $\Delta(C \times \Delta(C \times D))$). If instead smooth utility replaces weak order and strong continuity, Lipschitz continuity is replaced by Gateaux differentiability. The novelty of Theorem 1 is to establish that

$$v_a^1(m^1) = \int \phi(u(c) - r(a^1))dm^1(c), \tag{3.9}$$

$$v_a^2(m^{21}) = \int \left[r(m^1) + \beta \int \phi(u(c) - r(m^1))dm^1(c)\right] dm^{21}(m^1), \tag{3.10}$$

$$v_a^3(m^{22}) = \beta \int \int \left[r(m^1) + \beta V_m(m)\right] d\mu(m)dm^{22}(\mu), \tag{3.11}$$

where ϕ is $\beta-$compatible with u, r is a reference point map with the above-described properties. I will now outline the main steps of the proof establishing these three equations.

To this end, first note that independence and PERU implies that

$$V_a(m) = \int [\hat{v}_a^1(c) + \hat{v}_a^2(\hat{m}^1) + v_a^3(\hat{m}^2)]dm(c, \hat{m})$$

where $\hat{m}_2$ is the marginal distribution over $\Delta(D)$ given $\hat{m}$ and $h_a : D \to \mathbb{R}$ for $h_a \equiv \hat{v}_a^2 + v_a^3$ is a convex function. A result by Ergin and Sarver (2010a) (generalized by Ergin and Sarver (2010b, Theorem 2.4)) implies that if $h_a : D \to \mathbb{R}$ is Lipschitz continuous, then it is Gâteaux differentiable on a dense G_δ set of D.

I show in the proof to the upcoming Theorem 4 that, in addition to satisfying axiom 8 each $\succsim_a$ also satisfies dynamic consistency, then $h_a(m)$ is convex without the need of axiom 5. This is noteworthy as it suggests that the notion of optimal reference points stems from the

insistence of dynamic consistent preferences.[18] Moreover, although anticipation compensation is used together with anticipations continuity in Theorem 1 to show that $h_a = \beta h$ for all $a \in D$ with $\beta \in (0,1)$, I show in the proof of Theorem 4 that both these assumptions are not needed if axiom 8 is imposed.

Since h is convex, it can be represented as the supremum over a set of affine functions (see, e.g., Aliprantis and Border (2006) Theorem 7.6).[19] Every such function can be interpreted as a von Neumann-Morgenstern utility function. The fact that h is Lipschitz continuous implies that a theorem by Gale (1967) can be invoked to show that the supremum is always attained. The same is true if h is smooth.

Dynamic consistency and degenerate anticipations continuity then implies that the linear functional corresponding to the Gâteaux derivative at some point $a = (c, m)$ must be an affine transformation of the function V_a that represents $\succsim_a$.

Using axiom 8, it is possible to provide a representation theorem that does not impose any structure on the anticipation-dependent part of each preference.

Theorem 4. *A family of preference relations $\succsim$ satisfies Axioms 3, 4, 6 and 8 if and only if for any $a \in D$, there is a $\beta \in (0,1)$, and a continuous and smooth function $v_{a^1} : C \to \mathbb{R}$ with*

$$\{\hat{m}^1\} = \arg \max_{m^1 \in \Delta(C)} \left\{ \int v_{m^1}(c) d\hat{m}^1(c) \right\} \tag{3.12}$$

such that $\succsim_a$ can be represented by a continuous and smooth function $V_a : D \to \mathbb{R}$ defined by

$$V_a(m) = \int \left[v_{a^1}(c) + \beta V_{\hat{m}}(\hat{m}) \right] dm(c, \hat{m}) \quad \forall m \in D. \tag{3.13}$$

Proof: See Appendix 3.9.2.

Theorem 4 provides a characterization of smooth preferences that does not assume any particular structure on how preferences for the timing of resolution of uncertainty is evaluated.

[18] See Segal (1997) for a discussion of how conditional and dynamic consistent preferences can be interpreted as reference dependent. Moreover, Border and Segal (1994) show, in a slightly richer setting, that independence is implied by the just-mentioned axioms if preferences additionally satisfy a reduction of compound lotteries assumption.

[19] For other papers closely related both in terms of technique and conceptually, see Kreps and Porteus (1979), Machina (1984), Maccheroni (2002), Chatterjee and Krishna (2011), Ergin and Sarver (2015), and Sarver (2018).

It highlights the severe restrictions imposed by dynamic consistency when coupled with separability. However, as shown by the reference-dependent preferences studied in this paper, this class is rich enough to capture many of the behaviors discussed in the decision theoretic literature and offers a viable alternative to the recursive models that predominates in applications.

The last part discusses the role of the anticipation axiom in imposing structure on each v_{a^1}. Consider the ordering $\succsim^I$ defined above. It is possible to show that it together with the other axioms imply that

$$c\frac{1}{2}c' \succsim^I \hat{c}\frac{1}{2}\bar{c} \iff \frac{1}{2}u(c) + \frac{1}{2}u(c') \geq \frac{1}{2}u(\hat{c}) + \frac{1}{2}u(\bar{c}),$$

where $u(c) \equiv v_{\delta_c}(c)$. By uniqueness of additively separable representations, this implies that it is without loss of generality to write $v_{a^1} = \phi(u - r(a^1))$ for all $a \in D$ with $\phi : [-a, a] \to \mathbb{R}$ being continuous, $a = \max u - \min u$, and normalized without loss of generality such that $\phi(0) = 0$. Dynamic consistency then implies that ϕ must be β-compatible with u as the reference point is optimal given the anticipated distribution over consumption.

Lastly, the uniqueness part of Theorem 1 follows from the uniqueness result of the mixture space theorem.

3.4 Status quo bias

In this section, I show that AD preferences are status quo biased (Samuelson and Zeckhauser, 1988). Thus, my preferences can account for an endowment effect that also holds when the object in question is a risky prospect.[20]

Definition 20. A family of preferences $\succsim$ exhibits (strict) status quo bias if $m \succsim_{\hat{m}} \hat{m}$ implies $m \succsim_m (\succ_m)\hat{m}$ for all $m, \hat{m} \in D$ (whenever $\succsim_m \neq \succsim_{\hat{m}}$).

A decision-maker whose preferences exhibit (strict) status quo bias is (strictly) less willing

[20]The endowment effect was coined by Thaler (1980) and refers to people's tendency to ascribe additional value to an object simply because they own it. An endowment effect for risk has frequently been observed in experimental and field studies (see, e.g., Post et al. (2008), Isoni et al. (2011), and Sprenger (2015))

to accept a temporal lottery m over $\hat{m}$ when her reference point was formed anticipating $\hat{m}$ than she is to accept m over $\hat{m}$ when her reference point was formed anticipating m (if the associated references points are different).[21]

Proposition 12. *If $\succsim$ has an AD representation, then it exhibits status quo bias.*

Proof: First, remember that dynamic consistency postulates that $(c, m) \succsim_a (c, \hat{m})$ implies $m \succsim_m \hat{m}$. Since $r(m)$ is optimal it must be the case that

$$r(m^1) + \int \beta(\phi(u(c) - r(m^1)) + \beta V_{\bar{m}}(\bar{m}))dm(c, \bar{m}) \geq$$
$$r(\hat{m}^1) + \int \beta(\phi(u(c) - r(\hat{m}^1)) + \beta V_{\bar{m}}(\bar{m}))dm(c, \bar{m})$$

which implies that $r(m^1) \geq r(\hat{m}^1)$ if $m \succsim_{\hat{m}} \hat{m}$. This, in turn, implies that $(c, m) \succsim_a (c, \hat{m})$ and we get the desired result.

Q.E.D.

It is interesting to note that strict status quo bias is not only consistent with dynamic consistency but also follows as a direct consequence of a strict version of it together with completeness (for any $a, m, \hat{m} \in D$, either $m \succsim_a \hat{m}$ or $\hat{m} \succsim_a m$) without any additional axioms.

Axiom 9. (Strict Dynamic Consistency) For any $a \in D$, $(c, m) \succsim_a (c, \hat{m})$ implies $m \succsim_m \hat{m}$ with strict preference if $\succsim_m \neq \succsim_{\hat{m}}$.

The proposition below makes the above claim precise and highlights the additional restrictions on a family of preferences with an AD representation satisfying the strict dynamic consistency axiom.

Proposition 13.

1. *If $\succsim$ satisfies completeness and strict dynamic consistency, then it exhibits strict status quo bias.*

[21] Note that if $\succsim$ exhibits strict status quo bias, then $m \succsim_{\hat{m}} \hat{m}$ for $\succsim_m = \succsim_{\hat{m}}$ trivially implies $m \succsim_m \hat{m}$.

2. *Let* $\succsim$ *have an AD representation* (u, ϕ, r, β) *with* $\beta > \frac{1}{2}$. *Then it satisfies strict dynamic consistency if and only if*

$$\{r(m^1)\} = \arg\max_{r\in\mathbb{R}}\{r + \int \beta\phi(u(c) - r)dm^1(c)\} \quad \forall\, m^1 \in \Delta(C). \tag{3.14}$$

Proof: See Appendix 3.9.3.

Thus, strict dynamic consistency implies that there exists a unique optimal reference point for any temporal lottery. This has the straightforward implication that $\beta\phi(y) < y$ for all $y \neq 0$.

I now present a comparative measure of status quo bias. The definition below only addresses status quo bias for deterministic outcomes. Essentially, by observing how large the decision-maker's willingness to accept is conditional on her anticipating to keep the object, it is possible to determine how susceptible to status quo bias she is.

Definition 21. The family of preferences $\succsim^1$ *exhibits a stronger status quo bias* than $\succsim^2$ if, for any $\boldsymbol{c}, \hat{\boldsymbol{c}} \in C^{\mathbb{N}}$,

$$\boldsymbol{c} \succsim^2_{\boldsymbol{c}} \hat{\boldsymbol{c}} \Longrightarrow \boldsymbol{c} \succsim^1_{\boldsymbol{c}} \hat{\boldsymbol{c}}.$$

The following theorem characterizes this comparative measure of status quo bias in terms of the AD representation.

Theorem 5. *Let* $\succsim^1$ *and* $\succsim^2$ *have AD representations* $(u_1, \phi_1, r_1, \beta_1)$ *and* $(u_2, \phi_2, r_2, \beta_2)$, *respectively. Then,* $\succsim^1$ *exhibits a stronger status quo bias than* $\succsim^2$ *if and only if* $\beta_1 = \beta_2$, *there exist scalars* $\sigma \in \mathbb{R}$, $\alpha > 0$ *such that* $u_2 = \alpha u_1 + \sigma$ *and* $r_2 = \alpha r_1 + \sigma$, *and* $\phi_2(\alpha x) \geq \alpha\phi_1(x)$ *for all* $x \in [-a_1, a_1]$ *satisfying* $(1 - \beta_1)\phi_1(x) \geq -a_1$ *with* $a_1 = \max u_1 - \min u_1$.

Proof: See Appendix 3.9.3.

From Theorem 3, two families of preference relations with AD representations are comparable in this way if and only if the preferences for deterministic consumption streams are the same, meaning that their discount factors are the same and their intrinsic utility functions are cardinally equivalent. Since the optimal reference point equates the outcome when there

is no uncertainty, what matters is the shape of ϕ. It follows that, if $u_1 = u_2 + \sigma$ for $\sigma \in \mathbb{R}$, for $\succsim^1$ to exhibit a stronger status quo bias than $\succsim^2$, ϕ_1 must be pointwise dominated by ϕ_2.

It is interesting to note that the comparative measure of status quo bias *coincides* with that of risk aversion. Thus, in this class of AD preferences, status quo bias and risk aversion are inherently linked with each other.

Corollary 4. *Let $\succsim^1$ and $\succsim^2$ have AD representations $(u_1, \phi_1, r_1, \beta_1)$ and $(u_2, \phi_2, r_2, \beta_2)$, respectively. If $\beta_1 > 1 + a_1/\phi_1(-a_1)$ for $a_1 = \max u_1 - \min u_1$, then $\succsim^1$ exhibits a stronger status quo bias than $\succsim^2$ if and only if $\succsim^1$ is more risk averse than $\succsim^2$.*

Any comparative measure of status quo bias for stochastic outcomes is bound to fail in the following sense. For any reasonable such measure, it must be the case that the compared families of preference relations are represented by the same reference point map. However, it can be shown that this implies that both families include the same preference relations. This observation is interesting in and of itself because it implies that reference point formation is inherently connected to consumption preferences.

3.5 Choice behavior and interpretation

While Theorem 1 provides a revealed preference foundation for AD preferences, I will now provide an interpretation of the class of utility functions that characterize these preferences. To save on notation, in this section I will only consider a two-stage model that implicitly assumes that consumption in any period other than period 2 is held constant. The first and the second stages represent *ex ante* and *ex post* the formation of the reference point, respectively. The decision-maker's preferences are defined over stochastic outcomes in the ex post stage.[22]

I will confine the analysis to the study of preferences over monetary risk with $C = [0, 1]$. In the simplest case, all uncertainty over ex post wealth is resolved at the end of the ex post stage. I consider AD preferences (u, ϕ) where both u and ϕ are strictly increasing and

[22] In this setting, the model is closely related to the model developed by Gollier and Muermann (2010) and is a special case of the model developed by Sarver (2018). See also Ben-Tal and Teboulle (1986, 2007) for another closely related idea.

β is normalized to unity for convenience. The decision-maker's preferences can then be represented by

$$V(m^1) = \max_{r \in [\min u, \max u]} \left\{ r + \int \phi(u(c) - r) dm^1(c) \right\}.^{23} \tag{3.15}$$

When the ex post outcome is deterministic, as mentioned earlier, the compatibility of ϕ with u implies that the reference point will be formed such that it equals this outcome. Thus, the decision-maker already derives all the utility from the outcome in the ex ante stage whereas she is neither elated nor disappointed by the outcome ex post. This feature is intuitive. Consider, for example, a person who is told that she has earned a bonus that will be added to her next wage. In such a situation, it seems plausible that the felicity she obtains from the bonus is received when she first received the news about the bonus, not when she actually has the money in her bank account.

When there is uncertainty, the situation is quite different. In this case, there is always scope for being either elated or disappointed by the outcome. Thus, there is a trade-off between anticipatory utility and the risk of being disappointed. If there is an outcome that is more likely than the rest, the reference point will tend to converge to it, and any outcome above or below it will generate elation or disappointment (see Section 3.6 for an analysis of how the reference point is formed). The intuition from when there is no uncertainty carries over to the case under uncertainty. It seems plausible that a person who is facing a risky situation that is likely to give her a bad outcome would feel unhappy about it beforehand but should then also be positively surprised by obtaining a larger than anticipated outcome ex post.

Note that when ϕ is not piecewise linear, it might be the case that the optimal reference point is not on the support of the lottery that the decision-maker is facing. Thus, the formation of the reference point may involve some type of cognitive dissonance. However, this feature seems quite realistic. Consider a PhD student who is waiting for the response from a journal about a paper he has submitted. In such a situation, it is very likely that the student would be elated to hear that the journal has not rejected the paper but has requested

[23] Consider preferences over the set $\{(c, m^1 \times m^2) : m^1 \in \Delta(C)\}$ given fixed c and m^2 to obtain the desired representation.

a revision. Moreover, it is quite likely that the student would be disappointed to hear that his paper was rejected.

One could imagine a model in which the reference point was formed as in the above equation but where the decision-maker does not obtain anticipatory utility. The problem is that anticipatory utility is a crucial ingredient for keeping the decision-maker's preferences dynamically consistent. The main issue is that a decision-maker who does not benefit from having a high reference point, as given endogenously by the environment she is facing, has incentives to distort her behavior to 'manipulate' her reference point. For example, in such a model the decision-maker would have incentives to hold pessimistic beliefs about the future to form a low reference point.

Whether the decision-maker is consciously choosing the reference point herself, I leave as an open question. It seems intuitive that the decision-maker is somewhat aware and can partially control the reference point. For example, there exists experimental evidence that suggest that people find optimism, i.e. anticipatory utility, costly and that they seem to be aware of this (see, e.g., Van Dijk et al. (2003), Carroll et al. (2006), Sweeny and Shepperd (2010), Sweeny and Krizan (2013)).

Example 12. Consider a worker who has learned that she will receive a bonus with her next wage, where the amount depends on how her firm has performing over the last quarter. For simplicity, let ϕ be piecewise linear with slope $1+\kappa$ for losses and $1-\kappa$ for gains, and u is everywhere differentiable. With probability α, her bonus is c, and with probability $(1-\alpha)$, it is $\bar{c}$, where $u(c) > u(\bar{c})$. That is, she is facing the lottery $m_\alpha^1 = \alpha\delta_c + (1-\alpha)\delta_{\bar{c}}$ with α being the probability that she obtains a high bonus, i.e., the quarterly report announces that her firm has enjoyed excellent performance.

The worker's utility associated with this lottery is given by

$$V(m_\alpha^1) = \max_{r \in [\min u, \max u]} \left[r + \alpha\phi(u(c) - r) + (1-\alpha)\phi(u(\bar{c}) - r) \right].$$

Thus, r solves

$$\alpha\phi'_-(u(c)-r) + (1-\alpha)\phi'_-(u(\bar{c})-r) \geq 1 \geq \alpha u'(\hat{c})\phi'_+(u(c)-r) + (1-\alpha)\phi'_+(u(\bar{c})-r). \quad (3.16)$$

The left-hand side of equation (3.16) is the marginal disutility from the 'risk of being disappointed' by choosing a higher reference point. Between the inequalities is the marginal anticipatory utility from increasing the reference point, $u'(\hat{c})$. Finally, the right-hand side represents the marginal utility from choosing a lower reference point.

If $\alpha \geq \underline{\alpha}$ where $\underline{\alpha} \in (0,1)$ is the unique solution to

$$u(\bar{c}) + \underline{\alpha}(1-\kappa)(u(c) - u(\bar{c})) = u(c) - (1-\underline{\alpha})(1+\kappa)(u(\bar{c}) - u(c)),$$

then the optimal reference point is $u(c)$, and $u(\hat{c})$ otherwise. Both are optimal in the knife-edge case when $\alpha = \underline{\alpha}$. Thus, the reference point map is sometimes not uniquely specified. Note that $\underline{\alpha}$ depends on κ. Thus, the worker's reference point is formed differently depending on how disappointed (elated) she is from consuming below (above) her reference point. If she is quite certain that the firm did well in the previous quarter, she is likely to be disappointed by a small bonus. Moreover, if she is very loss averse, i.e. $(1+\kappa)/(1-\kappa)$ is large, she will tend to form lower reference points and be more likely to be positively surprised.

The discontinuous nature of the reference point highlights that AD preferences can account for the observation that small changes in experimental procedures sometimes have significant effects on the subjects' perceptions of gains and losses (ODonoghue and Sprenger, 2018). Moreover, as is seen above, this phenomenon does not imply counterintuitive welfare implications: although the reference point changes discontinuously around $\bar{\alpha}$, the utility from the lottery, $V(m^1_\alpha)$, is everywhere continuous in α. This implies that such small procedural changes cannot significantly affect welfare.

□

3.5.1 Endogenous expectations induced by choice

I now consider a slightly more general environment in which the decision-maker's preferences are defined on distributions over choice sets. That is, the situation in which the decision-maker has to make a choice is uncertain when the reference point is formed. The timing is such that when the situation in which the choice has to be made is realized, the reference point is already formed and is taken as given by the decision-maker. As is standard, the decision-maker has

rational expectations about her choice in any realized situation given her reference point.

Let P represent the decision-maker's probabilistic beliefs regarding the ex post choice set. The set of potential choice sets is given by $\{x_l\}_{l\in L}$, where $L \subseteq \mathbb{R}$ and $x_l \subset \Delta(C)$ for all $l \in L$. That is, the ex ante probability that the ex post choice set is x_l is given by $P(l)$.[24] In this setting, a decision-maker with AD preferences (u, ϕ) over choice sets as described by P can be represented by

$$W(P) = \max_{r\in[\min u, \max u]} \left\{ r + \int \max_{m^1\in x_l} U_r(m^1) dP(l) \right\} \tag{3.17}$$

where

$$U_r(m^1) = \int \phi(u(c) - r) dm^1(c). \tag{3.18}$$

Given a reference point r, ex post preferences over lotteries in a realized choice set, x_l, can without loss be represented by U_r. For any reference point r and choice set l, let $m_l^r \in \arg\max_{m^1\in x_l} U_r(m^1)$. Thus, every reference point, r, induces a distribution $\int m_l^r dP(l)$ over ex post consumption. Thus, $W(P) = V(\int m_l^r dP(l))$, where V is given in equation (3.15).

The interpretation of equation (3.17) is as follows: The decision-maker ponders how she would feel making a choice among lotteries in each of every potential choice set knowing the overall distribution over consumption these choices induce. Again, the formation of the reference point need not be a conscious process; the important part is that the decision-maker can correctly predict her preferences in the ex post stage. Since W is strictly increasing in ex post utility, U_r, her preferences are dynamically consistent.

