Aggregation and Convergence in Experimental General Equilibrium Economies Constructed from Naturally Occurring Preferences

Sean Crockett Daniel Friedman Ryan Oprea*

December 1, 2017

Abstract

Prior laboratory experiments have studied general equilibrium economies constructed from "induced preferences" for artificial goods. We introduce new methods that allow us to study economies constructed instead from subjects' actual, "homegrown" preferences. Our subjects reveal their preferences by choosing portfolios of Arrow securities from budget lines through fixed endowments for a series of prices. We then construct several different economies by sorting subjects according to their revealed preferences. The constructed economies exhibit a wide range of predicted outcomes, where predictions are competitive general equilibria given the revealed preferences. Perhaps surprisingly, in every one of our markets the predicted excess demand is well-behaved, and avoids the pathologies highlighted in the Sonnenschein-Mantel-Debreu theorem. (The main reason seems to be heterogeneity in revealed preferences.) Actual trade in the constructed economies using a tatonnement market institution closely tracks predictions in most markets. The exceptions occur in economies with severe wealth effects that generate excess demands that are flat relative to measured preference volatility.

Keywords: Experimental Economics, General Equilibrium, Aggregation, Portfolio Choice, Heterogeneity, Risk Preferences, tatonnement.

JEL codes: C91, C92, D03, D51, G11

*Crockett: Baruch College, New York, NY 10010, sean.crockett@baruch.cuny.edu; Friedman: University of California, Santa Cruz, Santa Cruz, CA 95064, dan@ucsc.edu; Oprea: Economics Department, University of California, Santa Barbara, Santa Barbara, CA, 95064, roprea@gmail.com.

1 Introduction

By the 1970s it became clear that general equilibrium theory, at that time widely regarded as the centerpiece of economics, suffers from two fundamental problems. First, as the Sonnenschein-Mantel-Debreu theorem (SMD) [Sonnenschein, 1973, Mantel, 1974, Debreu, 1974] showed, even very standard neoclassical preferences can aggregate to produce “pathological” economies, for which competitive equilibrium is non-existent, or multiple, or dynamically unstable. Second, the theory itself tells us very little about how, when and why markets should converge to competitive equilibria even when they do exist. Jointly, these fundamental problems call into question the power of the theory to predict and interpret observed market behavior.

In this paper we introduce new experimental methods that shed new empirical light on those longstanding problems. Our contribution is to study laboratory economies that are built entirely from the “homegrown” preferences that subjects bring with them into the lab. In the first phase of our experiment, subjects partially reveal those preferences by choosing final allocations along dozens of exogenously given budget lines generated by a grid of prices and a fixed initial endowment of two goods (building on methods introduced in Choi, Fisman, Gale, and Kariv [2007]). Using those revealed preferences, we construct economies by sorting the subjects into several distinct groups, and for each economy we calculate aggregate excess demand at these grid prices. These calculations require no parametric assumptions but still produce “revealed competitive equilibria” (RCE) predicting behavior in each economy we construct. Using this approach, we can study for the first time whether real preferences, in the aggregate, give rise to the sort of pathologies cautioned by SMD. In the second phase of the experiment subjects trade under a tatonnement institution, allowing us to compare actual market prices and allocations to the RCE predictions. The sorting algorithms enable us to construct economies whose RCE predictions span a wide range of prices and final allocations, enabling demanding tests of general equilibrium theory.

The extant experimental literature on general equilibrium, by contrast, relies on *induced preferences.* In such experiments (discussed in more detail below) subjects are endowed with portfolios of fictional goods and given artificial preferences in the form of non-linear monetary reward functions over their final portfolio (i.e. the portfolio they hold after trading in a market). Subjects are allowed to trade goods using a market institution, and the payment functions and endowments generate equilibrium predictions and stability properties that can be compared to the outcomes and dynamics of this trade. While these methods have proven useful in answering important questions about the equilibrating tendencies of markets, they can tell us little about how real preferences aggregate, since the preferences in such experiments are artificial. They offer no insight into whether aggregates of actual preferences suffer from the sorts of pathologies highlighted in the SMD theorem, and whether real preferences expressed in markets are sufficiently stable for markets to converge to predictable equilibria.

Our new methods can be applied to any sort of preferences measurable in the lab. Here we apply them to a class of preferences that is both of central importance to general equilibrium theory and is simple to study in the lab: preferences over risky lotteries. In each period of Phase 1 of our experiment, we endow subjects with an initial portfolio of Arrow securities,[1] (x, y): half of subjects are randomly assigned as x-types, endowed only with security x throughout the experiment, and the other half are y-types, endowed only with security y. We then provide subjects with budget lines generated by their endowments and a variety of prices in a fixed but "scrambled" sequence, and ask subjects to choose final portfolios on each of these lines. In Phase 2, we use subjects' raw choices from Phase 1 to classify each of them as either a "high convexity" preference type (called Hx or Hy depending on their endowment type) or a "low convexity" preference type (Lx or Ly). We then construct a series of experimental economies based on permutations of these types: HxHy markets have only high convexity subjects, LxLy markets low convexity subjects and LxHy and HxLy have the indicated mixes. Finally, subjects trade securities in these economies using the classical tatonnement market institution.

