CHAPTER 10

Game Theory

In this chapter we continue the study of the testable implications of models of collective behavior. We focus here on game-theoretic models. Our first results use a version of the abstract choice environments from Chapter 2, and discuss testing Nash equilibrium as the prediction of the choices made by a collection of agents. We then turn to models of bargaining and two-sided matching.

10.1 NASH EQUILIBRIUM

Let N be a finite set of agents of cardinality n. For each $i \in N$, consider some finite set $\bar{S}_i$. The idea is that $\bar{S}_i$ is some "global" set of strategies that may be available to agent i. In a specific (observable) instance, player i is restricted to choosing a strategy from $S_i \subseteq \bar{S}_i$, a "budget" of possible strategies. Consider the nonempty product subsets of $\prod_{i\in N} \bar{S}_i$, denoted by $\mathcal{S}$. A typical element of $\mathcal{S}$ has the form $S_1 \times S_2 \times \cdots \times S_n$; where each S_i is a nonempty subset of $\bar{S}_i$. These sets are called *game forms*.

An element of $\prod_{i\in N} \bar{S}_i$ is called a *strategy profile*. As usual in game theory we use (s_i, s_{-i}) to denote a strategy profile in which s_i is i's strategy and $s_{-i} \in \prod_{j\in N\setminus\{i\}} \bar{S}_j$ is a strategy profile for players in $N \setminus \{i\}$.

A *joint choice function* is a mapping $c : \Sigma \subseteq \mathcal{S} \to 2^{\prod_{i\in N} \bar{S}_i} \setminus \{\varnothing\}$ satisfying $c(S) \subseteq S$. Note that $c(S)$ need not have a product structure, but it is required to be nonempty. In particular, every joint choice function is a choice function in the sense of Chapter 2. We will be interested in notions of rationalization that reflect the product structure of $\mathcal{S}$.

Let $\{\succeq_i\}_{i\in N}$ be a family of preference relations, each over $\prod_{i\in N} \bar{S}_i$. Note that these preferences define a (normal-form) *game* $(N, (S_i, \succeq_i|_S)_{i\in N})$ for each $S = S_1 \times S_2 \times \cdots \times S_n \in \mathcal{S}$. By $\succeq_i|_S$ we mean the restriction of preferences $\succeq_i$ to S. A *strategy profile* $s = (s_i)_{i\in N} \in S$ is a *Nash equilibrium* of $(N, (S_i, \succeq_i|_S)_{i\in N})$ if $(s_i, s_{-i}) \succeq_i (s'_i, s_{-i})$ for all $s'_i \in S_i$ and all $i \in N$.

We say that the family of preference relations $\{\succeq_i\}_{i\in N}$ *strongly Nash rationalize* a choice function c if for all $S = \prod_{i\in N} S_i \in \mathcal{S}$, $c(S)$ coincides with

the set of Nash equilibria of the game $(N, (S_i, \succeq_i|_S)_{i\in N})$:

$$c(S) = \{(s_i)_{i\in N} : \text{ for all } j \in N \text{ and all } s'_j \in S_j, (s_j, s_{-j}) \succeq_j (s'_j, s_{-j})\}.$$

We say that c is *strongly Nash rationalizable* by a class of preference relations if there exist preference relations $\succeq_i$ in that class which strongly Nash rationalize c.

We say that a family of binary relations $\succeq_i$ *weakly Nash rationalize* c if for all $S \in \mathcal{S}$, $c(S) \subseteq \{(s_i)_{i\in N} : \text{ for all } j \in N \text{ and all } s'_j \in S_j, (s_j, s_{-j}) \succeq_j (s'_j, s_{-j})\}$. We will say c is *weakly Nash rationalizable* by a class of preference relations if there exist preference relations $\succeq_i$ in that class which weakly Nash rationalize c.

Clearly, weak Nash rationalizability by weak orders has no empirical content; thus, when we speak of weak Nash rationalization, we shall take interest in weak Nash rationalization by strict preference relations.

10.1.1 Choice from all game forms

We shall assume that we observe the outcomes of the strategic interaction of the agents in N for each possible game form. That is, $\Sigma = \mathcal{S}$.[1] The results in this subsection are due to Yves Sprumont.

The set $\mathcal{S}$ forms a lattice under the set inclusion relation; its meet is given by $S \wedge S' = S \cap S'$. Writing $S = \prod_{i\in N} S_i$ and $S' = \prod_{i\in N} S'_i$, the join $S \vee S'$ is given by $S \vee S' = \prod_{i\in N}(S_i \cup S'_i)$.

Say that a joint choice function c satisfies *persistence under expansion* if for all $S, S' \in \mathcal{S}$, $c(S) \cap c(S') \subseteq c(S \vee S')$. It is obvious from the definition of Nash equilibrium that persistence under expansion is a necessary condition for c to be strongly Nash rationalizable.

We will say that $S \in \mathcal{S}$ is a *line* if there exists $i \in N$ such that for all $j \neq i$, S_i is a singleton. A line can be understood as a single-agent decision problem. We say that a joint choice function c satisfies *persistence under contraction* if the following two properties are satisfied:

- for all $S, S' \in \mathcal{S}$ with $S \subseteq S'$, $c(S') \cap S \subseteq c(S)$.
- If S' is a line, $S \subseteq S'$, and $c(S') \cap S \neq \varnothing$, then $c(S) \subseteq c(S')$.

Note that the first property of persistence under contraction is condition α of Chapter 2. The second property, which is only required to hold for lines, is similar to condition β.

The following theorem characterizes strongly Nash rationalizable joint choice functions.

[1] This assumption is analogous to assuming that the choices from all possible budgets are observed in Chapter 2.