I think of P as representing a dynamic two-stage process involving an asymmetry regarding the time the decision-maker has before making a choice. The interpretation is that the decision-maker has enough time before the choice situation occurs that she is able to mentally adjust to the risk she is facing. By contrast, when she chooses a lottery in the realized choice set, she is given a limited amount of time, so that she takes her mental state as given. Figure 3.3 shows the timeline of this interpretation.

I illustrate the just-described setting by way of example. The example below shows that

[24] As AD preferences are dynamically consistent, it is without loss of generality to only consider temporal lotteries. If this axiom were to be relaxed, one would need to consider dynamic choice sets in the axiomatization. See Gul and Pesendorfer (2004) for such a framework.

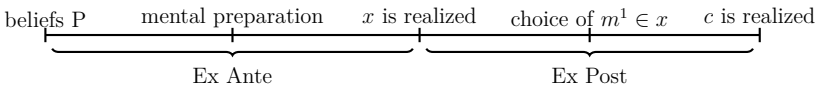

Figure 3.3: The timeline of the decision-set model.

a decision-maker with AD preferences can *seemingly* violate transitivity when choices of lotteries in the ex post stage are the only thing that is observed.

Example 13. Consider a person who has an appointment for a health check-up that can reveal whether the person is in the risk zone of contracting a certain disease. Immediately after the checkup, she can choose between signing up for insurance that will be pay for the medication is she were to develop the disease. Thus, if she chooses to take up the insurance, she will be as equally well off regardless of whether she develops the disease.

This situation is represented by a belief over ex post choice sets denoted by P. The degenerate choice set in which the test reveals that she is not in the risk zone of having the disease is represented by the (monetary) outcome c' and occurs with probability $1-\gamma$. With probability γ, the test reveals that she is in the risk zone. In this case, she can either opt to insure against the disease, represented by a deterministic outcome $c < c'$ (she will be completely compensated for having the disease if insured), or she can take the risk of not insuring herself, represented by a lottery m^1. If choosing not to insure herself, with probability α she develops the disease, represented by the outcome $\hat{c} < c$, and avoids it otherwise. The decision tree to the left in Figure 3.4 illustrates the situation; squares denote decision nodes, and circles denote chance nodes.

Now consider the situation in which she already knows that she is in the risk zone of contracting the disease and again faces the decision of whether to sign up for the supplementary health insurance plan. In this situation, she faces the degenerate belief, Q, associated with the ex post choice set given by the choice between c and m^1. The decision tree to the right in Figure 3.4 illustrates the situation.

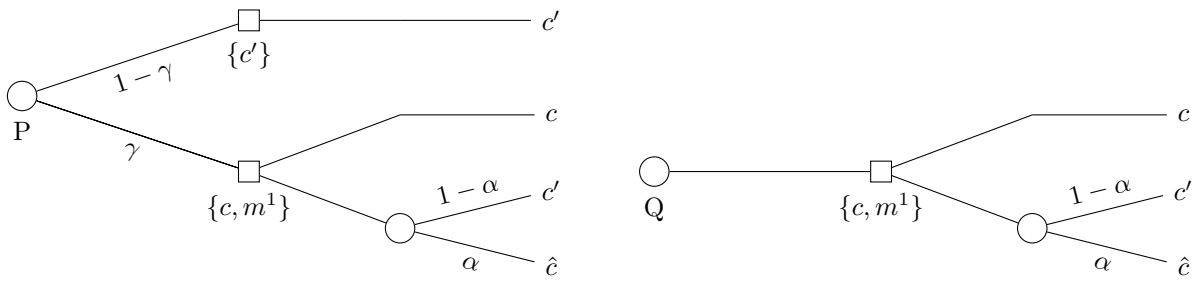

Figure 3.4: Decision trees for the ex ante beliefs P and Q.

The person's preferences can be represented by AD preferences where ϕ is piecewise linear as before and u is linear. It can be shown that for any $\alpha \in (\underline{\alpha}, 1)$ with

$$\underline{\alpha} = \max\left(1 - \frac{c - \hat{c}}{(1-\kappa)(c' - \hat{c})}, \frac{c' - c}{(1+\kappa)(c' - \hat{c})}\right),$$

she strictly prefers c over μ given ex post beliefs Q. Moreover, given any α, there exists a $\bar{\gamma}$ such that the optimal reference point when facing P is c' for any $\gamma \in (0, \bar{\gamma})$. She then strictly prefers μ over c if she happens to face the choice when her beliefs were P if and only if

$$\alpha < \frac{c - \hat{c}}{c' - \hat{c}}.$$

It is straightforward to show that for any $\hat{c} < c' < c$, setting κ sufficiently close to 1 yields $\frac{c - \hat{c}}{c' - \hat{c}} > \underline{\alpha}$. Thus, the model is able to rationalize the person's choice of μ over c conditional on her having just received the bad news about her health status, and her choice of c over μ conditional on her having the time to mentally prepare herself for her new health status.

This example shows that if the analyst fails to take the ex ante stage into account, she is led to infer that AD preferences may violate transitivity. To see this, note that if the person's reference point is c, she chooses c from the choice set $\{c, \mu\}$, while when it is c', she chooses μ from the same choice set. That is, as her reference point is optimal against μ, she maximizes utility by ex post choosing μ over c, even though she would have preferred c over μ had she anticipated to face c ex ante. This consideration is important because violations of transitivity lead to very different policy prescriptions than do preferences that may depend on past anticipations; a failure of transitivity might lead the policy-maker to focus on irrelevant features of the choice set itself, whereas AD preferences suggest that focus should be placed

on expectations management.

□

3.6 Reference point formation and loss aversion

I here consider prominent features of AD preferences from a behavioral economics perspective and discuss properties commonly associated with reference-dependent preferences. For simplicity, I focus only on the two-stage model with ex ante choices over lotteries with monetary outcomes. All the results in this section easily generalize to the temporal lotteries setting. The first part relates to an analysis of the objects in the AD representation with respect to risk attitudes in terms of stochastic dominance of different orders. The second part focuses on how the reference point is formed. Thereafter, I analyze AD preferences with respect to loss aversion and first-order risk aversion. Finally, I discuss three characterizations of special cases of AD preferences.

The result below follows from Cerreia-Vioglio et al.'s (2017) analysis of the expected utility core. Their results generalize the analysis in Machina (1982) because the functions do not need to be differentiable (see also related results in Chatterjee and Krishna (2011) for Lipschitz continuous preferences and Sarver's (2018) local expected utility analysis of upper semicontinuous utility functions that are convex in probabilities).

Proposition 14. *Let $\succsim$ have an AD representation (u, ϕ, r, β). Then,*

1. *V satisfies first-order stochastic dominance.*

2. *V satisfies second-order stochastic dominance if and only if $\phi(u - r)$ is concave on C for all $r \in [\min u, \max u]$.*

Proof: See Appendix 3.9.3.

An important question that has been postponed until now is what can be said about how the reference point is formed given AD preferences. The first result states that, given a lottery m^1, the optimal reference point is within the range of the intrinsic utility levels associated with a consumption level on its support. Moreover, if ϕ is piecewise linear, then the

consumption level c such that $u(c) = r(m^1)$ is always in m^1's support. Finally, the optimal reference point is generically uniquely specified.

Proposition 15. *Let $\succsim$ have an AD representation (u, ϕ, r, β). For any lottery $m^1 \in \Delta(C)$, $r(m^1)$ is such that $\underline{c} \leq r(m^1) \leq \bar{c}$, where $\underline{c} = \inf\{c \in supp(m^1)\}$ and $\bar{c} = \sup\{c \in supp(m^1)\}$, and it is generically unique. Moreover, if ϕ is piecewise linear, then $r(m^1) = u(c)$ for $c \in supp(m^1)$.*

Proof: The first part follows easily from ϕ being $\beta-$compatible with u. That $r(m^1)$ is unique on a dense set follows from Ergin and Sarver's (2010b) generalization of Mazur's Theorem on the generic Gâteaux differentiability of continuous and convex functions, which I utilize in the proof of Theorem 1 in Appendix 3.9.2. The last part follows from the observation that if $c \notin supp(m^1)$ for $r(m^1) = u(c)$, then increasing $r(m^1)$ by $\varepsilon > 0$ such that $\hat{c} \in supp(m^1)$ for $r(m^1) + \varepsilon = u(\hat{c})$ increases utility by $\eta\varepsilon > 0$ since it does not increase the risk of being disappointed.

Q.E.D.

Under the assumption that ϕ is twice continuously differentiable and strictly concave, the optimal reference point is always uniquely specified. In addition, if the distribution over outcomes is shifted towards another distribution, the optimal reference point also shifts towards the reference point of that distribution. Moreover, a lottery that first-order stochastically dominates another lottery has a higher optimal reference point. The same is true for second-order stochastic dominance if $\phi''' \geq 0$. Kreps and Porteus (1979) and Gollier and Muermann (2010) obtain similar results when analyzing preferences that are convex in probabilities.

Proposition 16. *Let $\succsim$ have an AD representation (u, ϕ, r, β) with $\phi'' < 0$. Then, the following is true:*

1. *The $r(m^1)$ is continuous in $m^1 \in \Delta(C)$.*
2. *For any $m^1, \hat{m}^1 \in \Delta(C)$ and $\alpha \in [0, 1]$, $r(\alpha m^1 + (1-\alpha)\hat{m}^1)$ converges monotonically to $r(m^1)$ as α tends to 1.*

3. *For any $m^1, \hat{m}^1 \in \Delta(C)$, if m^1 first-order stochastically dominates $\hat{m}^1$, then $r(m^1) \geq r(\hat{m}^1)$.*

4. *Assume in addition that $\phi''' \geq 0$. For any $m^1, \hat{m}^1 \in \Delta(C)$, if m^1 second-order stochastically dominates $\hat{m}^1$, then $r(m^1) \geq r(\hat{m}^1)$.*

Proof: For part 1, by Berge's Theorem of the Maximum and the strict concavity of the optimization program determining the reference point, the optimal reference point, $r(m^1)$, is unique and continuous in m^1. For part 2, the optimal reference point is given by the unique r that satisfies $\alpha\eta \int \phi'(u(c) - r)dm^1(c) + (1-\alpha)\eta \int \phi'(u(c) - r)d\hat{m}^1(c) = 1$. Totally differentiating this equation yields

$$\frac{\partial r}{\partial \alpha} = \frac{\int \phi'(u(c) - r)dm^1(c) - \int \phi'(u(c)r)d\hat{m}^1(c)}{\alpha \int \phi''(u(c) - r)dm^1(c) + (1-\alpha) \int \phi''(u(c)r)d\hat{m}^1(c)},$$

which is always either positive or negative. Parts 3 and 4 follow straightforwardly from the observation that $-\phi'(u(c) - r)$ satisfies FOSD (SOSD) iff $\phi'' \leq 0$ (and $\phi''' \geq 0$). Thus, as the left-hand side of $\beta \int \phi'(u(c) - r)dm^1(c) = 1$ is decreasing in r, any increase in the risk in an FOSD (SOSD) sense implies a lower reference point.

Q.E.D.

For any two AD representations $(u_1, \phi_1, r_1, \beta_1)$ and $(u_2, \phi_2, r_2, \beta_2)$ with $u_1 = u_2$, if ϕ_2 is twice continuously differentiable and concave, and ϕ_1 is a concave transformation of ϕ_2 with lower slope, then the optimal reference points are such that $r_1(m^1) \leq r_2(m^1)$ for all $m^1 \in \Delta(C)$.

Proposition 17. *Let $\succsim^1$ and $\succsim^2$ have AD representations $(u_1, \phi_1, r_1, \beta_1)$ and $(u_2, \phi_2, r_2, \beta_2)$, respectively, and $u_2 = \alpha u_1 + \sigma$ and $r_2 = \alpha r_1 + \sigma$ for $\alpha > 0$ $\sigma \in \mathbb{R}$. If $\phi_2'' < 0$ and $\beta_2 \rho(\phi_2(\alpha(\cdot))) = \beta_1 \alpha \phi_1$ with $\rho' \leq 1$ and $\rho'' \leq 0$, then $r_1(m^1) \leq r_2(m^1)$ for all $m^1 \in \Delta(C)$.*

Proof: Since $\phi_2'' < 0$, $\beta_2 \int \phi_2'(u_2(c) - r)dm^1(c)$ is strictly decreasing in $r \in \mathbb{R}$ and there is a unique intrinsic utility level, r_2, such that $\beta_2 \int \phi_2'(u_2(c) - r_2)dm^1(c) = 1$. Moreover,

$$\beta_1 \alpha \int \phi_1'(u_1(c) - r)dm^1(c) = \beta_2 \int \rho'(\phi_2(u_2(c) - \alpha r - \sigma))\phi_2'(u_2(c) - \alpha r - \sigma)dm^1(c).$$

Since $x < \beta_1\phi_1(x)$ for all $x \neq 0$, $\rho'(x) = 1$ for $x < 0$ and $\rho'(x) \leq 1$ for $x \geq 0$. Therefore, $\beta_2 \int \phi_2'(u_2(c) - r_2)dm^1(c) = 1$ implies $\beta_1 \int \phi_1'(u_1(c) - (r_2 - \sigma)/\alpha)dm^1(c) - 1 \leq 0$ where the left-hand side is strictly decreasing in r_2.

Q.E.D.

An important feature of reference-dependent preferences is that they allow for loss aversion. That is, the tendency for people to dislike losses relative to the reference point more than they like same-sized gains. Formally, I say that ϕ features *loss aversion* if $\phi(y) + \phi(-y) < 0$ for all $y \in (0, a]$.[25] Since ϕ is $\beta-$compatible, it has finite left and right derivatives around the origin. Therefore, *loss aversion for small stakes* is captured by the following property:

$$\lim_{y \to 0} \phi'(|y|) \equiv 1 \text{ and } \lim_{y \to 0} \phi'(-|y|) \equiv \lambda \text{ with } \lambda > 1.$$

Here, λ is referred to as the *coefficient of loss aversion*. The normalization of $\lim_{y \to 0} \phi'(|y|) \equiv 1$ is without loss of generality in the two-stage setting because there is no trade-off between consumption in different periods. Note that ϕ is β-compatible with some u then it must be the case that $\lambda \geq \beta^{-1} \geq 1$.

Proposition 18. *If (ϕ, u, r, β) is an AD representation, then ϕ features loss aversion.*

Proof: Assume to obtain a contradiction that $\beta\phi(y) \geq y$ for $y > 0$; then $\beta\phi(u(c) - r) - u(c) + r > 0$ for $u(c) \geq r$, implying that $\beta\phi(u(c) - r) + r \geq u(c)$ and contradicting ϕ being β-compatible with u. The argument for why $\beta\phi(-y) < -y$ for all $y > 0$ is symmetric. Thus, $\phi(y) + \phi(-y) < (y - y)/\beta = 0$ for all $y > 0$.

Q.E.D.

I will now show that loss aversion for small stakes completely captures the decision-maker's attitude towards small-stakes lotteries. The result relies on Segal and Spivak's (1990) analysis of first-order risk aversion. The idea is to measure the decision-maker's aversion to small-stakes lotteries using the notion of the so-called *risk premium*. Formally, given a utility function V, the certainty equivalent of the lottery $m^1 \in \Delta(C)$ is defined implicitly as the

[25] The notion of loss aversion considered here should not be understood as the component of risk attitudes popularized by Kahneman and Tversky (1979). Their notion requires that "the function is steeper in the negative than in the positive domain" (Tversky and Kahneman, 1991, p.1039).

outcome $c \in C$ such that $V(\delta_c) = V(m^1)$. The difference between the expected value and the certainty equivalent of a lottery m^1 is called the risk premium and is denoted by $\pi(m^1)$.

Definition 22. The family of preferences $\succsim$ exhibits first-order risk aversion at wealth level $c \in C$ if for all $m^1 \neq \delta_{\hat{c}}$ with $\int dm^1(c) = \hat{c}$, $\frac{\partial \pi(\delta_c + \varepsilon m^1)}{\partial \varepsilon}|_{\varepsilon=0^+} < \hat{c}$.[26]

The notion of first-order risk aversion has economic meaning, as $\frac{\partial \pi(\delta_c + \varepsilon m^1)}{\partial \varepsilon}|_{\varepsilon=0^+} < \hat{c}$ implies that the decision-maker is risk averse even if her exposure to the risk, as measured by $\varepsilon > 0$, approaches zero.

Proposition 19. *If $\succsim$ has an AD representation (u, ϕ, r, β), then it exhibits first-order risk-averse at all wealth levels if and only if ϕ features loss aversion for small stakes.*

Proof: See Appendix 3.9.3.

Note that since ϕ is $\beta-$compatible, a decision-maker with AD preferences is never first-order risk-loving. Moreover, if ϕ is differentiable at zero, the decision-maker is risk-neutral for small-stakes lotteries at almost all wealth levels, just like any expected utility maximizer with an increasing utility function.

I now present two special cases of the above model and relate it to the standard time- and state-separable expected utility model. As first observed by Kreps and Porteus (1979), an AD representation (u, ϕ, r, β) is neutral towards the resolution of uncertainty and can be represented by an additively separable expected utility function if and only if ϕ is a linear function with slope β^{-1}. To see this, note that $U_r = u(c) - r$ and

$$V(m^1) = \max_{r \in [\min u, \max u]} \int (r + u(c) - r)\, dm^1(c) = \int u(c) dm^1(c).$$

The first special case is when ϕ is piecewise linear; then, any AD preferences represented by (u, ϕ, r, β) is in the class of rank-dependent utility functions (Quiggin (1982), Schmeidler (1989)). Let F_{m^1} be the cumulative distribution function for the measure $m^1 \in \Delta(C)$. The following characterization was provided by Sarver (2018, Proposition S.1, Supplementary

[26] Here it is implicitly assumed that $\delta_c + \varepsilon m^1 \in \Delta(C)$ for all $\varepsilon > 0$ close to zero.

Material).

Proposition 20. *(Sarver, 2018) If $\succsim$ has an AD representation where ϕ is piecewise linear, then for any $m^1 \in \Delta(C)$,*

$$V(m^1) = \int u(c) d(g \circ F_{m^1})(c),$$

where

$$g(x) = \begin{cases} \lambda\beta x & \text{for} \quad x \leq \frac{1-\beta}{(\lambda-1)\beta}, \\ \beta x + 1 - \beta & \text{for} \quad x > \frac{1-\beta}{(\lambda-1)\beta}. \end{cases}$$

This result further highlights that expectations-based reference-dependent models with piecewise linear gain-loss utility are indistinguishable from rank-dependent models (see Masatlioglu and Raymond (2016) and Sarver (2018, Supplementary Material)).

Lastly, as shown by Ben-Tal and Teboulle (2007, Example 2.1) another special case is when

$$\phi(y) = (\theta\beta)^{-1}(1 - exp(-\theta y)).$$

Then, for any $m^1 \in \Delta(C)$,

$$V(m^1) = -\frac{\log\left(\int exp(-u(c)) dm^1(c)\right)}{\theta}.$$

3.7 Applications

3.7.1 Stochastic representative agent economy

In this subsection, I analyze a simple endowment economy with i.i.d. consumption growth. One goal is to elaborate the model's compatibility with the observed equity premium of 6% for the period 1889-1978. Mehra and Prescott (1985) have shown that this premium is incompatible with the additively separable expected-utility model with a reasonably low coefficient of relative risk aversion. In addition, I derive the magnitude of the model's implied *timing premium* associated with the data. The timing premium is defined as the share of

income a consumer would forgo to have all future uncertainty about consumption resolved tomorrow (Epstein et al., 2014).

Identical agents maximize preferences represented by an AD utility representation

$$V_a = \mathbb{E}\left[\beta\phi(u(c_1) - r(a^1)) + \sum_{t=1}^{\infty} \beta^t \max_{r_{t+1}\in\mathbb{R}} \left(r_{t+1} + \mathbb{E}\left[\beta\phi(u(c_{t+1}) - r_{t+1})|\mathcal{G}_t\right]\right)\right],$$

where $\mathcal{G}_t$ represents the information available to the agent at time t. Note that, due to the initial reference point, the implications of the model cannot fully be captured in a static setting.

I will assume that the representative agent's intrinsic utility function is of the CRRA form:[27]

$$u(c) = \begin{cases} \dfrac{c^{1-\rho}}{1-\rho} & \text{for} \quad \rho \geq 0, \\ \log(c) & \text{for} \quad \rho = 1. \end{cases} \tag{3.19}$$

I employ a one-parameter, $\kappa \in (0,1)$, version of the piecewise linear function

$$\phi(x) = \begin{cases} (1-\kappa)x/\beta & \text{for} \quad x \geq 0, \\ (1+\kappa)x/\beta & \text{for} \quad x < 0. \end{cases} \tag{3.20}$$

Log-consumption growth is given by the process

$$\log c_{t+1} - \log c_t = \mu_c + \sigma_c v_{t+1}, \; v_{t+1} \sim i.i.d.\, \mathcal{N}(0,1). \tag{3.21}$$

Denote the *stochastic discount factor*, or SDF, at time t by M_{t+1}. The risk-free rate is denoted by $R^f_{t+1} = \dfrac{1}{\mathbb{E}_t[M_{t+1}]}$, and the excess return on the risky asset is denoted by $R^m_{t+1} - R^f_{t+1}$. Using the basic Hansen-Jagannathan bound, any model that attempts to explain the equity premium puzzle must be able to generate an SDF where the Sharpe ratio, $\dfrac{\mathbb{E}_t[R^m_{t+1} - R^f_{t+1}]}{\sigma_t(R^m_{t+1})}$, is bounded by $\dfrac{\sigma_t(M_{t+1})}{\mathbb{E}_t[M_{t+1}]}$. Postwar data indicate that the Sharpe ratio is close to 0.5, which imposes a significant lower bound on the volatility of the SDF.