Preferences measured in Phase 1 allow us to make distinct equilibrium predictions for each of these economies. Crucially, we do not need to impose any parametric assumptions on preferences or choose an error structure for estimation to do this. Indeed, the experiment is designed to allow us to make general equilibrium predictions using only subjects' raw Phase 1 choices. Treating security y as the numeraire, we can compute for each Phase 2 economy the excess demand for security x revealed at each price seen in Phase 1 simply by summing the Phase 1 demands net of endowment over the economy's participants. The roots of these revealed excess demand functions for Phase 2 economies are the "revealed competitive equilibrium" (RCE) predictions. Moreover, because we observe subjects' choices twice at each price in Phase 1, we can estimate volatility in each participant's revealed preferences, and use this measure to predict whether aggregate preferences are volatile enough to interfere with equilibrating dynamics. Thus we are able to classify each RCE prediction, *ex ante*, as "fragile" or "robust" depending on how volatile aggregate preferences are relative to the slope of excess demand outside of equilibrium.

[1] Arrow security i pays one unit of consumption in state i and zero in all other states, so that holding one unit of each security type results in risk-free consumption of one unit. In our experiment we have two equally probable states and an Arrow security for each of them.

By constructing economies based on revealed convexities in preferences, we are able to use subjects' own preferences as treatment instruments to test general equilibrium comparative statics predictions. First, given our asymmetric aggregate endowments (total endowed y is less than total endowed x), we predict that more convex HxHy economies will generate lower prices than less convex LxLy economies. Second, we predict that due to stronger wealth effects, the equilibrating tendencies will be weaker in HxHy economies and so RCE will be more fragile than in LxLy economies

Individual preferences measured in Phase 1 are similar to those from the most closely related prior work (e.g. Choi, Fisman, Gale, and Kariv [2007]): a minority of subjects exhibit demand consistent with risk neutral preferences, most subjects display demands broadly consistent with risk aversion, and a large number of them seem to use distinctly non-neoclassical heuristics. More importantly for our purposes, we can study what these preferences imply for general equilibrium outcomes. At the individual level, preferences seem to confirm the worst suspicions raised by the SMD result: randomly pairing subjects into replica economies, we observe pathologies such as multiple equilibria or non-existent equilibrium nearly half the time. Remarkably, however, these pathologies disappear as economies become even moderately larger. In the 12-person economies we actually construct and study in the experiment, equilibrium exists and is unique in every single case.

Market-level results in Phase 2 are broadly supportive of general equilibrium predictions, comparative statics and classical dynamics. Prices usually converge to the RCE, and price levels universally obey comparative statics in each session and partition of subjects. As predicted, HxHy prices are lower in each session than LxLy prices. Moreover, prices virtually always converge to RCE pre-classified as robust, but do so less often in economies pre-classified as fragile. Thus even failures to equilibrate are predictable based on measured characteristics of Phase 1 behavior.

We close the paper by exploring the disconnect between the often poorly behaved demands of our individual subjects and the relatively well-behaved aggregate excess demands in our experimental economies. Our data indicate that economies become better behaved as they grow larger due to *heterogeneity* in individual preferences. That conclusion is based on simulations in which we sample data from our individual subjects to construct thousands of counterfactual economies of various sizes. In "replica" economies (those consisting of only one sampled agent on each side of the market), where heterogeneity is low, SMD pathologies – non-existence, multiplicity, instability – occur 40% of the time. By contrast, in economies of the size studied in our experiment (12 subjects) where heterogeneity is significantly larger, pathologies occur only 15% of the time. When we reach economies the size of our whole dataset, equilibrium exists and is unique and well behaved.

The simulation results support the theoretical conjecture [e.g., see Grandmont, 1987, 1992, Hildenbrand, 1994, Giraud and Quah, 2003, Hildenbrand and Kneip, 2005] that preference heterogeneity in the population can protect economies against SMD-style pathologies. Our simulation exercise indicates that preferences in our subject population are indeed sufficiently heterogeneous to tame aggregate excess demands as economies grow large.

A second simulation exercise shows that aggregate volatility, which can weaken equilibration, diminishes as markets grow large. Equilibria are fragile in 70% of replica economies, but only in 40% of 12 person economies and 12% of 60 person economies. Overall, due to heterogeneity and cancelling errors, economies are "well behaved" (have unique, robust equilibria) in only 1/5 of replica economies, half of 12 person economies and most 60 person economies. Thus, even though individual behavior is frequently far from neoclassical in our experiment, aggregation generally leads to remarkably neoclassical economies as the population grows large.