Theorem 10.1 *A joint choice function c is strongly Nash rationalizable by preference relations iff it satisfies persistence under expansion and persistence under contraction.*

Proof. It is easy to verify that if c is strongly Nash rationalizable by preference relations, then it satisfies persistence under expansion and persistence under contraction. Conversely, suppose c satisfies these two properties.

We first define a basic revealed preference, $\succeq_i^c$. For any pair (s_i, s_{-i}) and (s_i', s_{-i}), we define $(s_i, s_{-i}) \succeq_i^c (s_i', s_{-i})$ in the natural way; that is, $(s_i, s_{-i}) \succeq_i^c (s_i', s_{-i})$ iff $(s_i, s_{-i}) \in c\left(\{s_i, s_i'\} \times \prod_{j \neq i}\{s_j\}\right)$. Each $\succeq_i^c$ is clearly reflexive. Transitivity is also straightforward and follows similarly to the proof of Theorem 2.9. Hence, each $\succeq_c^i$ has an extension (by Theorem 1.5) to a weak order $\succeq_i$.

We claim that the orders $\succeq_i$ strongly Nash rationalize c. So, suppose we have given $S \in \mathcal{S}$ and $s \in S$ such that for all $i \in N$ and all $s_i' \in S_i$, $(s_i, s_{-i}) \succeq_i (s_i', s_{-i})$. Hence, for all $i \in N$ and all $s_i' \in S_i$, $(s_i, s_{-i}) \succeq_i^c (s_i', s_{-i})$, so that $(s_i, s_{-i}) \in c\left(\{s_i, s_i'\} \times \prod_{j \neq i}\{s_j\}\right)$. Now, $S = \bigvee_{i \in N} \bigvee_{s_i' \in S_i} \left(\{s_i, s_i'\} \times \prod_{j \neq i}\{s_j\}\right)$, so that by persistence under expansion (applied inductively), we have $s \in c(S)$. Conversely, suppose that $s \in c(S)$. Then it follows by persistence under contraction that $s \in c\left(\{s_i, s_i'\} \times \prod_{j \neq i}\{s_j\}\right)$ for all i and all s_i', so that $(s_i, s_{-i}) \succeq_i^c (s_i', s_{-i})$, or $(s_i, s_{-i}) \succeq_i (s_i', s_{-i})$.

The following result characterizes weak rationalizability by strict preference relations. We omit its proof, which is simple.

Theorem 10.2 *A joint choice function c is weakly rationalizable by strict preference relations iff for all $S \in \mathcal{S}$, all $s \in c(S)$, all $i \in N$, and all $s_i' \in S_i$, $c\left(\{s_i', s_i\} \times \prod_{j \neq i}\{s_j\}\right) = \{s\}$.*

The notion of Nash rationalizability is based on individual agents' incentives to choose strategies. We can instead consider what the group of agents N would jointly decide to do. One natural property is that the chosen strategy profiles should not be dominated: A joint choice function c is *Pareto rationalizable* by preference relations iff there are preference relations $\succeq_i$ on $\prod_{i \in N} \bar{S}_i$ such that for all $S \in \mathcal{S}$, $s \in c(S)$ iff there is no $s' \in S$ for which $s' \succeq_i s$ for all i, with at least one preference strict. Say it is *team rationalizable* if there exists a preference $\succeq$ over $\prod_{i \in N} \bar{S}_i$ such that for all $S \in \mathcal{S}$, $c(S)$ is the set of $\succeq$-maximal elements of S.

Theorem 10.3 *Suppose a joint choice function c satisfies $|c(S)| = 1$ for all $S \in \mathcal{S}$. Then c is Pareto rationalizable iff it is team rationalizable.*

Proof. If c is team rationalizable, it is clearly Pareto rationalizable. Conversely, suppose c is Pareto rationalizable, and let $\{\succeq_i\}_{i \in N}$ be the rationalizing relations. Define $s \succeq^* s'$ if it is not the case that $s \succeq_i s'$ for all $i \in N$ with at least one

preference strict. Clearly then, $s \succ^* s'$ iff $s \succeq_i s'$ for all $i \in N$ and there is $j \in N$ for which $s \succ_j s'$. Hence, $\succeq^*$ is quasitransitive. By definition, for every S, $c(S)$ coincides with the $\succeq^*$ maximal elements of S. The result now follows from Proposition 2.7.

In fact, Theorem 10.3 does not rely on the product structure of $\mathcal{S}$ and could be proved more generally.

The following theorem establishes that the implications of Nash rationalizability are stronger than those of Pareto rationalizability. We omit its proof, which is technical.

Theorem 10.4 *Suppose that $|N| = 2$, and that a joint choice function c satisfies $|c(S)| = 1$ for all $S \in \mathcal{S}$. Then if c is Nash rationalizable, it is Pareto rationalizable.*

10.1.2 Choice from a subset of game forms

We now discuss results in the revealed preference approach to game theory when not all possible game forms in $\mathcal{S}$ are observable.

We will say a collection $\Sigma \subseteq \mathcal{S}$ is *line-closed* if for all $S \in \Sigma$, all $s \in S$, and all $i \in N$, $S_i \times \prod_{i \neq j \in N} \{s_j\} \in \Sigma$. Line-closedness is enough to allow us to generate a meaningful revealed preference relation. A joint choice function can now be defined on Σ, and it is meaningful to discuss strong rationalizability by preferences.