[27] Note that there are two ways in which this model deviates from the representation in Section 3.2.2. First, the consumption space in this setting is $\mathbb{R}_+$, which is not compact. Second, since u is of the CRRA form, it is not Lipschitz continuous on this domain. This is done for technical convenience and the model highlights behavioral insights that could be generated within the axiomatic setting.

By solving the agents' maximization problem, the Euler equation gives the price of a claim to the consumption stream, which is interpreted as stocks. Thus, there is an SDF at time t, conditional on the reference point r_t, that is given by

$$M_{t+1} = \beta \left(\frac{c_{t+1}}{c_t}\right)^{-\rho} \left(\frac{1-\kappa+2\kappa\mathbb{1}_{\{u(c_{t+1})\leq r_{t+1}\}}}{1-\kappa+2\kappa\mathbb{1}_{\{u(c_t)\leq r_t\}}}\right).^{28}$$

It follows by Proposition 20 that the optimal reference point r_{t+1} satisfies $Pr(u(c_{t+1}) \leq r_{t+1}) = \frac{1}{2}$. Moreover, the Hansen-Jagannathan bound is given by

$$\frac{\mathbb{E}_t[R^m_{t+1} - R^f_{t+1}]}{\sigma_t(R^m_{t+1})} \leq \kappa$$

which can rationalize a Sharpe ratio below 0.3 with an empirically plausible coefficient of loss aversion $(1+\kappa)/(1-\kappa) = 1.86$ independent of $\rho \geq 0$. A potential problem with the model is that the standard deviation of the risk-free rate can be shown to be such that

$$\sigma(R^f_{t+1}) \geq \frac{\mathbb{E}_t[R^m_{t+1} - R^f_{t+1}]}{\sigma_t(R^m_{t+1})} \exp\left(-\rho^2\sigma_c^2/2\right) R^f,$$

implying that the risk-free rate is too variable to match historical data. This property is shared with other reference-dependent models (see, e.g., Abel (1990) and Pagel (2016))

The tractability of the model makes it possible to provide the average equity premium in closed form. The results of this section thus far are summarized by a proposition.

Proposition 21. *The unconditional risk-free rate and the market return are given by*

$$\mathbb{E}[R^f] = \frac{1}{\beta}\exp\left[\rho\mu_c - \frac{\rho^2}{2}\sigma_c^2\right]$$

and

$$\mathbb{E}[R^m] = \left[\frac{1}{2(1-\kappa)} + \frac{1}{2(1+\kappa)} + \frac{1-v}{v}\right]\exp\left[-\mu_c + \frac{1}{2}\sigma_c^2\right],$$

where $v = \beta\exp\left[-(1-\rho)\mu_c + \frac{(1-\rho)^2}{2}\sigma_c^2\right]$.

[28] I show in Appendix 3.9.3 how ϕ can be arbitrarily approximated using a smooth function so that the SDF is uniquely defined everywhere.

Proof: See Appendix 3.9.3.

Let $\kappa = 0.3$, $\rho = 1$, $\beta = 0.99$, and $\mu_c = 0.018$ with $\sigma_c^2 = 0.00127$ as in Mehra and Prescott (1985). Using these parameter values, I obtain an equity premium of 6.1 percent with a risk-free rate of 2.8 percent.

It should be noted that there is a voluminous literature that analyzes the impact of loss aversion on asset prices (see, e.g., Benartzi and Thaler (1995), Barberis et al. (2001), Yogo (2008), Barberis and Huang (2009), Pagel (2016), and Andries (2019)). Moreover, it is well known that it is possible to obtain a high equity premium together with low effective risk aversion (see, e.g., Epstein and Zin (1990)). Instead, the aim of this application is to illustrate the properties of the AD utility representation.

However, one interesting feature of my model is its implication for the pricing of long-term risk. I will now consider the timing premium associated with the above process. Epstein et al. (2014) consider a more complicated consumption process based on the process in Bansal and Yaron (2004). However, as I will make clear below, the results in this section also obtain with a consumption process as in the long-run risk literature with stochastic volatility (Bansal and Yaron, 2004).

Define the *timing premium* as the share of utility a person would be willing to forgo to have all risk resolved in the next period, given by $\pi^* = 1 - \frac{U_0}{U_0^*}$, where U_0 is the lifetime utility from the consumption process above, and U_0^* is the utility from the alternative process in which all risk is resolved at time 1.

It can be shown that U_0 is given by

$$U_0 = (1 - \kappa + 2\kappa \mathbb{1}_{\{u(c_t) \leq r_0\}}) \log(c_0) + \sum_{t=1}^{\infty} \beta^t \left(\log(c_0) - \kappa \sigma_c \sqrt{\frac{2}{\pi}} + t \cdot \mu_c \right)$$

and U_0^* is given by

$$U_0^* = (1 - \kappa + 2\kappa \mathbb{1}_{\{u(c_0) \leq r_0\}}) \log(c_0) + \beta \left(\log(c_0) + \mu_c - \kappa \sigma_c \sqrt{\frac{2}{\pi}} \right) + \sum_{t=2}^{\infty} \beta^t \left(\log(c_0) + t \cdot \mu_c \right).$$

It is evident that, although the per-period 'extra' risk premium relative to the standard model, $\kappa \sigma_c \sqrt{2/\pi}$, is significant, it is dwarfed by the benefit from the expected consumption increase

over time. Thus, the timing premium is negligible. It can be shown that $\sum_{t=1}^{\infty} \beta^t t \cdot \mu_c = \mu_c(1-\beta)^{-2} = 180$, whereas $(1-\beta)^{-1}\kappa\sigma_c\sqrt{2/\pi} \approx 1.21$ (assuming $c_0 > 1$). This implies that

$$\pi^* = 1 - \frac{U_0}{U_0^f} < \frac{180 - 1.21}{180} \approx 0.67\%.$$

By contrast, Epstein et al. (2014) show that the timing premium is between 20 and 30 percent in the Epstein-Zin model using Bansal and Yaron's consumption process. Since AD preferences are time-separable and only averse to uncertainty resolving over time just before the associated consumption takes place, the introduction of stochastic volatility does not change this result if the expected volatility is the same.

There are few experimental or field studies investigating the cost of the temporal resolution of uncertainty. An exception is the experimental investigation by Meissner and Pfeiffer (2018), who elicit preferences in a model-free way. They find that their subjects are willing to forgo on average 4.52 percent of their total consumption to have all uncertainty resolve immediately.

Remark 1. *As observed by Dillenberger et al. (2019), most models that has proposed in the asset pricing literature that can explain the equity premium puzzle while maintaining low effective risk aversion violates a property the call stochastic impatience. An example of this property can be described by the following two temporal lotteries.*

A With equal probability, permanently increase consumption by either 10 % starting today or by 20 % starting next year.

B With equal probability, permanently increase consumption by either 20 % starting today or by 10 % starting next year.

Here, it seems intuitively that any rational decision-maker would prefer lottery B over A. By contrast, Dillenberger et al. (2019) show that, e.g., the specification of Epstein-Zin preferences utilized by Bansal and Yaron (2004) is such that A is preferred over B, violating stochastic impatient. It is easy to show that any increasing AD preferences satisfy stochastic impatient as defined by Dillenberger et al. (2019), including a strict preference for B over A.

3.7.2 Life-cycle consumption

In this subsection, I analyze a simple life-cycle model with an infinite horizon. The agent is endowed with AD preferences. At the beginning of each period, she observes a permanent income shock and then decides how much to consume and how much to save. I will focus on two well-documented features of consumption behavior in response to shocks, namely excess sensitivity and excess smoothness. This phenomenon relates to predictions made by the standard additively separable model, which states that consumption growth between two periods $t-1$ and t cannot be explained by variables from period $t-1$ and earlier (see Jappelli and Pistaferri (2010) for a recent review).

A period is indexed $t \in \mathbb{N}_+$. As in the previous subsection, ϕ is piecewise linear and given by equation (3.20), but now intrinsic utility is given by an exponential utility function. This makes it possible to obtain closed-form solutions (see, e.g., Caballero (1990)). Thus,

$$u(c) = -\frac{1}{\theta} e^{-\theta c}$$

where the coefficient of risk aversion is $\theta \in (0, \infty)$. The additive income process is given by $Y_t = P_{t-1} + S_t$ where $S_t \sim \mathcal{N}(0, \sigma^2)$ with realization s_t and the permanent income is given by $P_t = P_{t-1} + S_t$. The intertemporal budget constraint is given by

$$c_{t+1} = y_{t+1} + (1+r)A_t - A_{t+1},$$

where c_t is consumption and A_t is nonhuman wealth in period t, respectively, and $1+r$ denotes the risk-free interest with $R = (1+r)^{-1}$ being the objective discount factor. In this model, agents maximize intertemporal utility given the constraint.

The first objective is to obtain a closed-form solution of the consumption function. In deriving this function, I follow the approach of Caballero (1990). The proposition below summarizes the feature of the consumption function associated with the just-described AD preferences.

Proposition 22. *The consumption function is given by*

$$c_t = y_t^p - (1-R)\sum_{i=1}^{\infty} R^i \sum_{j=1}^{i} \mathbb{E}_t[\Gamma(s_{t+j-1})] \tag{3.22}$$

where $y_t^p := (1-R)\left(A_t + \sum_{i=0}^{\infty} R^i \mathbb{E}_t[y_{t+i}]\right)$ *is permanent income and*

$$\Gamma_t(s_t) = \begin{cases} K - \dfrac{\log(1+\kappa)}{\theta} & \text{for} \quad s_t < \underline{S}, \\ K + (1+r)s_t & \text{for} \quad s_t \in [\underline{S}, \bar{S}], \\ K - \dfrac{\log(1-\kappa)}{\theta} & \text{for} \quad s_t > \bar{S}. \end{cases} \tag{3.23}$$

where $\underline{S} = -\dfrac{\log(1+\kappa)}{(1+r)\theta}$, $\bar{S} = -\dfrac{\log(1-\kappa)}{(1+r)\theta}$, $K = \mathbb{E}\left[(1-\kappa+2\kappa\mathbb{1}_{\{c_{t+1}<\hat{c}_{t+1}\}})e^{-\theta v_{t+1}} + \rho\right]/\theta$ *is a precautionary savings term,* $\rho = \log(1+r) + \log(\beta)$*, and* v_t *is an innovation with zero mean and an atom at zero.*

Proof: See Appendix 3.9.3.

The consumption function in Proposition (22) shares many similarities with the consumption function obtained using the standard model (i.e., by setting $\kappa = 0$), as derived by Caballero (1990). The permanent income term y_t^p is identical to the one in the standard model. What is different now is that (i) Γ_t is biased towards the 'status quo' and (ii) the precautionary savings term, K, is typically larger than the standard one and is a function of a 'biased' innovation term v_{t+1}. In the standard model, the innovation v_{t+1} is identically distributed to S_t whereas AD preferences' status quo bias implies that v_{t+1} has an atom at zero. Importantly, status quo bias implies that c_t is independent of $s_t \in [\underline{S}, \bar{S}]$.

Three other salient properties of the consumption function are summarized in the following corollary:

Corollary 5.

1. *The stochastic process of consumption is not a martingale with drift.*
2. *The precautionary savings term is increasing in the variance,* σ^2*, of* S_t*.*
3. *The precautionary savings term is positive, but depending on* s_t *and* $\hat{c}_t$*, it is*

possible that $\mathbb{E}_t[c_t - c_{t+1}] < 0$.

I now turn to the question regarding excess sensitivity and excess smoothness of consumption. These two observations are intrinsically related as explained by Campbell and Deaton (1989). I formalize these notions in the following way.

Definition 23. Consumption is excessively smooth on the interval $I \subset \mathbb{R}$ if $\dfrac{\partial c_t}{\partial s_t} < 1$ for all $s_t \in I$, and is excessively sensitive on I if $\dfrac{\partial (c_{t+1} - c_t)}{\partial s_t} > 0$ for all $s_t \in I$.

Given the above consumption function, the following proposition follows as a direct consequence.

Proposition 23. *The AD decision-maker's consumption is excessively smooth and sensitive only on* $[\underline{S}, \bar{S}]$.

An important feature of the above proposition is that excess sensitivity and smoothness disappear for permanent income shocks outside the interval $[\underline{S}, \bar{S}]$. This is supported by empirical evidence summarized by Jappelli and Pistaferri (2010) where it is noted that excess sensitivity and smoothness seem to vanish for *large* permanent income shocks. They call this pattern the *magnitude hypothesis* (see also the discussion in Chetty and Szeidl (2016)). Although the rather extreme predictions in Proposition (23) stem from the particular choice of intrinsic utility and gain-loss function, I expect qualitatively similar results to hold for a larger class of preferences.

Related to the literature, excess smoothness and sensitivity cannot be generated by the standard time- and state-separable model (see, e.g., Ludvigson and Michaelides (2001)). Here, I will briefly focus on the literature that can explain these stylized facts using preferences that generalizes the standard model. See Attanasio and Weber (2010) for a survey of the life-cycle consumption literature.

Models of habit formation can generate such features but, as noted by Chetty and Szeidl (2016), cannot explain the magnitude hypothesis. Moreover, Michaelides (2002) shows that the multiplicative habit model generates excess wealth accumulation when calibrated to

match consumption data. Such an exercise seems to be a natural next step in terms of further investigating AD preferences' ability to match consumption patterns. Focusing on a type of Markov equilibria, Pagel (2017) shows that the model developed by Kőszegi and Rabin (2009) can both explain excess sensitivity (and smoothness) and hump-shaped consumption profiles. The latter widely observed phenomenon can be explained by the model's ability to generate time-inconsistent behavior, which is ruled out by AD preferences. However, Pagel's (2017) analysis shows that the model cannot explain the magnitude hypothesis. Finally, Chetty and Szeidl (2016) builds on Grossman and Laroque (1990) using a model where durable goods feature adjustment costs. The model can generate excess sensitivity in consumption responses that vanishes for large shocks but only for durable goods.

3.8 Conclusion

In this paper, I have studied preferences over infinite-horizon temporal lotteries, conditional on recent anticipations, which can be represented as if the decision-maker dynamically evaluates outcomes as gains and losses relative to anticipations. I have shown that the utility representation can be uniquely identified even when the initial reference point is unobserved. The utility representation is consistent with observed behavior associated with reference dependence and is both portable and tractable. Moreover, its recursive formulation allows for unambiguous welfare comparisons of different policies across different contexts.

I have deliberately sought to keep the model as close to the standard time- and state-additive expected utility model as possible. This is done to isolate the implications of reference dependence with endogenously given reference points. It would be interesting to investigate elaborations of the model that allow for, e.g., inertia in the reference point or time-inconsistent preferences in which the reference point could be used as a commitment device. This and experimental tests of the model are left for future research.

3.9 Appendices

3.9.1 Appendix A: The Construction of D

Although the construction of D is standard (see, e.g., Epstein and Zin (1989) and Chew et al. (1991)), I will use a result that requires the homeomorphic space $\Delta(C \times D)$ to be viewed as a compact subset of a Banach space $(\mathcal{S}, ||f||_{BL^*})$, where $\mathcal{S}$ and the norm $||f||_{BL^*}$ is described in Appendix 3.9.2. Therefore, I will endow both spaces with a nonstandard metric that metricize the weak* topology and coincides with $||f||_{BL^*}$ on $\Delta(C \times D)$. Finally, I will briefly consider an (in my setting) equivalent set-up to D using filtrations.

Let $C^{\mathbb{N}} = C \times C \times C \times \ldots$ be endowed with the product metric, $\sum_{t=1}^{\infty} \frac{d_C}{2^t}$, given any compatible metric d_C on C, implying that $C^{\mathbb{N}}$ is a compact separable metric space. The domain D will be constructed inductively. Let $D_{-1} = C^{\mathbb{N}}$ and for each $t \geq 0$ define

$$D_t = \Delta(C \times D_{t-1})$$

and note that, for each t, D_t is also a compact metric space (the metric on $C \times D_{t-1}$ given inductively by $\frac{1}{2}d_C + \frac{1}{2}d_{D_{t-1}}$). Endow each D_t with the following metric: for any compact and metrizable space X the metric on $\Delta(X)$ is given by

$$d_{BL^*}(\mu, \nu) := \sup \left\{ \left| \int f d\mu - \int f d\nu \right| : f \in BL(X), ||f||_{BL} \leq 1 \right\},$$

where $BL(X)$ is the space of bounded Lipschitz continuous functions on X endowed with the norm

$$||f||_{BL} = \sup_{x \in X} |f(x)| + \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_X(x, y)}.$$

It can be shown that this metric is equivalent to the Kantorovich?Rubinstein metric (see, e.g., Bogachev (2007)). For any $d_1 \in D_1$, it is possible to collapse the uncertainty in d_1 to identify it with an element $d_0 \in \Delta(C^{\mathbb{N}})$. To do this, I define the function

$$f_0 : D_1 \to D_0, \; f_0(d_1)(B) = \mathbb{E}_{d_1}[T_B]$$

for any Borel measurable set B in the Borel sigma for $\Delta(C^{\mathbb{N}})$, where $T_B : C \times \Delta(C^{\mathbb{N}}) \to \mathbb{R}$ defined by

$$T_B(c, \nu) = \nu\{(c', c'', c''', \ldots) \in C^{\mathbb{N}} : (c, c', c'', c''', \ldots) \in B\}.$$

Thus, using induction, define $f_t : D_{t+1} \to D_t$ for $t \geq 1$ by

$$f_t(d_{t+1})(B_t) = d_{t+1}\{(c, d_t) \in C \times D_t : (c, f_{t-1}(d_t)) \in B_t\}$$

for any Borel measurable set B_t in the Borel sigma algebra $\mathcal{B}_t$ for D_t. It is now possible to define the space D by

$$D = \{d = (d_0, d_1, \ldots) : d_t \in D_t \text{ and } d_t = f_t(d_{t+1}) \; \forall t \geq 0\}.$$

Let the compact metrizable space $\prod_{t=0}^{\infty} D_t$ be endowed with the product metric in the following way $d_D = \sum_{t=1}^{\infty} \frac{d_{D_t}}{2^t}$. The metric space (D, d_D) is a subspace of $(\prod_{t=0}^{\infty} D_t, d_D)$ with the inherited metric.

Now consider the projection mapping $\pi_t : D \to D_t$ with $\pi_t^{-1}(\mathcal{B}_t) := \{\pi_t^{-1}(B_t) : B_t \in \mathcal{B}_t\}$ for all $t \geq 0$. Using this notation, it is for all $t \geq 1$ possible to define the map

$$P_{t+1} : \Delta(C \times D) \to \Delta(C \times D_t), \; P_{t+1}d(B_t) \equiv d(\pi_t^{-1}(B_t)), \forall \; B_t \in \mathcal{B}_t.$$

The following observation will be useful in the sequel. Using the above definition, it possible to define $g : D \to \Delta(C \times D)$ by setting $g(d)(B) = \int_B dm(c, \hat{d})$ for any Borel measurable set $B \subset C \times D$ where $P_t g(d) = d_t$ for all $t \geq 0$. It can be shown using an adaptation of the results in, e.g., Epstein and Zin (1989), that g so defined is an homeomorphism between D and $\Delta(C \times D)$.

Endow $\Delta(C \times D)$ with the metric d_{BL^*} and note that if a function $f : D \to \mathbb{R}$ is Lipschitz continuous then $f \circ g : \Delta(C \times D) \to \mathbb{R}$ is also Lipschitz continuous. The latter is a consequence of the equivalence of the Kantorovich-Rubinstein metric and the d_{BL^*} metric together with the homeomorphism g: If f is Lipschitz continuous (normalizing the Lipschitz constant to 1)

then

$$|f(d_1,\ldots)-f(\hat{d}_1,\ldots)|=|f(g(m))-f(g(\hat{m}))|\leq d_D((d_1,\ldots),(\hat{d}_1,\ldots))=$$
$$||d_{D_0}(g_0(m),g_0(\hat{m})),d_{D_1}(g_1(m),g_1(\hat{m})),\ldots||_p=\sum_{t=0}^{\infty}\frac{1}{2^{t+1}}\sup_{||h_t||_{BL}\leq 1}\left|\int h_t dd_t-\int h_t d\hat{d}_t\right|$$
$$\leq\sup_{||h||_{BL}\leq 1}\left|\int h dm-\int h d\hat{m}\right|=d_{\Delta(C\times D)}(m,\hat{m}),$$

where the last inequality follows from that $\hat{h}(c,d)=\sum_{t=0}^{\infty}\frac{\hat{h}_t(c,f_{t-1}(d_t))}{2^{t+1}}$ for $||\hat{h}_t||_{BL}\leq 1$ implies $||\hat{h}||_{BL}\leq 1$. In the rest of the paper, with some abuse of notation I denote $\int dm(c,\hat{d})$ by $\int dm(c,\hat{m})$ and so on.