Our experiment relates to several existing literatures. The experimental general equilibrium literature studies the mapping between *induced* individual preferences and aggregate market outcomes, focusing on the predictive power of equilibrium theory. That literature began with partial equilibrium experiments (e.g. Chamberlin [1948] and Smith [1962]) and later extended to general equilibrium (e.g. Williams, Smith, Ledyard, and Gjerstad [2000]). An influential strand of this literature studies economies with strong wealth effects in order to understand the dynamic laws governing price adjustment (e.g., Plott and George [1992], Plott [2000], Anderson, Plott, Shimomura, and Granat [2004], Hirota, Hsu, Plott, and Rogers [2005], Crockett, Oprea, and Plott [2011], Goeree and Lindsay [2016]). A closely related strand proposes and tests models of adjustment dynamics (e.g., Friedman [1984, 1991], Easley and Ledyard [1993], Gode and Sunder [1993], Cason and Friedman [1997], Gjerstad and Dickhaut [1998], Asparouhova, Bossaerts, and Ledyard [2011], Gjerstad [2013]). Our contribution to this literature is to replace artificial induced preferences with naturally occurring preferences, enabling us to pose new questions about the empirical properties of aggregate excess demand and the impact of preference instabilities on adjustment dynamics.

A second related literature is a series of papers in experimental finance in which, as in our experiment, subjects' valuations for risky assets are shaped by their homegrown preferences. A prominent strand of this literature tests broad predictions of CAPM (e.g., portfolio separation and mean-variance efficiency) by having subjects trade two risky assets together with a third riskless asset [e.g., Levy, 1997, Asparouhova, Bossaerts, and Plott, 2003, Bossaerts and Plott, 2004, Bossaerts, Plott, and Zame, 2007]. Biais, Mariotti, Moinas, and Pouget [2017] compare two static procedures for eliciting demand for a single risky asset: a BDM mechanism (one price is randomly selected for payment) and a market clearing rule (the price selected for payment is the one that minimizes excess demand summed across participants). They find that the two elicitation procedures are not statistically different in aggregate and that the equilibrium prices implied by pooled choices respond to aggregate risk in the direction consistent with portfolio theory. Our experiment differs from papers in this literature in that we first elicit preferences and use these to construct specific economies with a range of price and allocation predictions and then test the predictions against subsequent market behavior.

Experimental finance includes many other strands more distantly related to our paper, for example a large strand studying asset bubble formation in long-lived assets (e.g. Smith, Suchanek, and Williams [1988], Asparouhova, Bossaerts, Roy, and Zame [2016]). Breaban and Noussair [2015] and Crockett, Duffy, and Izhakian [2017] study general equilibrium economies and use separately elicited risk parameters to help explain when and how bubbles are formed.

There is also a vast literature on risk preference elicitation; see Cox and Harrison [2008], Charness, Gneezy, and Imas [2013], and Friedman, Isaac, James, and Sunder [2014] for contrasting overviews. One strand of that literature helped shape Phase 1 of our experiment. The subjects in Choi, Fisman, Gale, and Kariv [2007], like ours, chose portfolios of two Arrow securities on a sequence of budget lines. Our setup is different mainly in providing subjects with explicit and constant endowments. Hammond and Traub [2012] extend Choi et al. by allowing early budget line choices to impact the characteristics of future budget lines in order to sharpen tests of the generalized axiom of revealed preference (GARP). Halevy et al. [forthcoming] provide a design evaluating how preferences recovered from observed choices make out-of-sample predictions.[2] Our experiment uses budget line elicitation to make predictions rather than to test revealed preference axioms or to recover structural preferences.[3] We therefore see our findings as complementary, by showing that preferences revealed in individual choice can have strong predictive power in settings such as markets.

The remainder of the paper is organized as follows. Section 2 presents our experimental design and Section 3 develops theoretical predictions and hypotheses for the experiment. Section 4 presents

[2]A related strand of literature examines choices from budget lines not tied to explicit risky assets; see for example Andreoni and Miller [2002], Fisman, Kariv, and Markovits [2007] (other-regarding preferences) and Ahn, Choi, Gale, and Kariv [2007] (ambiguity).

[3]Indeed, our design does not allow us to study GARP violations because subjects make decisions on non-crossing budget lines. This is a direct consequence of the fact that subjects have a fixed endowment in all of the choices they make in the experiment.

the main results and Section 5 reports simulation exercises to better understand how aggregation repairs pathologies in excess demands. We conclude in Section 6. Four Online Supplemental Appendices provide additional details on the design, theory and analysis.