The following idea is due to Galambos. We define the following revealed preference pair. Define $\succeq_i^c$ by $s \succeq_i^c s'$ if for all $j \neq i$, $s_j = s'_j$, and there exists $S \in \Sigma$ for which $s, s' \in S$ and $s \in c(S)$. We say a joint choice function c defined on Σ satisfies *N-congruence* if for all $i \in N$ and all S, if for $s \in S$ we have $s \left(\succeq_i^c\right)^T (s'_i, s_{-i})$ for all i and all $s'_i \in S_i$, then $s \in c(S)$.[2]

Theorem 10.5 *Suppose that Σ is a line-closed domain. Then a joint choice function is strongly rationalizable by preference relations iff it satisfies N-congruence.*

Proof. Suppose a joint choice function is strongly Nash rationalizable by weak orders $\succeq_i$. Clearly, $\left(\succeq_i^c\right)^T \subseteq \succeq_i$. Consequently, if $s \in S$ satisfies $s \left(\succeq_i^c\right)^T (s'_i, s_{-i})$ for all $s'_i \in S_i$, it follows that $s \succeq_i (s'_i, s_{-i})$ for all $s'_i \in S_i$, so that $s \in c(S)$, since the $\{\succeq_i\}_{i \in N}$ strongly Nash rationalize c.

Conversely, suppose that choice function c satisfies N-congruence. If we define $\succ_i^c$ by $(s_i, s_{-i}) \succ_i^c (s'_i, s_{-i})$ if there exists S for which $s, (s'_i, s_{-i}) \in S$ and $s \in c(S)$ but $(s'_i, s_{-i}) \notin c(S)$, then by N-congruence, order pair $\langle \succeq_i^c, \succ_i^c \rangle$ is acyclic, so that by Theorem 1.5, there is a weak order $\succeq_i$ for which $\succeq_i^c \subseteq \succeq_i$ and $\succ_i^c \subseteq \succ_i$.

We claim that the $\{\succeq_i\}_{i \in N}$ strongly Nash rationalize c. Suppose that $s \in c(S)$. We claim that for all i and all $s'_i \in S_i$, $(s_i, s_{-i}) \succeq_i (s'_i, s_{-i})$. This follows as

[2] Recall that for a binary relation $\succeq$, $\succeq^T$ denotes the transitive closure.

$(s_i, s_{-i}) \succeq_i^c (s_i', s_{-i})$ and $\succeq_i^c \subseteq \succeq_i$. On the other hand, suppose that $s \in S$ satisfies $(s_i, s_{-i}) \succeq_i (s_i', s_{-i})$ for all i and all $s_i' \in S_i$. We claim that $s \in c(S)$. Since Σ is line-closed, it follows that $S_i \times \prod_{j \neq i} \{s_j\} \in \Sigma$. Then $s \in c\left(S_i \times \prod_{j \neq i} \{s_j\}\right)$. Otherwise, since c is nonempty-valued, there exists $s_i' \in S_i$ for which $(s_i', s_{-i}) \in \left(S_i \times \prod_{j \neq i} \{s_j\}\right)$, from which it follows that $(s_i', s_{-i}) \succ_i^c s$, contradicting the definition of $\succeq_i$. Consequently, we know that $s \succeq_i^c (s_i', s_{-i})$ for all $i \in N$, so that by N-congruence, $s \in c(S)$.

10.1.3 Zero-sum games

The preceding results impose no structure on the rationalizing games, but one may want to investigate the joint implications of Nash behavior and some properties of the strategic environment. One of the most natural restrictions is the property that the game be a zero-sum game. Zero-sum games are the subject of intense study in game theory; they also arise naturally if one believes that players care about relative, not absolute, payoffs.[3]

Fix $N = \{1, 2\}$. Suppose that $\Sigma = \mathcal{S}$ again, so that we are given choice from all possible game forms. A choice function c is *strongly Nash rationalizable by a zero-sum game* if there is a preference relation $\succeq$ on $\bar{S}_1 \times \bar{S}_2$ such that, for all $S \in \mathcal{S}$, $c(S)$ is the set of Nash equilibria of $(\{1, 2\}, (S_1, \succeq|_S), (S_2, \preceq|_S))$; where we denote by $\preceq$ the dual binary relation defined from $\succeq$, i.e. $x \preceq y$ iff $y \succeq x$.

It turns out that to characterize strong zero-sum rationalizability, all we need is one additional restriction on c. Say that c is *interchangeable* if, for all $S \in \mathcal{S}$, and all $s, s' \in c(S)$, $\{s\} \vee \{s'\} \subseteq c(S)$. Our next result is due to SangMok Lee.

Theorem 10.6 *A choice function is strongly Nash rationalizable by a zero-sum game iff it satisfies persistence under expansion, persistence under contraction, and it is interchangeable.*

We omit the proof of Theorem 10.6, which is somewhat technical.

10.2 BAYESIAN NASH EQUILIBRIUM

The next class of problems we tackle are motivated by mechanism design. Given a function from agents' types to outcomes, we want to know if the function could be an incentive-compatible direct revelation mechanism, for some preferences of the agents.

Formally, let N be a set of agents, and X a finite set of *outcomes*. For each agent $i \in N$, suppose we are given a finite *type space* T_i. A *direct revelation mechanism* is a function $g : T = \prod_{i \in N} T_i \to X$. The question we ask here is simple. Suppose we observe N, $T = \prod_{i \in N} T_i$, and g. Can we rationalize g as a strongly incentive-compatible direct revelation mechanism for some list

[3] For example, suppose that "material" payoffs are p_i and there are two players; $i = 1, 2$. If each player i cares about the difference $p_i - p_{3-i}$, then the sum of payoffs will be zero.

of preferences? In particular, is it possible that these mappings could be the unique Bayesian Nash equilibrium for some preferences in the direct revelation game? This answer and its question are due to John Ledyard.

Formally, let us define a *Bayesian environment* to be a pair of functions, one for each agent, denoted by $u_i : X \times T \to \mathbf{R}$ and $p_i : T_i \to \Delta(T_{-i})$. The function u_i is meant to be i's utility function over outcomes. The function p_i gives i's beliefs about other agents' types.