Finally, using π_0, it is possible to identify each $m\in D$ with a unique probability measure $\mu=\pi_0(m)\in\Delta(C^{\mathbb{N}})$. As in Subsection 3.2.2, let $\Omega=:C^{\mathbb{N}}$and $\{\mathcal{G}_t\}_t$ be a filtration on Ω where $\mathcal{B}$ is the Borel σ-algebra on C, $\mathcal{G}_0=\{\emptyset,\Omega\}$ and, for every $t>0$, $\mathcal{G}_t:=\mathcal{B}^t\times\{\emptyset,C\}^{\infty}$. Now, since Ω is a compact metrizable space, by Theorem 10.2.2 in Dudley (2002, p.345) regular conditional probabilities defined on $\mathcal{B}_0\times\Omega$ exist and are essentially unique. Let ω_t the element with place t in $\omega=(\omega_1,\omega_2,\ldots)$ and $p_{G_t}(\omega)$ is the conditional probability of $\omega\in\Omega$ occurring given $G_t\in\mathcal{G}_t$. Therefore, for all $a\in D$, $\succsim_a$ can be represented by

$$V_a(m)=\mathbb{E}_{\mu}\left[\phi(u(\omega_1)-r(a^1))+\sum_{t=1}^{\infty}\beta^{t-1}\max_{r\in[\underline{u},\bar{u}]}\left(r+\int_{G_{t-1}}\beta\phi(u(\omega_t)-r)dp_{G_{t-1}}(\omega)\right)\right]$$
$$=\mathbb{E}_m\left[\phi(u(c_1)-r(a^1))+\sum_{t=1}^{\infty}\beta^{t-1}\max_{r\in[\underline{u},\bar{u}]}\left(r+\beta\mathbb{E}\left[\phi(u(c_{t+1})-r)|\mathcal{G}_t\right]\right)\right].$$

3.9.2 Appendix B: Proof of Theorem 1 and 4

I begin with the converse which is similar for both Theorem 1 and 4.

Dekel et al. (2007, Lemma 1) show that if $\succsim_a$ has an affine representation V_a that is Lipschitz continuous, then $\succsim_a$ satisfies the Lipschitz continuity axiom. For degenerate expectations continuity, note that $u(c)\in\arg\max_{r\in\mathbb{R}}\{u(c)+\beta\phi(u(c)-r)\}$ by ϕ being β-consistent with u. Moreover, note that if $\beta\phi(y)<y$ for all $y\neq 0$, it is obvious that for any neighborhood $U\subset\Delta(C)$ of δ_c there exists a neighborhood $V\subset U$ such that $r:\Delta(C)\to\mathbb{R}$ is continuous on V. If $\beta\phi(y)=y$ for $y\neq 0$, then $\beta\phi(x)\geq\beta\phi(y+x)-y$ for all $x\in[\min u,\max u]$.

By the above reasoning then any potential candidate has to use the reference point $r \in [-a, a]$ such that $u(c) - r$ equals such an y. This implies that

$$\beta \int \phi(u(c') - u(c)) d\hat{m}^1(c') \geq \beta \int \phi(u(c') - u(c) + y) d\hat{m}^1(c') - y \quad \forall \hat{m}^1 \in \Delta(C) \Longrightarrow$$
$$u(c) + \beta \int_{C \setminus \{c\}} \phi(u(c') - u(c)) dm^1(c') \geq r + \beta \phi(u(c) - r) m^1(\{c\}) + \beta \int_{C \setminus \{c\}} \phi(u(c') - r) dm^1(c').$$

Thus, it is weakly optimal to use the reference point $u(c)$ over r so it is always possible to specify the optimal reference point such that it is continuous at degenerate expectations.

In the light of this, it is not hard to verify that axioms 1-7 hold for any family of binary relations that has an AD representation (u, ϕ, r, β).

The remainder of the proof shows that if a nondegenerate family of preferences satisfies axioms 1-7, then it has an AD representation. The analogous proof for Theorem 4 follows afterwards. The proof is divided into several steps. I begin with a series of Lemmas that implies that each $\succsim_a$ can be represented by an additively time-separable expected utility function that is Lipschitz continuous.

Lemma 2. *For any $a \in D$, if $\succsim_a$ satisfies weak order, von-Neumann and Morgenstern continuity and independence, then there exists an affine $V_a : D \to \mathbb{R}$ that represents $\succsim_a$ on D. Moreover, each V_a is unique up to an affine transformation.*

Proof. This result follows from the mixture space theorem (see, e.g., Fishburn (1970, Theorem 8.4, p.112)).

□

Lemma 3. *For any $a \in D$, if $\succsim_a$ can be represented by an affine function $V_a : D \to \mathbb{R}$, then V_a is Lipschitz continuous if $\succsim_a$ satisfies Lipschitz continuity.*

Proof. First, note that since $d_{\Delta(C \times D)}(m, \hat{m}) = d_D(g^{-1}(m), g^{-1}(\hat{m}))$ (i.e. D and $\Delta(C \times D)$ are isometric), there is no loss in continuing talking about the identification of m with elements in D (that is $f : D \to \mathbb{R}$ is Lipschitz continuous iff $f \circ g^{-1} : \Delta(C \times D) \to \mathbb{R}$ is Lipschitz continuous). It follows from Lemma 1 in Dekel et al. (2007) that V_a is Lipschitz continuous

as $d_{\Delta(C\times D)}(\lambda m, \lambda\hat{m}) = |\lambda| d_{\Delta(C\times D)}(m, \hat{m})$ for $\lambda \in \mathbb{R}$ and $d_{\Delta(C\times D)}$ is shift invariant (it can be seen as a norm on the Banach space consisting of the space of Borel signed measures with bounded variation on $C \times D$).

□

Lemma 4. *For any $a \in D$, if $\succsim_a$ can be represented by a continuous and affine $V_a : D \to \mathbb{R}$ and satisfies separability, there are Lipschitz continuous utility functions $w_a : C \to \mathbb{R}$ and $h_a : D \to \mathbb{R}$ such that $V_a(m) = \int (w_a(c) + h_a(\hat{m}))\, dm(c, \hat{m})$ represents $\succsim_a$.*

Proof. Following Gul and Pesendorfer (2004), $V_a\left(\frac{1}{2}(c,m) + \frac{1}{2}(\hat{c},\hat{m})\right) = V_a\left(\frac{1}{2}(c,\hat{m}) + \frac{1}{2}(\hat{c},m)\right)$ and the affinity of V_a implies that there exists a function $\hat{W}_a : C \times D \to \mathbb{R}$ such that $V_a(m) = \int \hat{W}(c,\hat{m}) dm(c,\hat{m})$ and, therefore, $V_a(c,m) = \hat{W}_a(c,\hat{m}) + \hat{W}_a(\hat{c},m) - \hat{W}_a(\hat{c},\hat{m})$. Thus,

$$\begin{aligned} V_a(m') &= \int \hat{W}_a(c,m) dm'(c,m) \\ &= \int \hat{W}_a(c,\hat{m}) dm'(c,m) + \int \hat{W}_a(\hat{c},m) dm'(c,m) - \int \hat{W}_a(\hat{c},\hat{m}) dm'(c,m). \end{aligned}$$

Therefore, setting $w_a = \hat{W}_a(\cdot,\hat{m}) - \hat{W}_a(\hat{c},\hat{m})$ and $h_a = \hat{W}_a(\hat{c},\cdot)$ gives us the result. It remains only to verify that both w_a and h_a are Lipschitz continuous given their respective domains and their implied metrics. This is easily seen by either fixing the distribution of period 1 consumption or the distribution of continuation lotteries.

□

Lemma 5. *For any $a \in D$, if $\succsim_a$ also satisfies anticipation compensation then it can be represented by*

$$V_a(m) = \int (v_a(c) + \beta h(\hat{m}))\, dm(c,\hat{m})$$

where $h(m_0) = 0$ for some $m_0 \in D$.

Proof. For any $a, \hat{a}, a' \in D$, given reference independence, there exists $c, \hat{c} \in \mathbb{R}$ such that $a = (c,m)$ and $\hat{a} = (\hat{c},\hat{m})$. By anticipation compensation, it is possible to find $c', \hat{c}' \in C$

(trivially, $c = c'$ and $\hat{c} = \hat{c}'$) such that

$$\frac{1}{2}(\bar{c}, \delta_{(c,m')}) + \frac{1}{2}(\bar{c}, \delta_{(\hat{c}',m')}) \sim_{a'} \frac{1}{2}(\bar{c}, \delta_{(c',m')}) + \frac{1}{2}(\bar{c}, \delta_{(\hat{c},m')})$$

for any $m' \in D$. Therefore, for any $m, \hat{m} \in D$

$$V_a(c, m) \geq V_a(c, \hat{m}) \iff V_{\hat{a}}(\hat{c}, m) \geq V_{\hat{a}}(\hat{c}, \hat{m})$$

implying that

$$h_a(m) \geq h_a(\hat{m}) \iff h_{\hat{a}}(m) \geq h_{\hat{a}}(\hat{m}).$$

By the uniqueness of additively separable utility representations (see Debreu (1960)), for any $a, \hat{a} \in D$, it is without loss of generality to set $h_a = \sigma + \gamma h_{\hat{a}}$ where $\sigma \in \mathbb{R}$ and $\gamma > 0$. Thus, letting $v_a = \dfrac{V_a - \sigma}{\gamma}$ for all $a \neq \hat{a} \in D$ such that $\beta h = h_a$ for $\beta > 0$ where h is normalized such that $\beta h(m_0) = 0$ for some fixed $m_0 \in D$. This gives $V_a = v_a + \beta h$, since V_a is unique up to an affine transformation.

□

Lemma 6. *For any $a \in D$, if $\succsim_a$ also satisfies preference for early resolution of uncertainty then h is a convex function.*

Proof. By the preference for early resolution of uncertainty axiom,

$$\begin{aligned} V_a(\alpha\delta_{(c,m)} + (1-\alpha)\delta_{(c,\hat{m})}) = \alpha V_a(c, m) + (1-\alpha)V_a(c, \hat{m}) &= v_a(c) + \alpha\beta h(m) + (1-\alpha)\beta h(\hat{m}) \\ &\geq V_a(c, \alpha m + (1-\alpha)\hat{m}) = v_a(c) + \beta h(\alpha m + (1-\alpha)\hat{m}) \end{aligned}$$

where the first equality follows from the affinity of V_a.

□

For any Banach space X, let X^* denote the space of all continuous linear functionals on X. The space X^* is called the (norm) dual of X with the duality given by $\langle x, x^* \rangle = \int x^*(a)dx(a)$. Since h is convex, it can be represented as the supremum over a set of affine functions (see,

e.g., Aliprantis and Border (2006) Theorem 7.6). If the supremum is attained, the set of affine functions that attains it are called its *subdifferential*. The subdifferential of a function $f : Y \to \mathbb{R}$ for $Y \subset X$ at $x \in X$ is defined by

$$\partial h(m) = \{x^* \in X^* : \langle y - x, x^* \rangle \leq f(x) - f(y) \text{ for all } y \in X\}.$$

Note that it may well be that this set is empty for a convex function defined on an infinite dimensional space. Finally, the *conjugate* function $f^* : X^* \to \mathbb{R} \cup \{+\infty\}$ is defined by $f^*(x^*) = \sup_{x \in Y}[\langle x, x^* \rangle - f(x)]$.

For any compact and metrizable set S, denote by $ca(S)$ the set of all signed (countable additive) Borel probability measures bounded in variation on S. Consider the set

$$\mathcal{D} := span\{\delta_s | s \in S\} = \left\{ \sum_{k=1}^{n} \alpha_k \delta_{s_k} : n \in \mathbb{N}, \alpha_k \in \mathbb{R}, s_k \in S \right\}$$

which can be embedded in the dual space of the set of bounded Lipschitz continuous functions on S, denoted BL^*, with the norm

$$||\cdot||_{BL^*} := \sup \left\{ \left| \int f d\mu - \int f d\nu \right| : f \in BL(S), ||f||_{BL} \leq 1 \right\}$$

(Hille and Worm, 2009, Lemma 3.5). Let $\mathcal{S}_{BL}$ be the closure of $\mathcal{D}$ in BL^* with respect to $||\cdot||_{BL^*}$. Theorem 3.11 in Hille and Worm (2009) shows that it is possible to identify $ca(S)$ with a $||\cdot||_{BL^*}$-dense subspace of $\mathcal{S}_{BL}$. Therefore, the $||\cdot||_{BL^*}$-closure of $ca(S)$ can be taken to be $\mathcal{S}_{BL}$.

It is easy to see that the metric d_{BL^*} on $\Delta(S)$ coincides with the norm $||\cdot||_{BL^*}$ on the same space. Moreover, Theorem 3.8 in Hille and Worm (2009) implies that $\Delta(S)$ is norm-closed in BL^* since S is complete. Finally, Theorem 3.7 in Hille and Worm (2009) states that the dual space of $\mathcal{S}_{BL}$ can be identified with $BL(S)$.

Now, take $S = C \times D$ and note that $\Delta(C \times D)$ is a convex and compact (in the relative topology) subset of $(\mathcal{S}_{BL}, ||\cdot||_{BL^*})$ and, hence, a Baire space. Note that the set of probability measures span the set of all signed measures. Therefore, using the normalization $h(m_0) = 0$ implies that the zero element, $\mathbf{0}$, is an element in $\Delta(C \times D)$, it follows that

$\text{span}(\Delta(C \times D)) = \text{aff}(\Delta(C \times D))$ is dense in $ca(C \times D)$ which in turn is dense in $\mathcal{S}_{BL}$.

By the above facts, and since h is convex and Lipschitz continuous on D, it is possible to use results from Ergin and Sarver (2010b, Lemma 1.1, Theorem 1.2, and Theorem 2.4) to show that:

Theorem 6. *Ergin and Sarver (2010b):*

(1) $h(m) = \max_{x^* \in BL(C \times D)}\{\langle m, x^* \rangle - f^*(x^*)\}$ *where* $f^* : BL(C \times D) \to \mathbb{R} \cup \{+\infty\}$ *is the conjugate function.*

(2) There exists a unique minimal weak closed and compact set* $\mathcal{M}_h$ *such that*

$$h(m) = \max_{x^* \in \mathcal{M}_h} \{\langle m, x^* \rangle - f^*(x^*)\}.$$

(3) The set of points m *where* $\partial h(m)$ *is a singleton is a dense* G_δ *set (in the relative topology) in* $\Delta(C \times D)$.

(4) $\mathcal{M}_h$ *is the closure in the weak* topology of the set*

$$\mathcal{N}_h = \{x^* \in BL(C \times D) : x^* \in \partial h(m) \text{ for some } m \in D \text{ s.t. } \partial h(m) \text{ is a singleton}\}.$$

Note that the completeness of $BL(C \times D)$ implies that $x^* \in \mathcal{M}_h$ is such that $x^* \in BL(C \times D)$. A standard result is that for a proper convex function $f : X \to \mathbb{R}$, the subdifferential is a singleton at a point x iff f is Gâteaux differentiable at x (where the derivate f' is such that $\partial f = \{f'\}$) (Aliprantis and Border, 2006, Corollary 7.17, p.268).[29] Thus, h is Gâteaux differentiable on a dense set.

In what follows, assume that for all $a \in D$, $\succsim_a$ also satisfies dynamic consistency, anticipation, and nondegeneracy.

Lemma 7. *For all* $c \in C$ *and* $m \in D$, $\partial h(c, m) \cap \mathcal{M}_h \subset \{\sigma + \gamma V_{(c,m)} : \sigma \in \mathbb{R}, \gamma > 0\}$.

[29] An example of a proper convex function is the extension of a continuous convex function on a nonempty compact subset Y of a Banach space extended to the whole space by letting $f(x) = +\infty$ on $x \in X \setminus Y$.

Proof. Fix arbitrary $c \in C$ and $m \in D$. Since $V_a(c,m)$ is continuous in a around each $e = (c,m) \in D$ and each V_a are continuous. Therefore, anticipation compensation implies that for any neighborhood U of (c,m) there exists a neighborhood $V \subset U$ containing (c,m) such that if $V_{(c,m)}(\hat{m}) > V_{(c,m)}(\bar{m})$ then $V_{m'}(\hat{m}) > V_{m'}(\bar{m})$ for all $m' \in V$.

Take any sequence $\{m_k\}_{k=1}^{\infty} \in V$ such that h is Gâteaux differentiable at each m_k. Since h is Gâteaux differentiable on a dense set of D, such a sequence can always be found. I will show that for any $m', \hat{m} \in D$ such that $V_{m_k}(m') > V_{m_k}(\hat{m})$, it must be the case that $x^*_{m_k}(m') \geq x^*_{m_k}(\hat{m})$ for $\{x^*_{m_k}\} = \partial h(m_k)$. Using Corollary B.3 in Ghirardato et al. (2004), if $x^*_{m_k}$ is nonconstant (V_{m_k} is nonconstant by nondegeneracy) this implies that there exists $\sigma_k \in \mathbb{R}$ and $\gamma_k > 0$ such that $V_{m_k} = \sigma_k + \gamma_k x^*_{m_k}$.

Assume towards a contradiction that $V_{m_k}(m') > V_{m_k}(\hat{m})$ and $x^*_{m_k}(m') < x^*_{m_k}(\hat{m})$ for $m', \hat{m} \in D$. Let $m_k \alpha m' = (1-\alpha)m_k + \alpha m'$ with $m_k \alpha \hat{m}$ defined similarly. By the above continuity observation, there exists an $\bar{\alpha} > 0$ such that $V_{m_k \alpha \hat{m}}(m_k \alpha \hat{m}) < V_{m_k \alpha \hat{m}}(m_k \alpha m')$ and $x^*_{m_k}(m_k \alpha m') < x^*_{m_k}(m_k \alpha \hat{m})$ for all $\alpha \in (0, \bar{\alpha})$. Since $x^*_{m_k}$ is the Gâteaux differential of h at m_k, it follows that

$$\lim_{\alpha \to 0} \frac{h(m_k \alpha m') - h(m_k) - h(m_k \alpha \hat{m}) + h(m_k)}{\alpha} = x^*_{m_k}(m' - \hat{m}) < 0.$$

This implies that $h(m_k \alpha \hat{m}) > h(m_k \alpha m')$ and $V_{m_k \alpha \hat{m}}(m_k \alpha \hat{m}) < V_{m_k \alpha \hat{m}}(m_k \alpha m')$, contradicting dynamic consistency. Finally, if $x^*_{m_k}$ is a constant function, then it must be the case that $m_k = \boldsymbol{c}_*$, that is, the worst possible element as $h(m) \geq x^*_{m_k}(m')$ for all $m, m' \in D$.

Now, notice that the sequence $\{x^*_{m_k}\}_{k=1}^{\infty}$ has a convergent subsequence since $\mathcal{M}_h$ is weak* compact. Without loss of generality suppose the sequence itself converges. I claim that $x^*_{m_k} \xrightarrow{w^*} x^*_{(c,m)}$ where $x^*_{(c,m)} = \sigma + \gamma V_{c,m}$ for $\sigma \in \mathbb{R}$ and $\gamma > 0$. Suppose not, then there is an $\varepsilon > 0$ and a convergent subsequence $\{x^*_{m_l}\}_{l=1}^{\infty}$ of $\{x^*_{m_k}\}_{k=1}^{\infty}$ with limit $x^*_{(c,m)}$ such that $||\sigma_l + \gamma_l x^*_{m_l} - V_{(c,m)}|| > \varepsilon$ for all l, $\sigma_l \in \mathbb{R}$ and $\gamma_l > 0$. By continuity of $V_{(\cdot)}(m')$ at (c,m) for any m', there exists an $L > 0$ such that $\sigma_l + \gamma_l x^*_{m_l} \neq V_{m_l}$ for all $l > L$, $\sigma_l \in \mathbb{R}$ and $\gamma_l > 0$, a contradiction to the construction of the sequence.

To summarize, $m_k \to (c,m)$, $x^*_{m_k} \xrightarrow{w^*} x^*_{(c,m)}$ and $x^*_{m_k} \in \partial h(m_k)$ for all k. What is left to show is that $x^*_{(c,m)} \in \partial h(c,m)$. Since, $\mathcal{M}_h$ is compact, the sequence $\{x^*_{m_k}\}_{k=1}^{\infty}$ is

norm-bounded. By the definition of the subdifferential and continuity of h, for any $\hat{m} \in D$,

$$\langle \hat{m} - (c,m), x^* \rangle = \lim_k \langle \hat{m} - m_k, x^*_{m_k} \rangle \leq \lim_k [h(\hat{m}) - h(m_k)] = h(\hat{m}) - h(c,m)$$

where the first inequality follows from a standard result (see, e.g., Ergin and Sarver (2010b, p.5)). Thus, conclude that $x^*_{(c,m)} \in \partial h(c,m)$.

Since any $x^* \in \mathcal{M}_h$ is either in $\mathcal{N}_h$ or is the limit of some sequence in $\mathcal{N}_h$ and the above result holds for any such sequence with limit (c,m), it must be the case that $\partial h(c,m) \cap \mathcal{M}_h \subset \{\sigma + \gamma V_{(c,m)} : \sigma \in \mathbb{R}, \gamma > 0\}$. Finally, since (c,m) was arbitrary conclude that it holds for all $(c,m) \in D$.

□

Lemma 8. *For all $c \in C$ and $m \in D$, $\partial h(c,m) \cap \mathcal{M}_h = \{\sigma_c + V_{(c,m)}\}$ where $\sigma_c \in \mathbb{R}$.*

Proof. Note that for all $c, \hat{c} \in C$ and all $m, \hat{m}, a$ future separability implies that

$$\frac{1}{2}(c', \delta_{(\hat{c},\hat{m})}) + \frac{1}{2}(c', \delta_{(c,m)}) \sim_a \frac{1}{2}(c', \delta_{(\hat{c},m)}) + \frac{1}{2}(c', \delta_{(c,\hat{m})}). \tag{3.24}$$

Now, assume to get a contradiction that

$$max_{x^* \in \mathcal{M}_h} \{f^*(x^*) + x^*(c,m)\} = \sigma + \gamma V_{\delta_{(c,m)}}(c,m)$$

and

$$max_{\hat{x}^* \in \mathcal{M}_h} \{f^*(\hat{x}^*) + \hat{x}^*(\hat{c},m)\} = \hat{\sigma} + \hat{\gamma} V_{\delta_{(\hat{c},m)}}(\hat{c},m)$$

where $\gamma \neq \hat{\gamma}$. By nondegeneracy, there exists an $\hat{m}$ with $h(m) \neq h(\hat{m})$ which implies that equation (3.24) depends on the choice of m and $\hat{m}$, contradicting separability. Thus, it must be the case that $\partial h(c',m) \cap \mathcal{M}_h = \{\sigma + \gamma V_{(c',m)}\}$ and $\partial h(\bar{c},m) \cap \mathcal{M}_h = \{\hat{\sigma} + \gamma V_{(\bar{c},m)}\}$ for any $m \in D$.

Since C is connected, this must be the case for all $(c,m) \in D$. Conclude that for any two $x^*, \hat{x}^*$ that are optimal against some $(c,m), (\hat{c},\hat{m}) \in D$ are such that the respective γ and $\hat{\gamma}$ associated with them are equal. Since h is multiplied by a constant $\beta > 0$, there is no loss of

generality to normalize $\gamma = 1$.

□

Lemma 9. $\mathcal{M}_h = \{\sigma_{\hat{c}} + V_{(\hat{c},m)} : \hat{c} \in C\}$.

Proof. Anticipation certainty implies that if $\succsim_m = \succsim_{\hat{m}}$, then

$$\alpha(c, m) + (1 - \alpha)(c, \hat{m}) \sim_a (c, \alpha m + (1 - \alpha)\hat{m})$$

for all $\alpha \in [0, 1]$. Since for all $m \in D$, $\succsim_m = \succsim_{(c,\hat{m})}$ for some $c \in C$ and $\hat{m} \in D$, anticipation certainty then implies that

$$\alpha(c, \delta_{(\hat{c},\hat{m})}) + (1 - \alpha)(c, m) \sim_a \left(c, \alpha\delta_{(\hat{c},\hat{m})} + (1 - \alpha)m\right)$$

for all $\alpha \in [0, 1]$.

Since $\partial h(\hat{c}, \hat{m}) \cap \mathcal{M}_h = \{\sigma_{\hat{c}} + V_{(\hat{c},\hat{m})}\}$ for some $\sigma_{\hat{c}} \in \mathbb{R}$ as $\partial h(\hat{c}, \hat{m})$ is independent of $\hat{m}$, it must be the case that $\partial h(m) \cap \mathcal{M}_h = \{\sigma_{\hat{c}} + V_{(\hat{c},\hat{m})}\}$ too, otherwise

$$\alpha(c, \delta_{(\hat{c},\hat{m})}) + (1 - \alpha)(c, m) \nsim_a \left(c, \alpha\delta_{(\hat{c},\hat{m})} + (1 - \alpha)m\right)$$

for some $\alpha \in [0, 1]$ contradicting anticipation certainty. Since $m \in D$ was arbitrary, it follows that $\mathcal{M}_h = \{\sigma_{\hat{c}} + V_{(\hat{c},\hat{m})} : \hat{c} \in C\}$.

□

Lemma 10. *For all $a \in D$, $v_a(c) = \hat{v}_a(u(c))$ where $u : C \to \mathbb{R}$ is nonconstant and Lipschitz continuous and $\hat{v}_a : [\min u, \max u] \to \mathbb{R}$ is such that $\hat{v}_a(u)$ is nonconstant and Lipschitz continuous.*

Proof. Due to Lemma 9, independence implies that for any consumption lotteries of the form $m' = \alpha\delta_{(c',c,m)} + (1 - \alpha)\delta_{(c',\hat{c},m)}$, $\succsim_a$ can be represented by

$$V_a(m') = v_a(c') + \alpha u(\hat{c}) + (1 - \alpha)u(c) + k(m) \tag{3.25}$$

where $u : C \to \mathbb{R}$ and $k : D \to \mathbb{R}$. For the rest of the lemma, I fix m and disregard k. Consider any two degenerate temporal lotteries $\hat{m} = (c, \delta_{(\hat{c},m)}), m' = (c, \delta_{(c',m)}) \in D$ and note that

$$|V_a(\hat{m}) - V_a(m')| = |u(\hat{c}) - u(c')| \leq M d_{\Delta(C\times\Delta(C\times D))}(\hat{m}, m') = M d_C(\hat{c}, c')$$

for $M > 0$, so u is Lipschitz continuous.

From equation (3.25), we see that any $c\frac{1}{2}\hat{c}' \sim^I c'\frac{1}{2}\hat{c}$ is equivalent to

$$\begin{aligned} V_a\left(\frac{1}{2}\delta_{(c^*,c,m^*)} + \frac{1}{2}\delta_{(c^*,\hat{c}',\bar{m}^*)}\right) &= v_a(c^*) + \beta\frac{1}{2}(u(c) + u(\hat{c}') \\ &= v_a(c^*) + \beta\frac{1}{2}(u(c') + u(\hat{c}')) \\ &= V_a\left(\frac{1}{2}\delta_{(c^*,c',\bar{m}^*)} + \frac{1}{2}\delta_{(c^*,\hat{c},m^*)}\right). \end{aligned}$$

That is, $u(c) = c_r + u(\hat{c}') - u(c')$. By anticipation, for any $c_r, c_{\hat{r}} \in C$ whenever $\bar{c}\frac{1}{2}c_r \sim^I c\frac{1}{2}c_{\hat{r}}$ and $\hat{c}\frac{1}{2}c_r \sim^I c'\frac{1}{2}c_{\hat{r}}$,

$$V_{(c_r,a)}(c, m) \geq V_{(c_r,a)}(c', \hat{m}) \iff V_{(\hat{r},\hat{a})}(\bar{c}, m) \geq V_{(\hat{r},\hat{a})}(\hat{c}, \hat{m}).$$

Now, defined the map $r : D \to \mathbb{R}$ that satisfies $r(a) = u(c)$ for all $\hat{a} \in D$ if $\succsim_a = \succsim_{(c,m)}$ and note that this map can always be made surjective.

Let c_x be the equivalent class for which $u(c) = x$ for all $c \in c_x$. Then $v_a(c) = v_a(\hat{c})$ for all $c, r \in \mathbb{R}_x$. To see this, note that by anticipation compensation,

$$v_a(c) = v_a(\hat{c}) \iff v_{\hat{a}}(c') = v_{\hat{a}}(\hat{c}')$$

for $u(c') = u(c) - r(a) + r(\hat{a})$ and $u(\hat{c}') = u(\hat{c}') - r(a) + r(\hat{a})$. Let $c = \hat{c}$, then for all $\bar{c} \in C$ such that $u(\bar{c}) = u(c')$ it must be the case that $v_{\hat{a}}(\bar{c}) = v_{\hat{a}}(c')$. This can be done for an arbitrary $c' \in C$. Therefore, it is without loss to write $v_a(c) = \hat{v}_a(u(c))$ where $\hat{v}_a : [\min u, \max u] \to \mathbb{R}$ is a continuous function.

Clearly, $\hat{v}_a(u)$ is Lipschitz continuous and the nonconstant part follows from nondegeneracy.

□

Lemma 11. *$\hat{v}_a(u) = \phi(u - r(a))$ with $\phi : [-a, a] \to \mathbb{R}$ for $a = \max u - \min u$ and normalized such that $\phi(0) = 0$, where $r : D \to \mathbb{R}$ satisfies $r(a) = u(c)$ for some $c \in C$ and $m \in D$ such that $\succsim_a = \succsim_{(c,m)}$.*

Proof. Given equation (3.25), let $c'\frac{1}{2}c_{r(a)} \sim^I c\frac{1}{2}c_{r(\hat{a})}$ and $\hat{c}\frac{1}{2}c_{r(a)} \sim^I \hat{c}'\frac{1}{2}c_{r(\hat{a})}$, where $u(c_{r(a)}) = r(a)$. Reference compensation then implies that

$$\begin{aligned}
&(c, m) \succsim_a (\hat{c}, \hat{m}) \\
\iff &(c', m) \succsim_{\hat{a}} (\hat{c}', \hat{m}) \\
\iff &\hat{v}_a(u(c)) + \beta h(m) \geq \hat{v}_a(u(\hat{c})) + \beta h(\hat{m}) \\
\iff &\hat{v}_{\hat{a}}(u(c')) + \beta h(m) = \hat{v}_{\hat{a}}(u(c) + r(\hat{a}) - r(a)) + \beta h(m) \geq \\
&\hat{v}_{\hat{a}}(u(\hat{c}')) + \beta h(\hat{m}) = \hat{v}_{\hat{a}}(u(\hat{c}) + r(\hat{a}) - r(a)) + \beta h(\hat{m}).
\end{aligned}$$

By uniqueness of additively separable representations, this implies that it is without loss of generality to write $v_a = \hat{v}_a(u) = \phi(u - r(a))$ for all $a \in D$ with $\phi : [-a, a] \to \mathbb{R}$ being continuous, $a = \max u - \min u$, and normalized such that $\phi(0) = 0$.

□

Lemma 12. *$\beta \in (0, 1)$.*

Proof. For any $a \in D$, consider any pair of lotteries $(c_{r(a)}, \hat{c}, m), (c_{r(a)}, c', m) \in D$ such that $(c_{r(a)}, \hat{c}, m) \succsim_a (c_{r(a)}, c', m)$. Such a pair exists by nondegeneracy. For any fixed m, define inductively a pair of sequences $\{\hat{m}_n\}$ and $\{\hat{m}'_n\}$ by:

$$\hat{m}_n = (c_{r(a)}, \underbrace{\hat{c}, \ldots, \hat{c}}_{n}, m) \text{ and } \hat{m}'_n = (c_{r(a)}, \underbrace{c', \ldots, c'}_{n}, m).$$

Note that the former sequence is always preferred to the latter. Since D is compact there exist convergent subsequences $\{\hat{m}_{n_k}\}$ and $\{\hat{m}'_{n_k}\}$ converging to $\hat{m}$ and $\hat{m}'$ in D, respectively. By continuity, $V_a(\hat{m}_{n_k})$ and $V_a(\hat{m}'_{n_k})$ converges to $V_a(\hat{m})$ and $V_a(\hat{m}')$, respectively. Thus, the difference $V_a(\hat{m}_{n_k}) - V_a(\hat{m}'_{n_k})$ converges to some real number. This difference can be written

as:

$$V_a(\hat{m}_{n_k}) - V_a(\hat{m}'_{n_k}) = \sum_{i=1}^{n_k} \beta^i u(\hat{c}) + \beta^{n_k+1} h(m) - \sum_{i=1}^{n_k} \beta^i u(c') - \beta^{n_k+1} h(m)$$
$$= \sum_{i=1}^{n_k} \beta^i (u(\hat{c}) - u(c')).$$

Since by construction $u(\hat{c}) > u(c')$, it is clear that this difference convergences to a real number only if $\beta \in (0, 1)$.

□

Lemma 13. *$\mathcal{M}_h = \{r(a) + V_a : a \in D\}$ with $V_a = \phi(u - r(a)) + \beta h$ for all $a \in D$.*

Proof. Notice that $r((c, m)) = \sigma_c + \phi(0)$ for all $c \in C$ and $m \in D$ and that $c \in \arg\max_{\hat{c} \in C}\{\sigma_c + \beta\phi(u(c) - u(\hat{c}))\}$. By the normalization $\phi(0) = 0$, this implies that $u(c) = \sigma_c$ for all $c \in C$. Moreover, it must be the case that $\beta\phi(x) \leq x$, otherwise $\hat{c} = c$ is not optimal against $c \in C$. Thus, it is without loss of generality to write

$$\mathcal{M}_h = \{r + \beta[\phi(u - r) + \beta h] : r \in \mathbb{R}\}, \tag{3.26}$$

where ϕ is extended to $\mathbb{R}$ such that it is never optimal to have a references point r not in $[\min u, \max u]$. Then, $r : D \to \mathbb{R}$ can always be defined such that

$$r(m) \in \arg\max_{r \in \mathbb{R}} [r + \beta \int \phi(u(c) - r) dm^1(c)]. \tag{3.27}$$

□

Lemma 14. *ϕ is β-compatible with u.*

Proof. Since $\phi(u - r)$ is Lipschitz continuous for all $r \in [\min u, \max u]$ and ϕ is nonconstant by nondegeneracy, what is left to show is that ϕ is nondecreasing and that $\beta\phi(y) = y$ implies $\beta\phi(x) \geq \beta\phi(y + x) - y$ for all $x \in [a_*, a^*]$ where $a^* = \min\{a - y, a\}$ and $a_* = -\min\{a, a + y\}$. For the former, note that $\bar{c}\frac{1}{2}c_a \succsim^I c\frac{1}{2}c_{\hat{a}}$ and $c'\frac{1}{2}c_{\hat{a}} \sim^I \hat{c}\frac{1}{2}c_a$ implies that $u(\bar{c}) - u(c_{\hat{a}}) \geq$

$u(c) - u(c_a)$ and $u(c') - u(c_{\hat{a}}) = u(\hat{c}) - u(c_a)$ for $c_a, c_{\hat{a}}, c', \bar{c} \in C$. Now assume towards a contradiction that ϕ is strictly decreasing on some nonempty interval $(x, y) \subseteq [-a, a]$. Let $x = u(c) - u(c_a)$ and $u(\bar{c}) - u(c_{\hat{a}}) \in (x, y)$, then

$$(c, m) \sim_{(c_a, a)} (c', \hat{m}) \iff \phi(u(c) - u(c_a)) = \phi(u(c') - u(c_a)) + \beta(h(\hat{m}) - h(m))$$

but $\phi(u(\bar{c}) - u(c_{\hat{a}})) < \phi(u(\hat{c}) - u(c_{\hat{a}})) + \beta(h(\hat{m}) - h(m))$, contradicting anticipation compensation as $(c, m) \succsim_{(c_a, a)} (c', \hat{m})$ but $(\hat{c}, \hat{m}) \succ_{(c_{\hat{a}}, \hat{a})} (\bar{c}, m)$.

Assume to get a contradiction that $\beta\phi(y) = y$ and $\phi(x) < \phi(y + x) - y$ for some y and $x \in [a_*, a^*]$ where $a^* = \min\{a - y, a\}$ and $a_* = -\min\{a, a + y\}$. Then, there exists $\bar{\alpha} > 0$ such that for any $\alpha \in (0, \bar{c})$,

$$u(c) \notin \arg\max_{r \in \mathbb{R}} \{r + \alpha\beta\phi(u(c') - r) + (1 - \alpha)\beta\phi(u(c) - r)\}$$

for some $c, \hat{c}' \in C$ such that $u(c) - u(\hat{c}') = y$ as

$$\begin{aligned} u(c) + \alpha\beta\phi(u(c') - u(c)) < u(c') + \alpha\beta\phi(u(c') - u(\hat{c}')) + (1 - \alpha)\beta\phi(u(c) - u(\hat{c}')) \\ \iff y + \beta\phi(u(c') - u(c)) < \beta\phi(u(c') - u(\hat{c}')), \end{aligned}$$

where $x = u(c') - u(c)$ and $y + x = u(c') - u(\hat{c}')$. This contradicts the continuity of $\succsim_a$ at (c, m).

□

Finally, to sum up the representation Theorem part of Theorem 1, I have shown that for each $a \in D$, $\succsim_a$ can be represented by

$$V_a(m) = \int [\phi(u(\hat{c}) - r(a)) + r(\hat{m}) + \beta V_{\hat{m}}(\hat{m})] \, dm(\hat{c}, \hat{m}), \ \forall m \in D. \tag{3.28}$$

Clearly, it is without loss of generality to redefine $r : \Delta(C) \to \mathbb{R}$ where $r(a^1) = r(a)$ with a^1 being the marginal distribution over C given a, as the continuation lottery does not matter for the optimal reference point.

I now show the uniqueness part of Theorem 1. First, by the mixture space theorem and

the uniqueness of additively separable representations, standard arguments imply that u and r are unique up to a joint affine transformation, β is unique, and $\phi(u - r(a))$ is unique up to an affine transformation (note here that the normalization of $\phi(0) = 0$ is without loss). Moreover, it must be the case that for two AD representations $(u_1, \phi_1, r_1, \beta_1)$ and $(u_2, \phi_2, r_2, \beta_2)$, if $u_1 = \sigma + \gamma u_2$ then

$$\phi_1(u_1 - r(a)) = \gamma\phi_2(u_2 - u - r(a)) = \phi_2(\gamma(u_1 - r(a)))$$

which is the case iff $\phi_2(\gamma(u_1 - r(a))) = \gamma\phi_2(u_1 - r(a))$, that is, if ϕ is homogenous of degree 1. This finishes the proof of Theorem 1.

I now continue with the last part of the proof of Theorem 4, that is, if each $\succsim$ satisfies axiom 3, 4, 6, 9, then it has a utility representation as in the statement of the theorem.

First, note that the antecedents in Lemma 2 and 4 holds if $\succsim$ satisfies axiom 3, 4 and 9. Thus, there are continuous and Gateaux differentiable utility functions $w_a : C \to \mathbb{R}$ and $h_a : D \to \mathbb{R}$ such that $V_a(m) = \int (w_a(c) + h_a(\hat{m}))\, dm(c, \hat{m})$ represents $\succsim_a$. Moreover, a stronger version of Lemma 7 holds if dynamic consistency (axiom 6) is imposed:

Lemma 15. *For all $\hat{m} \in D$, $\partial h_a(\hat{m}) = \{x^*_{\hat{m}}\} = \{V^*(\hat{m}) + V_{\hat{m}}\}$ where V^* is the conjugate function defined by $V^*(m) \equiv \sup_{m \in D}\{x^*_{\hat{m}}(m) - V_{\hat{m}}(m)\}$. Moreover, $h_m = h_{\hat{m}}$ and $V_m = V_{\hat{m}}$ for $\hat{m} = m^1 \times m^2 \times m^3$.*

Proof. The first part follows from the proof of Lemma 7. The second part is a straightforward implication of axioms 4, 6 and 9.

□

Lemma 16. *For all $a \in D$, there are continuous and Gateaux differentiable utility functions $v_a : C \to \mathbb{R}$ and $h : D \to \mathbb{R}$, and a scalar $\beta > 0$ such that $\succsim_a$ can be represented by*

$$V_a(m) = \int (v_a(c) + \beta h(\hat{m}))\, dm(c, \hat{m}).$$

Proof. First, by intermediate value theorem for Gateaux differentiable and continuous func-

tions (see, e.g., Cerreia-Vioglio et al. (2017)), if two continuous and Gateaux differentiable functions have the same Gateaux derivatives then they are affine transformations of each other. Thus, it is possible for all $a \in D$ to renormalize w_a such that $v_a = \frac{w_a - \sigma_a}{\gamma_a}$ and $\beta h = \frac{h_a}{\gamma_a}$ for $\beta > 0$.

□

Lemma 17. *h is a convex function and can be written as $h = v^2 + v^3$ where $v^3 : \Delta(\Delta(D))$ is linear in probabilities, i.e. $v^3(\nu) = \int \beta h(\mu) d(\nu)(\mu)$ for $\nu \in \Delta(\Delta(D))$ and $\mu \in \Delta(D)$. Moreover, $V_a = V_{\hat{a}}$ for all $\hat{a} = a^1 \times \hat{a}^1 \times \hat{a}^3$ for any $\hat{a}^2 \in \Delta\Delta(C)$ and $\hat{a}^3 \in \Delta(\Delta(D))$.*

Proof. The second part follows from the second part of the strong separability assumption (apply Lemma 4 again) and the fact that h is independent of a. The latter together with the unqiuness of additively separable representations implies that, since V_a satisfies independence for each $a \in D$, v^3 has to do this too. Moreover, note that this also implies that $V_a = V_{\hat{a}}$ for all $\hat{a} = a^1 \times \hat{a}^1 \times \hat{a}^3$ for any $\hat{a}^2 \in \Delta\Delta(C)$ and $\hat{a}^3 \in \Delta(\Delta(D))$ by the same reasoning.