Two points are worth mentioning here. First, we allow agent i's utility to depend on the entire profile of types. Second, p_i carries each type to a probability measure over the others' types. We denote the probability of type profile t_{-i} by $p_i(t_{-i}|t_i)$. Note, however, that we are not making any common prior assumption.

A Bayesian environment *strongly rationalizes* direct revelation mechanism g if for all $t_i, t_i' \in T_i$, if $g(t) \neq g(t_i', t_{-i})$ for some t_{-i}, then

$$\sum_{t_{-i} \in T_{-i}} u_i(g(t_i, t_{-i}), t) p(t_{-i}|t_i) > \sum_{t_{-i} \in T_{-i}} u_i(g(t_i', t_{-i}), t) p(t_{-i}|t_i). \tag{10.1}$$

Theorem 10.7 *For any direct revelation mechanism, there exists a Bayesian environment strongly rationalizing it.*

Proof. Rewrite Equation (10.1), so that we have:

$$\sum_{t_{-i} \in T_{-i}} \sum_{x \in X} u_i(x,t) p(t_{-i}|t_i) \mathbf{1}_{g(t_i,t_{-i})=x} > \sum_{t_{-i} \in T_{-i}} \sum_{x \in X} u_i(x,t) p(t_{-i}|t_i) \mathbf{1}_{g(t_i',t_{-i})=x};$$

equivalently:

$$\sum_{t_{-i} \in T_{-i}} \sum_{x \in X} u_i(x,t) p(t_{-i}|t_i) [\mathbf{1}_{g(t_i,t_{-i})=x} - \mathbf{1}_{g(t_i',t_{-i})=x}] > 0. \tag{10.2}$$

First, observe that given any t_i, we can find u_i and p_i (depending on t_i) which satisfy equation (10.2) iff there is a function $w_{t_i} : X \times T_{-i} \to \mathbf{R}$ such that for all t_i' for which $g(t_i, t_{-i}) \neq g(t_i', t_{-i})$ for some t_{-i}, then

$$\sum_{t_{-i} \in T_{-i}} \sum_{x \in X} w_{t_i}(x, t_{-i}) [\mathbf{1}_{g(t_i,t_{-i})=x} - \mathbf{1}_{g(t_i',t_{-i})=x}] > 0. \tag{10.3}$$

This follows because it is simple to then find $u_i(x,t)$ and $p_i(t_{-i}|t_i)$ for which $u_i(x,t)p_i(t_{-i}|t_i) = w_{t_i}(x, t_{-i})$ (for example, define $p_i(t_{-i}|t_i) > 0$ arbitrarily so that $\sum_{t_{-i}} p_i(t_{-i}|t_i) = 1$, and then define $u_i(x,t) = \frac{w_{t_i}(x,t_{-i})}{p_i(t_{-i}|t_i)}$). Hence, we turn to equation (10.3). For every t_i, there is a list of such equations.

If we can solve the corresponding list of equations for each t_i, then we have proved that the direct revelation mechanism is strongly rationalizable. So, fix t_i. For each t_i' for which there exists t_{-i} such that $g(t_i, t_{-i}) \neq g(t_i', t_{-i})$, there is an equation of type (10.3). Denote the set of such t_i' by $T_i'(t_i)$.

In equation (10.3), view w as a vector in $\mathbf{R}^{X \times T_{-i}}$. The quantities $[\mathbf{1}_{g(t_i,t_{-i})=x} - \mathbf{1}_{g(t_i',t_{-i})=x}]$ play the role of coefficients. Applying Lemma 1.12, if for some t_i,

a solution to the system described by equation (10.3) does not exist, then there exists, for each $t_i' \in T_i'(t_i)$, $\eta_{t_i'} \geq 0$, not all of which are zero, such that for all t_{-i}, x,

$$\sum_{t_i' \in T_i'(t_i)} \eta_{t_i'} [\mathbf{1}_{g(t_i,t_{-i})=x} - \mathbf{1}_{g(t_i',t_{-i})=x}] = 0. \tag{10.4}$$

By assumption, for any $t_i'' \in T_i'(t_i)$, there are t_{-i} and x for which $g(t_i, t_{-i}) = x \neq g(t_i'', t_{-i})$. Pick such t_{-i} and x. Because $\mathbf{1}_{g(t_i,t_{-i})=x} = 1$, we then have that $[\mathbf{1}_{g(t_i,t_{-i})=x} - \mathbf{1}_{g(t_i'',t_{-i})=x}] = 1$, and further, for any $t_i' \in T_i'(t_i)$, we have $[\mathbf{1}_{g(t_i,t_{-i})=x} - \mathbf{1}_{g(t_i',t_{-i})=x}] \geq 0$. Hence, equation 10.4 implies $\eta_{t_i'} = 0$. We conclude that $\eta = 0$, a contradiction.

10.3 BARGAINING THEORY

We now turn to a formulation of the revealed preference problem for bargaining theory. We can work out the empirical consequences of the most commonly used cooperative theories of bargaining.

Suppose n agents bargain over a fixed quantity of a single-dimensional resource: think of bargaining over a fixed monetary amount, which needs to be allocated among n agents. There is a given disagreement point, a point that specifies monetary outcomes for all the agents in the event that there is no agreement. We imagine that we observe similar agents bargaining in different circumstances, for example workers and firms in wage bargaining. The question is when observed outcomes can be rationalized as consistent with standard bargaining theory.

We shall first describe the theories under consideration, then define the kinds of data we might use to test them. Then, we present a result that characterizes the observable implications of these theories in the case where the disagreement points are fixed and the same for all agents. The surprising implication is that all of these theories are observationally equivalent. Each theory tries to capture a distinct economic phenomenon or criterion, but they turn out to have rather weak empirical consequences, to the point that they are all equivalent.