To establish that h is convex, first note that v^3 is convex as it is linear in probabilities. Thus, h is convex if and only if v^2 is a convex function. Assume to get a contradiction that there are two lotteries m^1 and $\hat{m}^1$ such that

$$(1-\alpha)v_a^2(m^1) + \alpha v_a^2(\hat{m}^1) < v_a^2((1-\alpha)m^1 + \alpha\hat{m}^1).$$

Since V_a is linear in probabilities, it is possible to find $m^2 \times m^3$ and $\hat{m}^2 \times \hat{m}^3$ such that $V_a(a) < V_a(\hat{m})$ where $a = ((1-\alpha)m^1 + \alpha\hat{m}^1) \times m^2 \times m^3$ and $\hat{m} = \hat{m}^1 \times \hat{m}^2 \times \hat{m}^3$ with

$$\int \beta h(x) d(m^2 \times m^3)(x) + (1-\alpha)(v_a(m^1) - v_a(\hat{m}^1)) = \int \beta h(x) d(\hat{m}^2 \times \hat{m}^2)(x) - \varepsilon$$

for $\varepsilon > 0$. Such a lottery $\hat{m} \in D$ exists for any $\alpha \in (0,1)$ by continuity of the preferences if m^1 and $\hat{m}^1$ are close enough. In fact, it is possible to find a bound $\delta > 0$ such that this is always the case for any lotteries $m', \hat{m}' \in D$ with $d_D(m', \hat{m}') < \delta$. By continuity, as ε tends to zero the resulting $\hat{m}$ is such that $h(a) > h(\hat{m})$, violating dynamic consistency as $(c, a) \succsim_{\bar{a}} (c, \hat{m})$ but $\hat{m} \succ_a a$. Since v^2 is convex on any neighborhood, it is sufficient to ensure

that it is also convex on the entire D.

□

The last lemma concludes the proof of Theorem 4.

3.9.3 Appendix C: Remaining Proofs

Proof of Proposition 11: For any complete metric space $(\mathcal{V}, d)$, Banach Fixed Point Theorem implies that any contraction mapping, $T : \mathcal{V} \to \mathcal{V}$, has a unique fixed point V^*. The set of Lipschitz continuous real-valued functions, $BL(D \times C)$, on the compact metric space $D \times C$ is complete when endowed with the metric $d_{BL}(f, g) = ||f - g||_\infty + ||f - g||_d$ for all $f, g \in BL(D \times C)$, where

$$||f||_d = \sup\{||f(s) - f(t)||(d_{D\times C}(s, t))^{-1} : s, t \in D \times C, s \neq t\}$$

and $|| \cdot ||_\infty$ is the sup-norm (see, e.g., Bogachev (2007)).

First, note that any $\succsim$ with an AD representation (u, ϕ, β, r) is such that for all $a \in D$, $\succsim_a$ can be represented by

$$\hat{V}_{\hat{r}(a^1)}(m) = \int \left[\phi(u(c) - u(\hat{r}(a^1))) + u(\hat{r}(\hat{m}^1)) + \beta V_{\hat{r}(\hat{m}^1)}(\hat{m})\right] dm(c, \hat{m}) \quad \forall m \in D, \quad (3.29)$$

where $\hat{r} : \Delta(C) \to C$ is such that $r(a^1) = u(\hat{r}(a^1))$ for all $a^1 \in \Delta(C)$.

The operator $T : BL(D \times C) \to BL(D \times C)$ where

$$T\hat{V}_{\hat{c}}(m) = \int \left[\phi(u(c) - u(\hat{r}(\hat{c}))) + u(\hat{r}(\hat{m}^1)) + \beta V_{\hat{r}(\hat{m}^1)}(\hat{m})\right] dm(c, \hat{m})$$

is well-defined and a contraction mapping using Blackwell's sufficient conditions. That is, (i) $V'_c(m) \leq V_c(m)$ for all $(m, a) \in D \times C$ implies $T\hat{V}'_c(m) \leq T\hat{V}_c(m)$ for all $(m, c) \in D \times C$, and (ii) $T(\hat{V}_c(m) + x) \leq T\hat{V}_c(m) + \beta x$ for all $\hat{V} \in BL(D \times C)$, $x \in \mathbb{R}_+$ and $(m, c) \in D \times C$. Thus, $T\hat{V}_{(\cdot)} = \hat{V}_{(\cdot)}$ is unique and satisfies equation (3.29). This implies that, for any $a \in D$, u, ϕ, β, and r, there exists a unique Lipschitz continuous function V_a satisfying equation (3.4).

Q.E.D

Proof of Theorem 3: I start with the necessity part. The proof is an adaption of the proof of Theorem 3 in Sarver (2018, p.1376-1379). First, I show that that for all $\boldsymbol{c} = (c_1, c_2, c_3, \ldots) \in C^{\mathbb{N}}$, $V_a^2(\boldsymbol{c}) = \alpha V_a^1(\boldsymbol{c}) + \lambda$. Since $\succsim_1$ is more risk averse than $\succsim_2$, $V_a^1(\boldsymbol{c}) \geq V_a^1(\boldsymbol{c'})$ implies $V_a^2(\boldsymbol{c}) \geq V_a^2(\boldsymbol{c'})$ for any $\boldsymbol{c}, \boldsymbol{c'} \in D$.

Lemma 18. *If $V_{\hat{a}}^1(\boldsymbol{c}) \geq V_{\hat{a}}^1(\boldsymbol{c'})$ implies $V_{\hat{a}}^2(\boldsymbol{c}) \geq V_{\hat{a}}^2(\boldsymbol{c'})$ for any $\boldsymbol{c} = (c, c, c, \ldots), \boldsymbol{c'} = (c', c', c', \ldots) \in D$, then there exist $\lambda \in \mathbb{R}$, $\alpha > 0$ such that $\beta_1 = \beta_2$, $u_2 = \alpha u_1 + \lambda$ and $V_{\hat{a}}^2(\boldsymbol{c}) = \alpha V_{\hat{a}}^1(\boldsymbol{c}) + \lambda(1-\beta)^{-1}$.*

Proof. It is easy to see that $u_1(c) \geq u_1(c')$ implies $u_2(c) \geq u_2(c')$ since $V_{\hat{a}}^1(\boldsymbol{c}) \geq V_{\hat{a}}^1(\boldsymbol{c'})$ implies $V_{\hat{a}}^2(\boldsymbol{c}) \geq V_{\hat{a}}^2(\boldsymbol{c'})$. The next order of business is showing that $u_1(c) > u_1(c')$ implies $u_2(c) > u_2(c')$. By nondegeneracy of $\succsim^1$ and $\succsim^2$, there exists $c^*, c_* \in C$ such that $u_2(c^*) > u_2(c_*)$. For any $T \in \mathbb{N}$, take two consumption streams $\boldsymbol{c}_T = (c_1, c_2, c_3, \ldots)$ and $\boldsymbol{c}'_T = (c'_1, c'_2, c'_3, \ldots)$ where $c_1 = c'_1$, $c_t = c$ and $c'_t = c'$ for all $1 < t < T$, $c_T = c_*$ and $c'_T = c^*$, and $c_\tau = c'_\tau$ for all $\tau > T$. Fix T such that $V_{\hat{a}}^1(\boldsymbol{c}_T) > V_{\hat{a}}^1(\boldsymbol{c}'_T)$, which exists since the latter is equivalent to

$$\beta_1(1-\beta_1^{T-2})u_1(c) + \beta_1^T u_1(c_*) > \beta_1(1-\beta_1^{T-2})u_1(c') + \beta_1^T u_1(c^*).$$

By the hypothesis, $V_{\hat{a}}^2(\boldsymbol{c}_T) \geq V_{\hat{a}}^2(\boldsymbol{c}'_T)$ which implies that $u_2(c) > u_2(c')$ as

$$\beta_2(1-\beta_2^{T-2})u_2(c) + \beta_2^T u_2(c_*) \geq \beta_2(1-\beta_2^{T-2})u_2(c') + \beta_1^T u_2(c^*).$$

and $u_2(c^*) > u_2(c_*)$. Conclude that u_1 and u_2 are ordinally equivalent.

What is left to show is that $V_{\hat{a}}^1(\boldsymbol{c}) > V_{\hat{a}}^1(\boldsymbol{c'})$ implies $V_{\hat{a}}^2(\boldsymbol{c}) > V_{\hat{a}}^2(\boldsymbol{c'})$ for any $\boldsymbol{c} = (c_1, c_2, c_3, \ldots), \boldsymbol{c'} = (c'_1, c'_2, c'_3, \ldots) \in D$. Assume on the way to a contradiction that $V_{\hat{a}}^2(\boldsymbol{c}) = V_{\hat{a}}^2(\boldsymbol{c'})$ instead. Since u_1 is continuous and C is connected, there exists an $\boldsymbol{c''} = (c''_1, c''_2, c''_3, \ldots) \in D$ where c''_t is such that $u_1(c_t) > u_1(c''_t) > u_1(c'_t)$ and $c''_\tau = c'_\tau$ for all $\tau \neq t$. This implies that $V_{\hat{a}}^1(\boldsymbol{c}) > V_{\hat{a}}^1(\boldsymbol{c''}) > V_{\hat{a}}^1(\boldsymbol{c'})$ but, since $u_2(c''_t) > u_2(c'_t)$, $V_{\hat{a}}^2(\boldsymbol{c}) = V_{\hat{a}}^2(\boldsymbol{c'}) > V_{\hat{a}}^2(\boldsymbol{c''})$, a contradiction. Thus, also $V_{\hat{a}}^1(\boldsymbol{c})$ and $V_{\hat{a}}^2(\boldsymbol{c})$ are ordinally equivalent.

By the uniqueness up to an affine transformation of additive separable utility functions (see, e.g., Debreu (1960)), it must be the case that $\beta_1 = \beta_2$, $u_2 = \alpha u_1 + \lambda$ and $V_{\hat{a}}^2(\boldsymbol{c}) =$

$\alpha V^1_{\hat{a}}(\boldsymbol{c}) + \lambda(1-\beta)^{-1}$.

□

By the separability axiom, it suffices to show that

$$\max_{r\in\mathbb{R}}\{r+\beta\int\phi_1(u_1(c)-r)dm^1(c)\}\geq u_1(c)\Longrightarrow \max_{r\in\mathbb{R}}\{r+\beta\int\phi_2(u_2(c)-r)dm^1(c)\}\geq u_2(c)$$

for any $m\in D$ to establish that

$$(c_1,m)\succsim^1_{\hat{a}}(c_1,c_2,c_3,\ldots)\Longrightarrow(c_1,m)\succsim^2_{\hat{a}}(c_1,c_2,c_3,\ldots).$$

It is shown in the proof of Theorem 3 in Sarver (2018) (easily adapted to my setting) that the continuity axiom implies that for any $m\in D$ there exists an $c^1_{m^1}\in C$ such that $u_1(c^1_{m^1})=\max_{r\in\mathbb{R}}\{r+\beta\int\phi_1(u_1(c)-r)dm^1(c)\}$ for any AD representation (u,ϕ,β). Therefore, since $\succsim^1_{\hat{a}}$ is more risk averse than $\succsim^2_{\hat{a}}$,

$$\begin{aligned}\max_{r\in\mathbb{R}}&\{r+\beta\int\phi_2(u_2(c)-r)dm^1(c)\}\\&\geq u_2(c^1_{m^1})\\&=\alpha u_1(c^1_{m^1})+\lambda\\&=\max_{r\in\mathbb{R}}\{r+\beta\int\alpha\phi_1(u_1(c)-r)dm^1(c')+\lambda\}.\end{aligned}$$

Clearly, this holds if $\phi_2(x)\geq\alpha\phi_1(\alpha^{-1}x)$ for all $x\in[-a,a]$. To prove the converse, assume to get a contradiction that $\phi_2(y)<\alpha\phi_1(\alpha^{-1}y)$ for all y in some nonempty interval $[\underline{y},\bar{y}]\subset[-a,a]$. Then, there exists some $m^1=\gamma c+(1-\gamma)c'\in\Delta(C)$ for $\gamma>0$ close to 1 and $u(c)-u(c')\in(\underline{y},\bar{y})$ such that $\max_{r\in\mathbb{R}}\{r+\beta\int\phi_2(u_2(c)-r)dm^1(c)\}<\max_{r\in\mathbb{R}}\{r+\beta\int\alpha\phi_1(u_1(c)-r)dm^1(c')+\lambda\}$ since $x<\beta\phi_2(x)$ for all $x\neq0$.

Q.E.D.

Proof of Proposition 13: For part 1., if $m\succsim_{\hat{m}}\hat{m}$ then $\neg[\hat{m}\succ_{\hat{m}}m]$ which implies, by completeness and strict dynamic consistency, that $\neg[(c,\hat{m})\succ_a(c,m)]$ for any $a\in D$. Thus, $(c,m)\succsim_a(c,\hat{m})$ for all $a\in D$ implying that $m\succ_{\hat{m}}\hat{m}$.

For part 2, if $\beta > \frac{1}{2}$ then for any $m^1 \in \Delta(C)$ there exists a temporal lottery $m \in D$ given by $m = m^1 \times \delta_{m^1} \times \delta_{\delta_{m^1}} \ldots$ such that

$$\beta V_m(m) = \frac{\beta U^*}{1-\beta} > U^*$$

where $U^* = r(m^1) + \beta \int \phi(u(c) - r(m^1))dm^1(c)$. Thus, no matter how much preferable m^1 is than $\hat{m}^1$, by continuity it is possible to find $m^2 \times m^3, \hat{m}^2 \times \hat{m^3} \in \Delta\Delta(C \times D)$ such that $(c, m) \sim_a (c, \hat{m})$ for $m = m^1 \times m^2 \times m^3$ and $\hat{m} = \hat{m}^1 \times \hat{m}^2 \times \hat{m}^3$ (independent of $a \in D$).

Now assume that

$$\arg\max_{r \in \mathbb{R}} \{r + \int \beta\phi(u(c) - r)dm^1(c)\} \quad \forall\, m^1 \in \Delta(C).$$

contains both $r(m^1)$ and $r(\hat{m}^1) \neq r(m^1)$, and $(c, m) \sim_a (c, \hat{m})$. Then, since $r(m^1)$ is also optimal against $\hat{m}$, it must be the case that $m \sim_m \hat{m}$ violating strict dynamic consistency.

Conversely, assume that $(c, m) \succsim_a (c, \hat{m})$ but $m \sim_m \hat{m}$ with $r(m^1) \neq r(\hat{m}^1)$. This implies that

$$r(m^1) + \beta \int \phi(u(c) - r(m^1))dm^1(c) = r(m^1) + \beta \int \phi(u(c) - r(m^1))d\hat{m}^1(c),$$

hence $r(m^1) \in \arg\max_{r \in \mathbb{R}} \{r + \int \beta\phi(u(c) - r)d\hat{m}^1(c)\}$ as $(c, m) \succsim_a (c, \hat{m})$.

Q.E.D.

Proof of Theorem 5: It follows from the proof of Theorem 3 in Appendix 3.9.3 that $\succsim^1$ exhibits a stronger status quo bias than $\succsim^2$ if and only if $\beta_1 = \beta_2$ and there exist scalars $\sigma \in \mathbb{R}$, $\alpha > 0$ such that $u_2 = \alpha u_1 + \sigma$ and $r_2 = \alpha r_1 + \sigma$.

Assume to get a contradiction that $\phi_2(\alpha x^*) < \alpha\phi_1(x^*)$ for some $x^* \geq -b$ where $b > 0$ is such that $\phi_1(-b) = -a/(1-\beta)$ for $a = \max u_1 - \min u_1$. By the definition of a and connectedness of C, it must then be the case that

$$\phi_1(u_1(c) - r) + \frac{u_1(c^*)}{1-\beta} \geq \phi_1(0) + \frac{u_1(c')}{1-\beta} \tag{3.30}$$

and

$$\phi_2(u_2(c) - \alpha r - \sigma) + \frac{u_2(c^*)}{1-\beta} < \phi_2(0) + \frac{u_2(c')}{1-\beta} \tag{3.31}$$

for some $(c, c^*, c^*, c^*, \ldots), (\hat{c}, c', c', c', \ldots) \in C^{\mathbb{N}}$ and $u_1(c) - r = x^*$. This contradicts the fact that $\succsim^1$ exhibits a stronger status quo bias than $\succsim^2$. Conclude that the stated condition is necessary for $\succsim^1$ to exhibit a stronger status quo biased than $\succsim^2$.

Conversely, assume that $u_1(c) - r < -b$, then

$$\phi_1(u_1(\hat{c}) - u_1(c)) + h(m) < \phi_1(0) + h(\hat{m})$$

for all $m, \hat{m} \in D$ so in particular $(c, c', \ldots) = \boldsymbol{c} \succsim^1_{\boldsymbol{c}} (\hat{c}, \bar{c}, \ldots) = \hat{\boldsymbol{c}}$. Second, for any other $u_1(c) - u_1(\hat{c}) \in [-b, a]$, it must be the case that

$$\phi_2(0)+h_2(c', \ldots) \geq \phi_2(u_2(\hat{c})-u_2(c))+h_2(\bar{c}, \ldots) \Longrightarrow \phi_1(0)+h_1(c', \ldots) \geq \phi_1(u_1(\hat{c})-u_1(c))+h_1(\bar{c}, \ldots)$$

$$\boldsymbol{c} \succsim^2_{\boldsymbol{c}} \hat{\boldsymbol{c}} \Longrightarrow \boldsymbol{c} \succsim^1_{\boldsymbol{c}} \hat{\boldsymbol{c}}.$$

Therefore, $\succsim^1$ exhibits a stronger status quo bias than $\succsim^2$.

Q.E.D.

Proof of Proposition 14: Corollary S.1 (Sarver, 2018, p.2, Supplementary Appendix) implies that if $\mathcal{W} : \Delta([0,1]) \to \mathbb{R}$ is continuous in the topology of weak convergence and convex. Then the following are equivalent: (i) $\mathcal{W}$ is monotone with respect to FOSD (SOSD) if and only if it satisfies

$$\mathcal{W}(\mu) = \max_{\rho \in \Phi} \int \rho(c) d\mu(c)$$

for some collection of nondecreasing (and concave) continuous functions $\rho : [0,1] \to \mathbb{R}$. Clearly, for a monotonic AD representation, (u, ϕ, r, β), $\phi(u - r)$ is always increasing for all $r \in [\min u, \max u]$. The SOSD requirement then boils down to $\phi(u - r)$ being concave for all $r \in [\min u, \max u]$.

Q.E.D.

Proof of Proposition 19: The risk premium is given implicitly by

$$u(c-\pi(\delta_c+\varepsilon\mu))=r_\varepsilon+\int\phi(u(c+\varepsilon c')-r_\varepsilon)\mu(c'). \tag{3.32}$$

The optimal reference point is a function of ε as the solution to

$$\lim_{r\to r_\varepsilon^-}\int\phi'(u(c+\varepsilon c')-r)\mu(c')\geq 1\geq\lim_{r\to r_\varepsilon^+}\int\phi'(u(c+\varepsilon c')-r)\mu(c'). \tag{3.33}$$

If both sides holds with equality any small impact of ε is of second order and, thus, has an negligible effect on r_ε. If not, it is optimal to set the reference point to $r_\varepsilon=u(c+\varepsilon c^*)$ for some c^* with $\mu(c^*)>0$. By differentiating equation (3.32) then we get

$$\begin{aligned}\frac{\partial\pi(\delta_c+\varepsilon\mu)}{\partial\varepsilon}=-(u'(c-\pi(\delta_c+\varepsilon\mu))^{-1}\Bigg[&c^*u'(c+\varepsilon c^*)\left[1-\int\phi'(u(c+\varepsilon c')-r)\mu(c')\right]\\&+\int\phi'(u(c+\varepsilon c')-r)u'(c+\varepsilon c')c'\mu(c')\Bigg]\end{aligned} \tag{3.34}$$

Let ε tend to zero from above. Then, if equation (3.33) holds with equalities, the first two terms are zero and $\frac{\partial\pi(\delta_c+\varepsilon\mu)}{\partial\varepsilon}|_{\varepsilon=0^+}=[\lambda-1]\eta\int_{c'>0}c'\mu(c')$. If not, we have

$$\begin{aligned}\frac{\partial\pi(\delta_c+\varepsilon\mu)}{\partial\varepsilon}|_{\varepsilon=0^+}=-c^*+c^*\eta\int_{c'>c^*}\mu(c')+c^*\lambda\int_{c'\leq c^*}\mu(c')\\-\eta\int_{c'>c^*}c'\mu(c')-\lambda\eta\int_{c'\leq c^*}c'\mu(c').\end{aligned} \tag{3.35}$$

There are two cases, either $c^*>\hat{c}$ or $c^*<\hat{c}$. I will start with the former case. The RHS of equation (3.35) can be reordered to give

$$\frac{\partial\pi(\delta_c+\varepsilon\mu)}{\partial\varepsilon}|_{\varepsilon=0^+}=c^*(\lambda\eta-1)-\hat{c}^*(\lambda-1)\eta\int_{c'>\hat{c}^*}\mu(c')+(\lambda-1)\eta\int_{c'>\hat{c}^*}\hat{c}\mu(c')>\hat{c}$$

Finally, when $\hat{c}^*<0$ the RHS of equation (3.35) can be reordered to give

$$\frac{\partial\pi(\delta_c+\varepsilon\mu)}{\partial\varepsilon}|_{\varepsilon=0^+}=-\hat{c}^*(1-\eta)-\hat{c}^*(\lambda-1)\eta\int_{c'\leq\hat{c}^*}\mu(c')+(1-\lambda)\eta\int_{c'\leq\hat{c}^*}\hat{c}\mu(c')>\hat{c}.$$

Q.E.D.