In fact, we uncover a particularly striking form of observational equivalence. We find preferences (utility functions) that serve to rationalize the data as coming from *any* of the theories. We might expect that two theories are observationally equivalent because a given dataset can be rationalized by one theory or by the other, but normally each rationalization will involve different values for the unobservable variables in the theories (preferences in our case). In the case of bargaining theory, it turns out that we can choose the unobservables so that they work for all rationalizations.

The model is as follows. We assume some quantity $m \in \mathbf{R}_+$ of money, and a vector $d = (d_1, \ldots, d_n)$ that represents the disagreement point. The set

$$B(m,d) = \{(x_1, \ldots, x_n) \in \mathbf{R}_+^n : \sum_{i=1}^n x_i \leq m \text{ and, for all } i, x_i \geq d_i\}$$

is the set of all allocations of m amongst n agents, in which everyone gets at least their disagreement outcomes. The set B is therefore a set of feasible and "individually rational" allocations.

A bargaining theory uses information on agents' preferences to predict an outcome in $B(m,d)$. Suppose that each agent i is described by a strictly increasing and concave utility function $u_i : \mathbf{R}_+ \to \mathbf{R}$. We shall focus on three theories: *utilitarianism*, *Nash bargaining*, and *egalitarianism.*

The *utilitarian* theory calls for maximizing the sum of agents' utilities. It predicts that m is allocated so as to maximize the sum $\sum_{i=1}^n u_i(x_i)$ over $B(m,d)$. In fact, for reasons that will become clear later, we consider a generalization of the utilitarian theory, where for some function $g : A \subseteq \mathbf{R} \to \mathbf{R}$, the sum $\sum_{i=1}^n g(u_i(x_i) - u_i(d_i))$ is maximized over $B(m,d)$.

The *Nash bargaining* theory predicts a choice in $B(m,d)$ that maximizes

$$\prod_{i=1}^n [u_i(x_i) - u_i(d_i)].$$

The expression being maximized is termed the *Nash product*. Note that the Nash bargaining theory is a special case of our generalization of the utilitarian theory, letting $g = \log$.

Finally, the *egalitarian* (or maxmin) theory says that $x \in B(m,d)$ should be chosen to maximize

$$\min_{i \in N} [u_i(x_i) - u_i(d_i)].$$

We assume a set of K observations of bargaining outcomes. Each outcome represents a split of some monetary quantity. We assume that the disagreement points are fixed and the same for all agents: we can, for all intents and purposes, take the disagreement outcome to be zero for all agents. A *dataset* is then a set $D = \{x^k : k = 1, \ldots, K\}$. Each observation k specifies an allocation $x^k = (x_1^k, \ldots, x_n^k) \in \mathbf{R}_+^n$ of the total amount of money $\sum_{i=1}^n x_i^k$. Let $N = \{1, \ldots, n\}$.

Let $g : \mathbf{R}_+ \to \mathbf{R} \cup \{-\infty\}$ be a strictly increasing, smooth, and concave function. We say that data $\{x^k\}_{k=1}^K$ are *g-rationalizable* if there exist strictly increasing, smooth, and strictly concave functions u_i for which $u_i(0) = 0$ and $u_i'(0) = \infty$ (*Inada conditions*), and for which $\sum_{i \in N} g(u_i(x_i^k)) \geq \sum_{i \in N} g(u_i(y_i))$ for all allocations $(y_1, \ldots, y_n) \in B(\sum_i x_i^k, 0)$ and $k = 1, \ldots, K$. The utilitarian and Nash models are special cases of g-rationalizability. Note that the assumption of rationalizability already reflects our assumption that the disagreement point is fixed and the same for all agents.

On the other hand, data $\{x^k\}_{k=1}^K$ are *maxmin rationalizable* if there exist strictly increasing and strictly concave u_i, normalized so that $u_i(0) = 0$, for

which $\min_{i\in N} u_i(x_i^k) \geq \min_{i\in N} u_i(y_i)$ for all $(y_1,\ldots,y_n) \in B(\sum_i x_i^k, 0)$ and $k = 1,\ldots,K$.

We say that data $\{x^k\}_{k=1}^K$ are *comonotonic* if for all $i,j \in N$ and all k,l, $x_i^k < x_i^l$ implies $x_j^k < x_j^l$, and for all $i,j \in N$, $x_i^k = 0$ iff $x_j^k = 0$. Comonotonicity requires that outcomes are perfectly strictly ordinally correlated (when 0 is also considered an outcome).

The following result characterizes the data that are g-rationalizable or maxmin rationalizable.

Theorem 10.8 *Given data $\{x^k\}_{k=1}^K$ and a strictly increasing concave g, the following are equivalent:*

I) *The data are comonotonic.*
II) *The data are g-rationalizable.*
III) *The data are maxmin rationalizable.*

The proof shows more than is stated here. In the proof we construct rationalizing utilities that work for any function g, as well as for the maxmin model. The resulting observational equivalence is therefore unusually strong. We can find unobservable preferences that rationalize the data using any of the models under consideration.

Proof. It follows from the first-order conditions that if the data are either g-rationalizable or maxmin rationalizable, then they are comonotonic.