Proof of Proposition 21: I now derive the equity premium. The market portfolio is defined as the claim to aggregate consumption. This gives the following expression of the price-dividend ratio

$$p_t = c_t \sum_{\tau=t+1}^{\infty} \mathbb{E}_t \left[\beta \left(\frac{c_\tau}{c_t} \right)^{1-\rho} \frac{\xi_\tau}{\xi_t} \right] = c_t \sum_{\tau=t+1}^{\infty} \mathbb{E}_t \left[\prod_{i=t}^{\tau-1} \beta \left(\frac{c_{i+1}}{c_i} \right)^{1-\rho} \frac{\xi_{i+1}}{\xi_i} \right]$$

where ξ_t equals $\lambda\eta$ if $u(c_t) \leq r_t$ and η otherwise.

Since consumption growth is i.i.d. but ξ_t depends on c_t/c_{t-1}, this can be rewritten as

$$\frac{p_t}{c_t} = \frac{1}{\xi_t} \mathbb{E}_t \left[\sum_{\tau=t+1}^{\infty} \xi_\tau \prod_{i=t}^{\tau-1} \beta \left(\frac{c_{i+1}}{c_i} \right)^{1-\rho} \right] = \frac{1}{\xi_t} \frac{v}{1-v}$$

where $v = \beta \mathbb{E} \left[\left(\frac{c_{u+1}}{c_u} \right)^{1-\rho} \right] < 1$. The return of the market portfolio is given by

$$R^m_{t+1} = \frac{c_{t+1}}{c_t} \left(\frac{1 + p_{t+1}/c_{t+1}}{p_t/c_t} \right) = \left(\frac{\xi_t}{\xi_{t+1}} + \xi_t \frac{1-v}{v} \right) \cdot \frac{c_{t+1}}{c_t}.$$

which implies

$$\mathbb{E}[R^m] = \mathbb{E} \left[\left(\frac{1}{\xi_{t+1}} + \frac{1-v}{v} \right) \cdot \frac{c_{t+1}}{c_t} \right].$$

To avoid problems with the kink in ϕ, approximate ϕ by

$$\phi_\alpha(x) = \begin{cases} \lambda\eta x & \text{for} \quad x < -\alpha, \\ (1 - (1-\eta)(x/2\alpha))\, x & \text{for} \quad x \in [-\alpha, \alpha], \\ \eta x & \text{for} \quad x > \alpha. \end{cases} \tag{3.36}$$

with α tending to zero.

Calculations for the Sharpe ratio and the variance of the risk-free rate:

$$\mathbb{E}_t[M^{r_t}_{t+1}] = \beta \frac{\exp\left(-\rho\mu_c + \frac{\rho^2}{2}\sigma_c^2 \right)}{\eta + (\lambda\eta - \eta)\mathbb{1}_{\{u(c_t) \leq r_t\}}}$$

$$\mathbb{E}_t\left[\left(M_{t+1}^{r_t}\right)^2\right] = \beta^2 \frac{\eta^2 + 2\eta(\lambda\eta - \eta)Pr(u(c_{t+1}) \leq r_{t+1}) + (\lambda\eta - \eta)^2 Pr(u(c_{t+1}) \leq r_{t+1})}{\left(\eta + (\lambda\eta - \eta)\mathbb{1}_{\{u(c_t) \leq r_t\}}\right)^2} \exp\left(-2\rho\mu_c + 2\rho^2\sigma_c^2\right)$$
$$= \beta^2 \frac{\eta + \lambda\eta - \lambda\eta^2}{\left(\eta + (\lambda\eta - \eta)\mathbb{1}_{(u(c_t) \leq r_t)}\right)^2} \exp\left(-2\rho\mu_c + 2\rho^2\sigma_c^2\right)$$

$$\sigma(R_{t+1}^*) = \frac{\sqrt{(\eta + \lambda\eta - \lambda\eta^2)\exp\left(\rho^2\sigma_c^2\right) - 1}}{\beta\exp(-\rho\mu_c + \rho^2\sigma_c^2)} \geq \frac{\mathbb{E}_t[R_{t+1}^m - R_{t+1}^*]}{\sigma_t(R_{t+1}^m)} \exp\left(-\rho^2\sigma_c^2/2\right) R^*.$$

The proof is completed by noting that $\lambda\eta = 1 + \kappa$ and $\eta = 1 - \kappa$.

Q.E.D.

Proof of Proposition 22:

If a solution exists, it must satisfy the following Euler inequality

$$\eta e^{-\theta c_t} \leq (1 + r)\beta\mathbb{E}\left[(\eta + (\lambda - 1)\eta\mathbb{1}_{\{u(c_{t+1}) < r_{t+1}\}})e^{-\theta c_{t+1}}\right] \leq \lambda\eta e^{-\theta c_t} \tag{3.37}$$

where r_{t+1} is the median of the induced consumption process. I will guess and verify the consumption process compatible with the above inequalities. Clearly, given that the above inequalities hold with equality, the process must be linear in levels. This gives the following process:

$$c_{t+1} = \phi_t c_t + \Gamma_t(S_t) + v_{t+1} \tag{3.38}$$

where $\Gamma_t : \mathbb{R} \to \mathbb{R}$, ϕ_t and v_{t+1} are to be determined. Assume that the Euler inequality holds with equality, then clearly ϕ_t has to equal 1 otherwise would be determined by this equation regardless of the budget constraint. Plugging in equation (3.38) in (3.37) gives:

$$\eta \leq \exp[-\theta\Gamma_t(S_t)]\mathbb{E}\left[(\eta + (\lambda - 1)\eta\mathbb{1}_{\{u(c_{t+1}) < r_{t+1}\}})e^{-\theta v_{t+1}}\right] \leq \lambda\eta.$$

Note that when the Euler equation does not hold with equality, it must be the case that any

income shock will be pushed towards the future. Thus, Γ_t is a function of S_t given by

$$\Gamma_t(S_t) = \begin{cases} \dfrac{\mathbb{E}\left[(\eta + (\lambda-1)\eta \mathbb{1}_{\{u(c_{t+1})<r_{t+1}\}})e^{-\theta v_{t+1}}\right] - \log(\lambda\eta) + \rho}{\theta} & \text{for} \quad S_t < \underline{S}, \\ \dfrac{\mathbb{E}\left[(\eta + (\lambda-1)\eta \mathbb{1}_{\{u(c_{t+1})<r_{t+1}\}})e^{-\theta v_{t+1}}\right] + \rho}{\theta} + (1+r)S_t & \text{for} \quad S_t \in [\underline{S}, \bar{S}], \\ \dfrac{\mathbb{E}\left[(\eta + (\lambda-1)\eta \mathbb{1}_{\{u(c_{t+1})<r_{t+1}\}})e^{-\theta v_{t+1}}\right] - \log(\eta) + \rho}{\theta} & \text{for} \quad S_t > \bar{S}. \end{cases} \tag{3.39}$$

where $\rho = \log(1+r) + \log(\beta)$, $\underline{S} = -\dfrac{\log(\lambda\eta)}{(1+r)\theta}$, and $\bar{S} = -\dfrac{\log(\eta)}{(1+r)\theta}$.

Notice that $\mathbb{E}_t[y_{t+1}] - \mathbb{E}_{t-1}[y_{t+1}] = S_t$ and rewrite the intertemporal budget constraint as follows (using $R = (1+r)^{-1}$):

$$\sum_{i=0}^{\infty} R^i(c_{t+i} - y_{t+i}) = A_t = \sum_{i=0}^{\infty} R^i c_{t+i} + \sum_{i=0}^{\infty} R^i(y_{t+i} - \mathbb{E}_t[y_{t+i}]) - \sum_{i=0}^{\infty} R^i \mathbb{E}_t[y_{t+i}].$$

This can be written as

$$\sum_{i=0}^{\infty} R^i c_t + \sum_{i=1}^{\infty} R^i \sum_{j=1}^{i} \Gamma_{t+j-1}(S_{t+j-1}) - \sum_{i=0}^{\infty} R^i \mathbb{E}_t[y_{t+i}] + \sum_{i=1}^{\infty} R^i \sum_{j=1}^{i} v_{t+j} - \sum_{i=1}^{\infty} R^i \sum_{j=1}^{i} S_{t+j} = A_t. \tag{3.40}$$

Taking expectations condition on information in period t gives:

$$c_t = y_t^p - (1-R)\sum_{i=1}^{\infty} R^i \sum_{j=1}^{i} \mathbb{E}_t[\Gamma_{t+j-1}(S_{t+j-1})] \tag{3.41}$$

where the first term is permanent income $y_t^p := (1-R)\left(A_t + \sum_{i=0}^{\infty} R^i \mathbb{E}_t[y_{t+i}]\right)$.

Plugging back equation (3.41) in (3.40) gives

$$\sum_{i=1}^{\infty} R^i \sum_{j=1}^{i} \left(\Gamma_{t+j-1}(S_{t+j-1}) - \mathbb{E}_t[\Gamma_{t+j-1}(S_{t+j-1})]\right) + \sum_{i=1}^{\infty} R^i \sum_{j=1}^{i} (v_{t+j} - S_{t+j}) = 0.$$

Since this equation has to hold for all t, the following condition has to hold

$$\begin{aligned} 0 =& \sum_{i=1}^{\infty} R^i \left[\Gamma_{\tau-1}(S_{\tau-1}) - \mathbb{E}_{\tau-j-1}[\Gamma_{\tau-1}(S_{\tau-1})] + v_\tau - S_\tau\right] \iff \\ v_\tau =& L(S_\tau) - \gamma \log(\lambda\eta)/\theta - \alpha \log(\eta)/\theta + (1-\gamma-\alpha)\mathbb{E}[S_\tau | S_\tau \in [\underline{S}, \bar{S}]] \end{aligned}$$

where $\gamma = Pr\left(S_\tau > -\frac{\log(\lambda\eta)}{\theta(1+r)}\right)$, $\alpha = Pr\left(S_\tau > -\frac{\log(\eta)}{\theta(1+r)}\right)$, and

$$L(S_\tau) = \begin{cases} \frac{\log(\lambda\eta)}{\theta} + S_\tau & \text{for} \quad S_\tau < \underline{S}, \\ 0 & \text{for} \quad S_\tau \in [\underline{S}, \bar{S}], \\ \frac{\log(\eta)}{\theta} + S_\tau & \text{for} \quad S_\tau > \bar{S}. \end{cases} \tag{3.42}$$

Again, the proof is completed by noting that $\lambda\eta = 1 + \kappa$ and $\eta = 1 - \kappa$.

Q.E.D.

3.9.4 Appendix D: Additional Results

Following Kőszegi and Rabin (2007) I formalize another property related to the endowment effect in the ex ante/ex post setting and show that it is satisfied by any AD preferences (u, ϕ, r, β) if u is linear, ϕ is piecewise linear and $C = [0, 1]$. Proposition 24 below states that the decision-maker, in the ex-post stage, is no more willing to accept a lottery $\mu \in \Delta(C)$ on top of some wealth level c when her reference point is c, then she is to accept μ when she is already facing a lottery ν given any reference point $\hat{c}$.

Letting $\hat{c}$ be the optimal reference point when facing μ with $\nu = \delta_c$, it is evident that the proposition implies an endowment effect for risk. Remember, ex post utility given the reference point $\hat{c}$ is $U_{\hat{c}}(\mu) = \int \phi(u(c) - u(\hat{c}))d\mu(c)$ and for any $\mu, \nu \in \Delta(C)$ let $\mu + \nu$ denote the convolution (that is, $(\mu + \nu)(c) = \int \mu(c - \hat{c})d\nu(\hat{c})$) of μ and ν assumed to live in $\Delta(C)$.

Proposition 24. *Let $\succsim$ have an AD representation (u, ϕ, β) where ϕ is piecewise linear and u is linear. For any lotteries $\mu, \nu \in \Delta(C)$ and any $c, \hat{c} \in C$, if $U_c(\mu) \geq U_c(\delta_c)$, then $U_{\hat{c}}(\mu + \nu) \geq U_{\hat{c}}(\nu)$.*

Proof of Proposition 24: The decision-maker prefers $\hat{c} + \mu$ over μ when the reference point is $\hat{c}$ if and only if

$$\int \phi(c)d\mu(c) \geq 0.$$

She prefers $\mu+\nu$ over ν when the reference point is z if and only if

$$\int\int \phi(c+c'-z)d\mu(c)d\nu(c') \geq \int \phi(c'-z)d\nu(c').$$

or equivalently as

$$\int\int (\phi(c+c'-z)-\phi(c'-z))d\mu(c)d\nu(c') \geq 0 \tag{3.43}$$

As observed bc Kőszegi and Rabin (2007, Proposition 1.) for $c \geq 0$, $\phi(c+c'-z)-\phi(c'-z) \geq \eta c$, and for $c \leq 0$, $\phi(c+c'-z)-\phi(c'-z) \geq \lambda\eta c$. Thus we can rewrite (3.43) as

$$\begin{aligned}\int\int (\phi(c+c'-z)-\phi(c'-z))d\mu(c)d\nu(c') \\ \geq \int\int \left(\eta \max\{c,0\} + \lambda\eta \min\{c,0\}\right) d\mu(c)d\nu(c') = \int \phi(c)d\mu(c) \geq 0.\end{aligned}$$

Q.E.D.

Proof of Proposition 20: This results is a slight generalization of a result by Sarver (2018)[Supplementary Appendix]. It is here provided for completeness.

For simplicity, I provide the result for the two-stage model and, in addition, abstracting from ex post choices. It is straightforward to then extend it to the general temporal lottery model.

When ϕ is piecewise linear, the indirect utility function given a lottery $\mu \in \Delta(C)$ is obtained by maximizing

$$V(\mu) = r + \int_{u(x)\leq r} \lambda\eta[u(x)-r]dF_\mu(x) + \int_{u(x)>r} \eta[u(x)-u(r)]dF_\mu(x)$$

with respect to r (where F_μ is the cumulative distribution function for μ). Differentiating with respect to r gives the first order condition

$$\begin{aligned}1 &= \int_{u(x)\leq r} \lambda\eta dF_\mu(x) + \int_{u(x)>r} \eta dF_\mu(x) \\ \Leftrightarrow \frac{1-\eta}{(\lambda-1)\eta} &= \int_{u(x)\leq r} dF_\mu(x) = F_\mu(r).\end{aligned}$$

$u(x)$ is strictly increasing, and so is its inverse, therefore the objective function is first

increasing in r and then decreasing implying that the FOC equal to 0 is necessary and sufficient for the optimum.

By the above, the optimal reference point r^μ given a lottery μ is such that $\lim_{u(x)\to r^{\mu-}} F_\mu(x) \leq \frac{1-\eta}{(\lambda-1)\eta} \leq F_\mu(r^\mu)$. To show that the AD representation is a special case of RDU notice that, given an optimal reference point r, $V(\mu)$ is equal to

$$
\begin{aligned}
&r^\mu + \int_{u(x)\leq r^\mu} \lambda\eta(u(x)-r^\mu)dF_\mu(x) + \int_{u(x)>r^\mu} \eta(u(x)-r^\mu)dF_\mu(x)\\
=&r^\mu\left[1-(\lambda-1)\eta F_\mu(r^\mu)-\eta\right] + \int_{u(x)\leq r^\mu} u(x)d\lambda\eta F_\mu(x) + \int_{u(x)>r^\mu} u(x)d\eta F_\mu(x)\\
=&r^\mu\left[1-(\lambda-1)\eta F_\mu(r^\mu)-\eta\right] + r^\mu\left[\lambda\eta F_\mu(r^\mu)) - \lambda\eta \lim_{u(x)\to r^{\mu-}} F_\mu(x)\right]\\
&+ \int_{u(x)<r^\mu} u(x)d(g\circ F_\mu)(x) + \int_{u(x)>r^\mu} u(x)d(g\circ F_\mu)(x)\\
=&r^\mu\left[1-\eta+F_\mu(r^\mu) - \lambda\eta \lim_{u(x)\to r^{\mu-}} F_\mu(x)\right] + \int_{u(x)<r^\mu} u(x)d(g\circ F_\mu)(x) + \int_{u(x)>r^\mu} u(x)d(g\circ F_\mu)(x)\\
=&r^\mu\left[g(F_\mu(r^\mu)) - g\left(\lim_{u(x)\to r^{\mu-}} F_\mu(x)\right)\right] + \int_{u(x)<r^\mu} u(x)d(g\circ F_\mu)(x) + \int_{u(x)>r^\mu} u(x)d(g\circ F_\mu)(x)\\
=&\int u(x)d(g\circ F_\mu)(x),
\end{aligned}
$$

where g is defined as in the proposition.

Q.E.D.

To show that $\succsim$ cannot satisfy a *strict* preference for late resolution of uncertainty, note that axioms 1-8 with PLRU replacing PERU implies that each $\succsim_a$ can be represent by an AD representation where the reference point is the worst possible outcome instead of being optimal. By the reasoning in Lemma 17 that, for completeness is reproduced here, this will lead to a violation of dynamic consistency.

First, remember that an AD representation can be written as

$$V_a(m) = \int [v_a^1(c) + v^2(\hat{m}^1) + v^3(\hat{m}^2 \times \hat{m}^3)]dm(c,\hat{m}), \tag{3.44}$$

where v^2 is concave by PLRU and v^3 is an affine function. Assume to get a contradiction

that there are two lotteries m^1 and $\hat{m}^1$ such that

$$(1-\alpha)v^2(m^1)+\alpha v^2(\hat{m}^1)<v^2((1-\alpha)m^1+\alpha\hat{m}^1),$$

where $\alpha \in (0,1)$, that is, v^2 is strictly concave for two lotteries. Since v^3 is linear in probabilities, it is possible to find $m^2 \times m^3$ and $\hat{m}^2 \times \hat{m}^3$ such that $V_a(a) < V_a(\hat{m})$ where $a = ((1-\alpha)m^1+\alpha\hat{m}^1) \times m^2 \times m^3$ and $\hat{m} = \hat{m}^1 \times \hat{m}^2 \times \hat{m}^3$ with

$$\int \beta h(x)d(m^2 \times m^3)(x) + (1-\alpha)(v^2(m^1)-v^2(\hat{m}^1)) = \int \beta h(x)d(\hat{m}^2 \times \hat{m}^2)(x) - \varepsilon$$

for $\beta h = v^2 + v^3$ and $\varepsilon > 0$. Such a lottery $\hat{m} \in D$ exists for any $\alpha \in (0,1)$ by continuity of the preferences if m^1 and $\hat{m}^1$ are close enough. By continuity, as ε tends to zero the resulting $\hat{m}$ is such that $h(a) > h(\hat{m})$, violating dynamic consistency as $(c,a) \succsim_{\bar{a}} (c,\hat{m})$ but $\hat{m} \succ_a a$.

Bibliography

Abel, A. B. (1990): "Asset Prices under Habit Formation and Catching up with the Joneses," *The American Econoic review*, 80, papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association. [46], [86]

Aliprantis, C. D. and K. C. Border (2006): *Infinite dimensional analysis*, Springer. [68], [99], [100]

Andries, M. (2019): "Risk pricing under gain-loss asymmetry," *Working paper*. [87]

Attanasio, O. P. and G. Weber (2010): "Consumption and saving: models of intertemporal allocation and their implications for public policy," *Journal of Economic literature*, 48, 693–751. [91]

Backus, D. K., B. R. Routledge, and S. E. Zin (2004): "Exotic preferences for macroeconomists," *NBER Macroeconomics Annual*, 19, 319–390. [47]

Balkenborg, D. (1994): "Strictness and evolutionary stability," *Discussion Paper 52, The Center for Rationality and Interactive Decision Theory, The Hebrew University of Jerusalem*. [39]

Balkenborg, D., J. Hofbauer, and C. Kuzmics (2015): "The refined best-response correspondence in normal form games." *International Journal of Game Theory*, 44, 165 – 193. [10], [20]

Balkenborg, D. and K. H. Schlag (2001): "Evolutionarily stable sets," *International Journal of Game Theory*, 29, 571–595. [40]

Bansal, R. and A. Yaron (2004): "Risks for the long run: A potential resolution of asset pricing puzzles," *The journal of Finance*, 59, 1481–1509. [49], [87], [88]

Barberis, N. and M. Huang (2009): "Preferences with frames: a new utility specification that allows for the framing of risks," *Journal of Economic Dynamics and Control*, 33, 1555–1576. [87]

Barberis, N., M. Huang, and T. Santos (2001): "Prospect theory and asset prices," *The quarterly journal of economics*, 116, 1–53. [87]

Basu, K. and J. Weibull (1991): "Strategy subsets closed under rational behavior," *Economics Letters*, 36, 141 – 146. [1], [2], [7], [8], [17]

Becker, G. S. and K. M. Murphy (1988): "A theory of rational addiction," *Journal of political Economy*, 96, 675–700. [51]

Bell, D. E. (1985): "Disappointment in decision making under uncertainty," *Operations research*, 33, 1–27. [47], [50], [63]

Ben-Tal, A. and M. Teboulle (1986): "Expected utility, penalty functions, and duality in

stochastic nonlinear programming," *Management Science*, 32, 1445–1466. [72]

——— (2007): "An old-new concept of convex risk measures: The optimized certainity equivalent," *Mathematical Finance*, 17, 449–476. [72], [84]

Benartzi, S. and R. H. Thaler (1995): "Myopic loss aversion and the equity premium puzzle," *The quarterly journal of Economics*, 110, 73–92. [87]

Bogachev, V. I. (2007): *Measure theory*, vol. 2, Springer Science & Business Media. [93], [110]

Border, K. C. and U. Segal (1994): "Dynamic consistency implies approximately expected utility preferences," *Journal of Economic Theory*, 63, 170–188. [68]

Bowman, D., D. Minehart, and M. Rabin (1999): "Loss aversion in a consumption–savings model," *Journal of Economic Behavior & Organization*, 38, 155–178. [51]

Brunnermeier, M. K. and J. A. Parker (2005): "Optimal expectations," *The American Economic Review*, 95, 1092–1118. [51], [55]

Caballero, R. J. (1990): "Consumption puzzles and precautionary savings," *Journal of monetary economics*, 25, 113–136. [89], [90]

Campbell, J. and A. Deaton (1989): "Why is consumption so smooth?" *The Review of Economic Studies*, 56, 357–373. [91]

Campbell, J. Y. and J. H. Cochrane (1999): "By force of habit: A consumption-based explanation of aggregate stock market behavior," *Journal of political Economy*, 107, 205–251. [46], [51]

Caplin, A. and J. Leahy (2001): "Psychological expected utility theory and anticipatory feelings," *The Quarterly Journal of Economics*, 116, 55–79. [51]

Carroll, P., K. Sweeny, and J. A. Shepperd (2006): "Forsaking optimism," *Review of general psychology*, 10, 56–73. [74]

Cerreia-Vioglio, S., F. Maccheroni, and M. Marinacci (2017): "Stochastic dominance analysis without the independence axiom," *Management Science*, 63, 1097–1109. [79], [109]

Chatterjee, K. and R. V. Krishna (2011): "A nonsmooth approach to nonexpected utility theory under risk," *Mathematical Social Sciences*, 62, 166–175. [68], [79]

Chetty, R. and A. Szeidl (2016): "Consumption commitments and habit formation," *Econometrica*, 84, 855–890. [51], [91], [92]

Chew, S. H. and L. G. Epstein (1991): "Recursive utility under uncertainty," in *Equilibrium theory in infinite dimensional spaces*, Springer, 352–369. [56]

Chew, S. H., L. G. Epstein, and U. Segal (1991): "Mixture symmetry and quadratic utility,"

Econometrica: Journal of the Econometric Society, 139–163. [56], [93]

Cho, I.-K. and D. M. Kreps (1987): "Signaling games and stable equilibria," *The Quarterly Journal of Economics*, 102, 179–221. [10]

Constantinides, G. M. (1990): "Habit formation: A resolution of the equity premium puzzle," *Journal of political Economy*, 98, 519–543. [51]

Debreu, G. (1960): "Topological methods in cardinal utility theory," Tech. rep., Cowles Foundation for Research in Economics, Yale University. [98], [111]

Dekel, E., B. L. Lipman, A. Rustichini, and T. Sarver (2007): "Representing Preferences with a Unique Subjective State Space: A Corrigendum 1," *Econometrica*, 75, 591–600. [61], [67], [95], [96]

Demichelis, S. and K. Ritzberger (2003): "From Evolutionary to Strategic Stability," *Journal of Economic Theory*, 113, 1–25. [15], [21]

Dillenberger, D., D. Gottlieb, and P. Ortoleva (2019): "Stochastic Impatience and the separation of Time and Risk Preferences," . [88]

Dudley, R. M. (2002): *Real analysis and probability*, Cambridge University Press. [95]

Epstein, L. G., E. Farhi, and T. Strzalecki (2014): "How much would you pay to resolve long-run risk?" *American Economic Review*, 104, 2680–97. [49], [85], [87], [88]

Epstein, L. G. and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–969. [51], [56], [93], [94]

——— (1990): "'First-order'risk aversion and the equity premium puzzle," *Journal of monetary Economics*, 26, 387–407. [87]

Ergin, H. and T. Sarver (2010a): "A unique costly contemplation representation," *Econometrica*, 78, 1285–1339. [67]

——— (2010b): "The unique minimal dual representation of a convex function," *Journal of Mathematical Analysis and Applications*, 370, 600–606. [67], [80], [100], [102]

——— (2015): "Hidden actions and preferences for timing of resolution of uncertainty," *Theoretical Economics*, 10, 489–541. [68]

Evdokimov, P. and A. Rustichini (2016): "Forward induction: thinking and behavior," *Journal of Economic Behavior & Organization*, 128, 195–208. [33]

Fishburn, P. C. (1970): *Utility theory for decision making*, Wiley, 99, illustrated, reprint ed. [96]

Gal, D. and D. D. Rucker (2018): "The loss of loss aversion: Will it loom larger than its gain?"

Journal of Consumer Psychology, 28, 497–516. [48]

Gale, D. (1967): "A geometric duality theorem with economic applications," *The Review of Economic Studies*, 34, 19–24. [68]

Ghirardato, P., F. Maccheroni, and M. Marinacci (2004): "Differentiating ambiguity and ambiguity attitude," *Journal of Economic Theory*, 118, 133–173. [101]

Gollier, C. and A. Muermann (2010): "Optimal Choice and Beliefs with Ex Ante Savoring and Ex Post Disappointment," *Management Science*, 56, 1272–1284. [51], [52], [72], [80]

Govindan, S. and R. Wilson (2009): "On forward induction," *Econometrica*, 77, 1–28. [33]

Grossman, S. T. and G. Laroque (1990): "Asset pricing and optimal portfolio choice in the presence of illiquid durable consumer goods," *Econometrica*. [92]

Gul, F. (1991): "A theory of disappointment aversion," *Econometrica: Journal of the Econometric Society*, 667–686. [50]

Gül, F., D. Pearce, and E. Stacchetti (1993): "A bound on the proportion of pure strategy equilibria in generic games," *Mathematics of Operations Research*, 18, 548–552. [15]

Gul, F. and W. Pesendorfer (2004): "Self-control and the theory of consumption," *Econometrica*, 72, 119–158. [56], [61], [76], [97]

Hanany, E. and P. Klibanoff (2007): "Updating preferences with multiple priors," *Theoretical Economics*. [62]

——— (2009): "Updating ambiguity averse preferences," *The BE Journal of Theoretical Economics*, 9. [62]

Harsanyi, J. C. (1973a): "Games with randomly disturbed payoffs: A new rationale for mixed-strategy equilibrium points," *International journal of game theory*, 2, 1–23. [33]

——— (1973b): "Oddness of the number of equilibrium points: A new proof," *International Journal of Game Theory*, 2, 235–250. [15]

Hauk, E. and S. Hurkens (2002): "On Forward Induction and Evolutionary and Strategic Stability," *Journal of Economic Theory*, 106, 66 – 90. [22], [24], [33], [34], [35]

Hille, S. C. and D. T. Worm (2009): "Embedding of semigroups of Lipschitz maps into positive linear semigroups on ordered Banach spaces generated by measures," *Integral Equations and Operator Theory*, 63, 351–371. [99]

Hofbauer, J., P. Schuster, K. Sigmund, et al. (1979): "A note on evolutionary stable strategies and game dynamics," *Journal of Theoretical Biology*, 81, 609–612. [39]

Hurwicz, L. (2008): "But Who Will Guard the Guardians?" *The American Economic Review*, 98,

577–585. [23]

Isoni, A., G. Loomes, and R. Sugden (2011): "The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations: Comment," *American Economic Review*, 101, 991–1011. [69]

Jappelli, T. and L. Pistaferri (2010): "The consumption response to income changes," *Annu. Rev. Econ.*, 2, 479–506. [49], [55], [89], [91]

Jiang, J.-H. (1963): "Essential component of the set of fixed points of the multivalued mappings and its application to the theory of games," *Scientia Sinica*, 12, 951. [3], [6], [15], [26], [30]

Kahneman, D. and A. Tversky (1979): "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the econometric society*, 263–291. [48], [50], [58], [82]

Kalai, E. and D. Samet (1984): "Persistent equilibria in strategic games," *International Journal of Game Theory*, 13, 129–144. [20]

Kibris, O., Y. Masatlioglu, and E. Suleymanov (2018): "A theory of reference point formation," *Working Paper*. [50]

Kohlberg, E. and J.-F. Mertens (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1003–1037. [2], [3], [19], [24], [27], [32], [33]

Kőszegi, B. (2010): "Utility from anticipation and personal equilibrium," *Economic Theory*, 44, 415–444. [51], [62]

Kőszegi, B. and M. Rabin (2006): "A model of reference-dependent preferences," *The Quarterly Journal of Economics*, 121, 1133–1165. [46], [50], [63]

——— (2007): "Reference-dependent risk attitudes," *The American Economic Review*, 97, 1047–1073. [50], [119], [120]

——— (2009): "Reference-dependent consumption plans," *The American Economic Review*, 99, 909–936. [50], [55], [92]

Kreps, D. M. and E. L. Porteus (1978): "Temporal resolution of uncertainty and dynamic choice theory," *Econometrica: journal of the Econometric Society*, 185–200. [48], [51], [56], [61]

——— (1979): "Temporal von neumann-morgenstern and induced preferences," *Journal of Economic Theory*, 20, 81 – 109. [68], [80], [83]

Kuzmics, C., D. Balkenborg, J. Hofbauer, et al. (2013): "Refined best-response correspondence and dynamics," *Theoretical Economics*, 8. [20], [21]

Laslier, J.-F. and K. V. d. Straeten (2004): "Electoral competition under imperfect information," *Economic Theory*, 24, 419–446. [9]

Lewis, D. (1969): *Convention: A Philosophical Study*, Harvard University Press. [2]

Ljungqvist, L. and H. Uhlig (2015): "Comment on the Campbell-Cochrane habit model," *Journal of Political Economy*, 123, 1201–1213. [46]

Loewenstein, G. (1987): "Anticipation and the valuation of delayed consumption," *The Economic Journal*, 97, 666–684. [51]

Loomes, G. and R. Sugden (1986): "Disappointment and dynamic consistency in choice under uncertainty," *The Review of Economic Studies*, 53, 271–282. [50]

Ludvigson, S. C. and A. Michaelides (2001): "Does buffer-stock saving explain the smoothness and excess sensitivity of consumption?" *American Economic Review*, 91, 631–647. [91]

Maccheroni, F. (2002): "Maxmin under risk," *Economic Theory*, 19, 823–831. [68]

Machina, M. J. (1982): ""Expected Utility" Analysis without the Independence Axiom," *Econometrica*, 50, 277–323. [79]

——— (1984): "Temporal risk and the nature of induced preferences," *Journal of Economic Theory*, 33, 199 – 231. [68]

——— (1989): "Dynamic consistency and non-expected utility models of choice under uncertainty," *Journal of Economic Literature*, 27, 1622–1668. [61], [62]

Mailath, G. J., L. Samuelson, and J. M. Swinkels (1993): "Extensive Form Reasoning in Normal Form Games," *Econometrica*, 61, 273–302. [2]

Markowitz, H. (1952): "The utility of wealth," *Journal of political Economy*, 60, 151–158. [46]

Masatlioglu, Y. and C. Raymond (2016): "A behavioral analysis of stochastic reference dependence," *The American Economic Review*, 106, 2760–2782. [84]

Matsui, A. (1992): "Best response dynamics and socially stable strategies," *Journal of Economic Theory*, 57, 343 – 362. [28], [29]

Maynard Smith, J. (1982): *Evolution and the theory of games*, Cambridge: Cambridge U. P. [25], [39]

Maynard Smith, J. and G. Price (1973): "The Logic of Animal Conflict," *Nature*, 246, 15. [23], [32], [39], [42]

McClennen, E. F. (1988): "Dynamic choice and rationality," in *Risk, decision and rationality*, Springer, 517–536. [61]

McClennen, E. F., E. F. M. McClennen, et al. (1990): *Rationality and dynamic choice: Foundational explorations*, Cambridge university press. [61]

McLennan, A. (2016): "The Index +1 Principle," *Mimeo*. [14], [19]

Mehra, R. and E. C. Prescott (1985): "The equity premium: A puzzle," *Journal of monetary Economics*, 15, 145–161. [84], [87]

Meissner, T. and P. Pfeiffer (2018): "Measuring Preferences Over the Temporal Resolution of Consumption Uncertainty," *Working Paper*. [88]

Mertens, J.-F. (1989): "Stable equilibria–A reformulation. Part 1. Definition and basic properties," *Mathematics of Operations Research*, 14, 575 – 625. [3], [15]

——— (1991): "Stable equilibria–A reformulation. Part II. Discussion of the definition and further results," *Mathematics of Operations Research*, 16, 694 – 753. [3], [15]

Michaelides, A. (2002): "Buffer stock saving and habit formation," *Available at SSRN 302079*. [91]

Myerson, R. (1978): "Refinements of the Nash equilibrium concept," *International Journal of Game Theory*, 7, 73–80. [2], [3], [7], [26]

Myerson, R. and J. Weibull (2015): "Tenable Strategy Blocks and Settled Equilibria," *Econometrica*, 83, 943–976. [1], [2], [3], [7], [10], [11], [13], [17], [22], [23], [24], [27], [35], [36], [37], [40], [41], [43]

Nash, J. (1950): "Non-cooperative Games," Ph.D. thesis, Chichester;Princeton, N.J;, 53-84. [2], [23]

Neilson, W. S. (2006): "Axiomatic reference-dependence in behavior toward others and toward risk," *Economic Theory*, 28, 681–692. [47], [63]

ODonoghue, T. and C. Sprenger (2018): "Reference-dependent preferences," *Handbook of Behavioral Economics-Foundations and Applications*, 1, 1. [46], [47], [49], [75]

Ok, E. A., P. Ortoleva, and G. Riella (2015): "Revealed (p) reference theory," *American Economic Review*, 105, 299–321. [50]

Pagel, M. (2016): "Expectations-based reference-dependent preferences and asset pricing," *Journal of the European Economic Association*, 14, 468–514. [86], [87]

——— (2017): "Expectations-based reference-dependent life-cycle consumption," *The Review of Economic Studies*, 84, 885–934. [92]

Post, T., M. J. Van den Assem, G. Baltussen, and R. H. Thaler (2008): "Deal or no deal? decision making under risk in a large-payoff game show," *American Economic Review*, 98, 38–71. [69]

Quiggin, J. (1982): "A theory of anticipated utility," *Journal of Economic Behavior & Organization*, 3, 323–343. [83]

Rabin, M. (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica*, 68, 1281–1292. [49]

Ritzberger, K. (1994): "The theory of normal form games from the differentiable viewpoint," *International Journal of Game Theory*, 23, 207–236. [4], [14], [15], [34]

——— (2002): *Foundations of non-cooperative game theory*, Oxford: Oxford Univ. Press. [14]

Ritzberger, K. and J. Weibull (1995): "Evolutionary Selection in Normal-Form Games," *Econometrica*, 63, 1371–1399. [8], [21]

Rozen, K. (2010): "Foundations of intrinsic habit formation," *Econometrica*, 78, 1341–1373. [47], [51], [63]

Ryder, H. E. and G. M. Heal (1973): "Optimal growth with intertemporally dependent preferences," *The Review of Economic Studies*, 40, 1–31. [46], [51]

Samuelson, W. and R. Zeckhauser (1988): "Status quo bias in decision making," *Journal of risk and uncertainty*, 1, 7–59. [48], [69]

Sarver, T. (2018): "Dynamic Mixture-Averse Preferences," *Econometrica*, 86, 1347–1382. [51], [55], [68], [72], [79], [83], [84], [111], [112], [114], [120]

Schelling, T. C. (1960): *The strategy of conflict*, Harvard university press. [2]

Schmeidler, D. (1989): "Subjective probability and expected utility without additivity," *Econometrica: Journal of the Econometric Society*, 571–587. [83]

Schmidt, U. (2003): "Reference dependence in cumulative prospect theory," *Journal of Mathematical Psychology*, 47, 122–131. [47], [63]

Segal, U. (1997): "Dynamic consistency and reference points," *journal of economic theory*, 72, 208–219. [68]

Segal, U. and A. Spivak (1990): "First order versus second order risk aversion," *Journal of Economic Theory*, 51, 111–125. [82]

Selten, R. (1975): "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory*, 4, 25–55. [2]

——— (1980): "A note on evolutionarily stable strategies in asymmetric animal conflicts," *Journal of Theoretical Biology*, 84, 93–101. [32]

Sprenger, C. (2015): "An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points," *Journal of Political Economy*, 123, 1456–1499. [69]

Sweeny, K. and Z. Krizan (2013): "Sobering up: A quantitative review of temporal declines in expectations." *Psychological Bulletin*, 139, 702. [74]

Sweeny, K. and J. A. Shepperd (2010): "The costs of optimism and the benefits of pessimism." *Emotion*, 10, 750. [74]

Swinkels, J. M. (1992a): "Evolution and strategic stability: From maynard smith to kohlberg and mertens," *Journal of Economic Theory*, 57, 333 – 342. [27], [28], [41]

——— (1992b): "Evolutionary stability with equilibrium entrants," *Journal of Economic Theory*, 57, 306 – 332. [22], [24], [26], [27], [28], [29], [30], [41]

Tercieux, O. (2006a): "p-Best response set," *Journal of Economic Theory*, 131, 45 – 70. [20]

——— (2006b): "p-Best response set and the robustness of equilibria to incomplete information," *Games and Economic Behavior*, 56, 371 – 384. [20]

Thaler, R. (1980): "Toward a positive theory of consumer choice," *Journal of Economic Behavior & Organization*, 1, 39–60. [69]

Thomas, B. (1985): "On evolutionarily stable sets," *Journal of mathematical Biology*, 22, 105–115. [25], [39], [41]

Tserenjigmid, G. (2018): "Choosing with the worst in mind: A reference-dependent model," *Journal of Economic Behavior & Organization*. [50]

——— (2019): "On the characterization of linear habit formation," *Economic Theory*, 1–45. [47], [51], [63]

Tversky, A. and D. Kahneman (1991): "Loss aversion in riskless choice: A reference-dependent model," *The quarterly journal of economics*, 106, 1039–1061. [82]

van Damme, E. (1989): "Stable equilibria and forward induction," *Journal of Economic Theory*, 48, 476 – 496. [22], [24]

van Damme, E. (1989): "Stable equilibria and forward induction," *journal of Economic Theory*, 48, 476–496. [33], [35]

van Damme, E. (1991): *Stability and perfection of Nash equilibria*, New York;Berlin;: Springer-Verlag, 2nd, rev. and enl. ed. [12], [30], [33]

Van Dijk, W. W., M. Zeelenberg, and J. Van der Pligt (2003): "Blessed are those who expect nothing: Lowering expectations as a way of avoiding disappointment," *Journal of Economic Psychology*, 24, 505–516. [74]

Voorneveld, M. (2004): "Preparation," *Games and Economic Behavior*, 48, 403 – 414. [20]

——— (2005): "Persistent retracts and preparation," *Games and Economic Behavior*, 51, 228 – 232. [20]

Weibull, J. (1995): *Evolutionary game theory*, Cambridge, Mass;London;: MIT Press. [39]

Weil, P. (1990): "Nonexpected utility in macroeconomics," *The Quarterly Journal of Economics*, 105, 29–42. [51]

Wikman, P. (2020): "Nash Blocks," *Working Paper*. [34], [40]

Wu, W.-T. and J.-H. Jiang (1962): "Essential equilibrium points of n-person noncooperative games," *Scientia Sinica*, 11, 1307–1322. [2]

Xu, Z. (2019): "Tenable Blocks and Settled Equilibria: a Bridge between Evolutionary Stability and Strategic Stability," *Working Paper*. [41]

Yechiam, E. (2018): "Acceptable losses: The debatable origins of loss aversion," *Psychological research*, 1–13. [48]

Yogo, M. (2008): "Asset prices under habit formation and reference-dependent preferences," *Journal of Business & Economic Statistics*, 26, 131–143. [51], [87]

Young, H. P. (1993): "The Evolution of Conventions," *Econometrica*, 61, 57–84. [2], [21]