For the other direction, we show something slightly stronger: If the data are comonotonic, then there exist strictly concave, continuous, and increasing functions u_i such that, if $\varphi : [0,\infty) \to \mathbf{R} \cup \{-\infty\}$ is an increasing, symmetric, and quasiconcave function, then $\varphi(u_1(x_1^k),\ldots,u_n(x_n^k)) \geq \varphi(u_1(y_1),\ldots,u_n(y_n))$ for all allocations $(y_1,\ldots,y_n)$ satisfying $\sum_{i\in N} x_i^k = \sum_{i\in N} y_i$.[4] As a special case, we have $\varphi(z_1,\ldots,z_n) = \sum_{i=1}^n g(z_i)$. Note the order of the quantifiers used above: the same profile of utility functions $u_1,\ldots,u_n$ works across all φ.

To this end, we suppose the data are comonotonic, and ignore replications as well as points where every agent consumes 0. Without loss of generality, let us suppose that $x_i^1 < x_i^2 < \ldots < x_i^K$ for all $i \in N$ (that this is possible follows from comonotonicity). Below we construct a profile of utility functions $u_1,\ldots,u_n$ with the property that for all $k = 1,\ldots,K$, $\sum_{i\in N} u_i(x_i^k)$ is maximal across all allocations $y_1,\ldots,y_n$ for which $\sum_{i\in N} x_i^k = \sum_{i\in N} y_i$, and $\min_{i\in N} u_i(x_i^k)$ is also maximal across all such allocations; it follows that, since each u_i is strictly increasing, $u_i(x_i^k) = u_j(x_j^k)$ for all $i,j \in N$.

We first argue that such a construction suffices to establish the result: Let φ be as above, and suppose by way of contradiction that there is a k and a feasible allocation $(y_1,\ldots,y_n)$ for which $\varphi(u_1(y_1),\ldots,u_n(y_n)) > \varphi(u_1(x_1^k),\ldots,u_n(x_n^k))$. Note then, by symmetry of φ, that for any permutation of the agents $\sigma : N \to N$,

[4] Symmetry means that if σ is a permutation on $\{1,\ldots,n\}$ then $\varphi(x_{\sigma(1)},\ldots,x_{\sigma(n)}) = \varphi(x_1,\ldots,x_n)$. Increasing here means that if $x_i > y_i$ for all i, then $\varphi(x_1,\ldots,x_n) > \varphi(y_1,\ldots,y_n)$.

$\varphi(u_{\sigma(1)}(y_{\sigma(1)}),\ldots,u_{\sigma(n)}(y_{\sigma(n)})) = \varphi(u_1(y_1),\ldots,u_n(y_n))$. Quasiconcavity of φ then implies that

$$\varphi\left(\sum_{i\in N}\frac{u_i(y_i)}{n},\ldots,\sum_{i\in N}\frac{u_i(y_i)}{n}\right) > \varphi(u_1(x_1^k),\ldots,u_n(x_n^k)).$$

By the strictly increasing property of φ, and using the fact that $u_i(x_i^k) = u_j(x_j^k)$ for all $i,j \in N$, this implies that

$$\sum_{i\in N}\frac{u_i(y_i)}{n} > \sum_{i\in N}\frac{u_i(x_i^k)}{n},$$

contradicting

$$\sum_{i\in N}u_i(x_i^k) \geq \sum_{i\in N}u_i(y_i)$$

for all feasible allocations $y_1,\ldots,y_n$.

We finish the proof by constructing, for each i, a strictly decreasing, continuous, and positive function f_i, with the property that if we set u_i to be the integral of f_i, then the profile of utility functions $(u_1,\ldots,u_n)$ works as required by the first part of the proof.

We proceed by induction. We ensure that, for each $i \in N$ and each k, the following are true:

I) $\int_0^{x_i^k} f_i(x)dx = \int_0^{x_j^k} f_j(x)dx$
II) $f_i(x_i^k) = f_j(x_j^k)$.

In the first place, for $k = 1$, we define for each agent j, $f_j(0) = +\infty$. The construction is done in a series of steps, labeled (I) to (VI).

I) For K, define $f_i(x_i^K) = 1$ for all $i \in N$;
II) for $x > x_i^K$, we define $f_i(x)$ to be any strictly decreasing function, taking values everywhere less than 1 and making f_i continuous.
III) We proceed by induction. Let $k > 1$ be arbitrary, and suppose that $f_i(x)$ has been defined for all $x \geq x_i^k$. We assume that for all $k' \geq k$, $f_i(x_i^{k'}) = f_j(x_j^{k'})$ and

$$\int_{x_i^k}^{x_i^K} f_i(x)dx = \int_{x_j^k}^{x_j^K} f_j(x)dx \text{ for all } i,j \in N.$$

Recall that we have $x_i^1 < x_i^2 < \ldots < x_i^K$. We choose a finite $f_j(x_j^{k-1})$ but we must choose it to be sufficiently large. Specifically, let z be large enough so that there is $\varepsilon > 0$ for which $z(x_j^k - x_j^{k-1}) - \varepsilon > \max_{i\in N} f_i(x_i^k)(x_i^k - x_i^{k-1}) + \varepsilon$ for all j. We can then set $f_j(x_j^{k-1}) = z$ for all j.

IV) Observe that, given $f_j(x_j^{k-1})$ and $f_j(x_j^k)$, for any $\varepsilon > 0$ and any

$$y \in \left(f_j(x_j^k)(x_j^k - x_j^{k-1}) + \varepsilon, f_j(x_j^{k-1})(x_j^k - x_j^{k-1}) - \varepsilon\right),$$

we may define f_j continuous and decreasing on $x \in (x_j^{k-1}, x_j^k)$ so that

$$\int_{x_j^{k-1}}^{x_j^k} f_j(x)dx = y.$$

This follows as we may choose the integral as close as possible to $f_j(x_j^{k-1})(x_j^k - x_j^{k-1})$ by taking a sequence of decreasing continuous functions approaching the constant value $f_j(x_j^{k-1})$ pointwise in (x_j^{k-1}, x_j^k); likewise we may choose the integral as close as possible to $f_j(x_j^k)(x_j^k - x_j^{k-1})$.

V) Complete $f_j(x)$ on $x \in (x_j^{k-1}, x_j^k)$ so that

$$\int_{x_j^{k-1}}^{x_j^k} f_j(x)dx$$

is equalized across all agents, by picking

$$y \in \bigcap_{i \in N} \left(f_i(x_i^k)(x_i^k - x_i^{k-1}) + \varepsilon, f_i(x_i^{k-1})(x_i^k - x_i^{k-1}) - \varepsilon \right)$$

and choosing $f_j(x)$ on $x \in (x_j^{k-1}, x_j^k)$ so that

$$\int_{x_j^{k-1}}^{x_j^k} f_j(x)dx = y.$$

VI) In the case of $k = 1$, we must also maintain that

$$\int_0^{x_j^1} f_j(x)dx < +\infty.$$

The functions f_j so constructed satisfy the conditions we ask for: that for all k, $f_j(x_j^k)$ is equalized across j, and

$$\int_0^{x_j^k} f_j(x)dx$$

is equalized across j. By setting

$$u_j(x) = \int_0^x f_j(x)dx,$$

we have the required u_j.

10.4 STABLE MATCHING THEORY

We now turn to stable matching theory. Stable matchings find very important normative applications in economics, but the theory provides a basic predictive framework as well. Many markets, such as labor markets and the marriage

"market," have two sets of agents who pair up and who may have preferences over who they form a pair with. We describe the basic notion of a stable matching, and carry out a simple revealed preference exercise.

Our version of the model assumes a set M of *types of men* and a set W of *types of women*. The sets M and W are finite and disjoint. We assume a number K_m of men of type m, and K_w of women of type w. The primitives of the model are then given by a tuple $\langle M, W, P, K\rangle$, in which M and W denote sets as before, $K = (K_i)_{i\in M\cup W}$ is a list of non-negative integers, and P is a *preference profile*: a list of preferences $>_m$ for every $m \in M$ and $>_w$ for every $w \in W$. Each $>_m$ is a linear order over W, and each $>_w$ is a linear order over M.[5]

The standard prediction concept, or theory of which matchings to expect, is the notion of stable matching. A *matching* is an $|M| \times |W|$ matrix $X = (x_{m,w})$ such that $x_{m,w} \in \mathbf{Z}_+$, $\sum_w x_{m,w} = K_m$ for all m, and $\sum_m x_{m,w} = K_w$ for all w. The number $x_{m,w}$ is the number of men of type m matched to women of type w.

Stability requires the definition of blocking. A pair $(m, w) \in M \times W$ is a *blocking pair for* X if there are m' and w' such that $m >_w m'$, $w >_m w'$, $x_{m,w'} > 0$, and $x_{m',w} > 0$. The matching X is *stable* if there are no blocking pairs for X. To keep the presentation simple, we ignore individual rationality and single (non-matched) agents.

A matching X is *stable-rationalizable* if there exists a preference profile $P = ((>_m)_{m\in M}, (>_w)_{w\in W})$ such that X is a stable matching in $\langle M, W, P, K\rangle$.

A (undirected) *graph* is a pair $G = (V, L)$, where V is a set and $L \subseteq V \times V$ is a non-reflexive and symmetric binary relation on V. Elements of V are referred to as *vertices* and elements of L as *edges*. A *path* in G is a sequence $p = \langle v_0, \ldots, v_N\rangle$ such that $(v_n, v_{n+1}) \in L$ for all $n \in \{0, \ldots, N-1\}$. We denote by $v \in p$ that v is a vertex in p. A path $\langle v_0, \ldots, v_N\rangle$ *connects* the vertices v_0 and v_N. A path $\langle v_0, \ldots, v_N\rangle$ is *minimal* if there is no proper subsequence of $\langle v_0, \ldots, v_N\rangle$ which also connects v_0 and v_N.

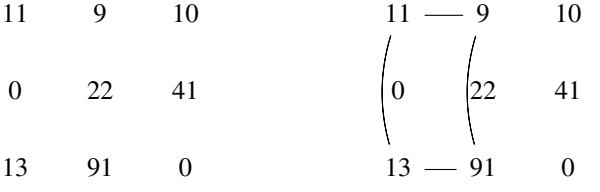
A *cycle* in G is a path $c = \langle v_0, \ldots, v_N\rangle$ with $v_0 = v_N$. A cycle is *minimal* if for any two vertices v_n and $v_{n'}$ in c, the paths in c from v_n to $v_{n'}$ and from $v_{n'}$ to v_n are distinct and minimal. If c and c' are two cycles, and there is a path from a vertex in c to a vertex in c', then we say that c and c' are *connected*.

For a matching X, we consider the graph defined by letting the vertices be all the nonzero elements of X; and by letting there be an edge between two vertices when they lie on the same row or column of X. Formally, to each matching X we associate a graph (V, L) defined as follows. The set of vertices V is $\{(m, w) : m \in M, w \in W \text{ such that } x_{m,w} > 0\}$, and an edge $((m, w), (m', w')) \in L$ is formed for every pair of vertices (m, w) and (m', w') with $m = m'$ or $w = w'$ (but not both).

[5] The most basic model assumes that there is only one agent of each type, but the revealed preference question is more interesting with many agents of each type.

Theorem 10.9 *A matching is stable-rationalizable iff its associated graph does not contain two connected distinct minimal cycles.*

We omit the proof of Theorem 10.9, but the intuition behind the necessity direction is simple, and worth conveying here. Consider for example the matching on the left below:



The matching has three types of men and three types of women. There are 11 men of type 1 matched to women of type 1, 9 men of type 1 are matched to women of type 2, and so on. The graph associated to this matching contains a cycle, shown on the right. In fact, it has more than one cycle, and they are connected, but focus now on the cycle depicted.

If we are to find rationalizing preferences, we need to decide how men of type 1 rank women of type 1 and 2. Say that women of type 2 are preferred to 1; we can denote this preference by orienting the horizontal edge on the graph as pointing from 11 to 9. Now consider the preferences of women of type 2. Is it possible for them to rank men of type 1 above men of type 3? The answer is negative, as there are type 2 women matched to type 3 men, and, simultaneously, type 1 men matched to type 2 women. If women of type 3 were to prefer type 1 men, then some of the latter would form blocking pairs with type 1 men who are matched to type 1 women. Thus the vertical edge from 9 to 91 must point down, *away* from the direction of the first edge, which points from 11 to 9. In fact, the important implication of stability is that for any two consecutive edges which form a right angle in the graph, one must be oriented to point away from the other.

The reasoning above extends to the horizontal edge between 91 and 13, and then to the vertical edge between 13 and 11. The former must point to the left, and the latter must then point up. As a result, a cycle must be oriented as a "flow." If we start, as we did, by having type 1 men prefer type 2 to type 1 women, then we obtain a clockwise flow. If we instead had started with the opposite preference, then we would have obtained a counterclockwise flow.

Now, we can consider a path leaving the cycle, for example the horizontal edge between 11 and 10 in the graph. An orientation of this edge amounts to specifying type 1 men's preferences between women of type 1 and 3. Recall that we have oriented the cycle in a clockwise fashion, so that the edge between 13 and 11 points up. Then the edge leaving the cycle, and going from 11 to 10, cannot be oriented to the left. The reason is that we would then have a blocking pair using some of the 10 men of type 1 who are matched with women of type

3, and some women of type 1 who are matched to men of type 3. The general principle is that any path leaving a cycle must point away from the cycle.

It should now be clear that we cannot have two cycles connected by a minimal path. Such a path would have to point away from both cycles. Then there would be two consecutive edges on the path, such that each edge point to their common vertex. This situation would imply the existence of some blocking pair, just as in the examples we discussed above.

10.5 CHAPTER REFERENCES

Persistence under expansion was introduced by Yanovskaya (1980), while Theorem 10.1 is a generalization of her result due to Sprumont (2000). Persistence under expansion and persistence under contraction are closely related to the consistency concepts discussed by Peleg and Tijs (1996). Persistence under expansion is clearly related to condition γ from social choice theory (see, e.g., Sen, 1971). The first part of persistence under contraction is similar to condition α for single-agent choice functions, while the second part is related to condition β. Theorems 10.2, 10.3, and 10.4 are also from Sprumont (2000). Ray and Zhou (2001) establishes related results for extensive-form games and subgame perfection. Xu and Zhou (2007), Bossert and Sprumont (2013), Rehbeck (2014), and Xiong (2013) study the extensive-form question when the game itself is not observable. Instead, the primitive is a classical choice function.

Theorem 10.5 and the notion of N-congruence are due to Galambos (2010). Galambos (2010) describes a more sophisticated version of Theorem 10.5 which appears in his dissertation. It need not rely on the assumption of a line-closed domain.

Theorem 10.6 is due to Lee (2011). Interchangeability is a well-known property of the Nash equilibria of zero-sum games. The contribution in Lee's work is to show that it is *all* that the property of zero-sum adds, from the revealed preference perspective.

Theorem 10.7 is due to Ledyard (1986). More complicated theorems appear there, dealing with certain restrictions on the forms of the u_i and p_i functions. For example, he obtains the necessary and sufficient conditions required for a direct revelation mechanism to be strongly rationalized by a Bayesian environment when the ordinal, but not cardinal, structure of each u_i is known conditional on each type profile t (as would be the case in standard private-valued single-dimensional consumption environments). Such characterizations are based on Lemma 1.12. However, a few important questions seem to remain unresolved. More importantly, the results described here consist of a single observation. It is not terribly surprising that anything is rationalizable in this case. For an analogy with choice theory, a choice function defined only on one budget is always rationalizable, unless there is sufficient structure on the class of rationalizing relations. An interesting idea would be

to understand what happens when multiple observations are possible, possibly when some parameters of the environment change, but others remain fixed. In fact, a simple method of doing this would be to consider strategy spaces S_i, one for each agent, and consider Bayesian strategies $\sigma_i : T_i \to S_i$. It is then meaningful to discuss notions of strict Bayesian Nash equilibrium, and by varying the spaces of strategies, one may come up with interesting testable implications.

Section 10.3 is based on Chambers and Echenique (2014b). The paper includes results for the cases where disagreement points may vary in an observable way (when disagreement points are unobservable the theories become non-testable). The proof of Theorem 10.8 and the observation that the same list of utility functions works to rationalize each environment, taken here from Chambers and Echenique (2014b), were suggested to us by an anonymous referee. For a continuous version of the problem, see Chiappori, Donni, and Komunjer (2012). We discuss other approaches in Chapter 12.

Theorem 10.9 and the discussion in Section 10.4 is taken from Echenique, Lee, Shum, and Yenmez (2013). That paper also includes results on which matchings are rationalizable as optimal for one side of the market, and rationalizable with transfers. These notions turn out to be observationally equivalent: a matching is rationalizable using monetary transfers iff it is rationalizable as stable and optimal for one side of the market (men or women).

We have assumed that observed matches consist of matrices of non-negative integers. We can instead suppose that multiple matching among the same individuals (or types of individuals) are observed. Then we can insist on rationalizing preferences that make all of them stable. This exercise is carried out in Echenique (2008). The same problem for a model with transfers is worked out in Chambers and Echenique (2014a).