2 Design

At each decision in our experiment, the subject

1. receives a fixed endowment $\omega = (\omega_x, \omega_y)$ of Arrow securities for two equally likely states, X and Y,
2. faces a price $p = \frac{p_x}{p_y}$ for security X (with Y as numeraire) that varies across decisions,
3. chooses a non-negative portfolio (x, y) on the budget line $px + y = p\omega_x + \omega_y$, and
4. earns x (resp. y) tokens if the decision is paid and state X (resp. Y) occurs.

At the beginning of each session, each of 28 subjects is randomly and permanently assigned to one of two "endowment types." The 14 x-type subjects enter every decision node with allocation $\omega = (100, 0)$, and the 14 y-type subjects begin with allocation $\omega = (0, 76)$.

These endowments are represented by a red dot at one end of their budget line as in Figure 1, which shows one decision problem for a y-type subject. The subject can click different points on the budget line to see text boxes displaying the corresponding (x, y) values; the last point clicked becomes her actual choice when she clicks the Confirm bar (or when a countdown clock expires); subjects who never click simply receive their endowment.[4] The realized state is determined only at the end of the session.

The experiment is divided into two Phases. In Phase 1 (periods 1-50) each subject makes a series of 50 individual choices at exogenously given prices, one choice per period. In Phase 2 (periods 51 and 52), subjects are partitioned into markets and make choices at endogenously and dynamically determined prices, many choices per period (we'll refer to each choice in a given Phase 2 period as a "round"). As explained in more detail below, subjects are paid for only one choice, determined at random at the end of the experiment, with each phase being equally likely to determine payment.

[4] The clock expired at 120 seconds for choices 1-2 and the expiration time gradually decreased down to 40 seconds for the 23rd choice and beyond. Whenever all subjects confirmed their choices before expiration, the next choice was presented immediately.

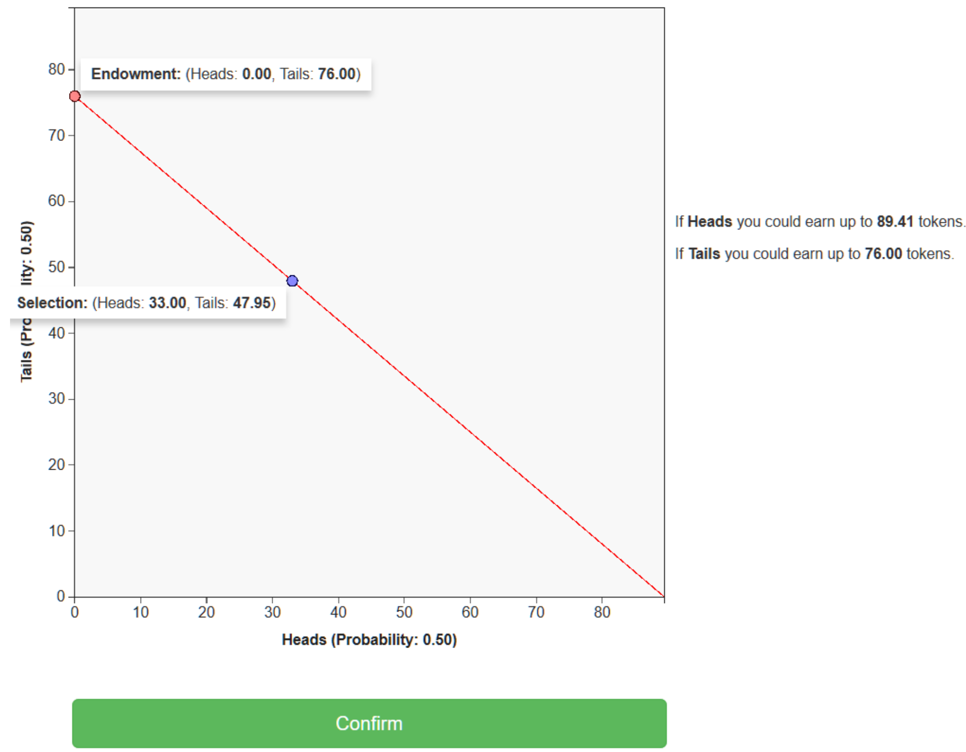


Figure 1: User interface. States X and Y are referred to as Heads and Tails respectively. Endowment here is $\omega = (0, 76)$, price is $p = 0.85$ and intercept values are shown in bold on the right. User clicks any point (x, y) on budget line, e.g., (33.00, 47.96), to check contingent payoffs, and clicks the green Confirm bar to make it her final choice.

2.1 Phase 1: Revealed Preferences

Phase 1 (periods 1-50) elicits subjects' individual preferences over portfolios. During the first 25 periods (referred to below as Phase 1P, for "practice"), each subject chooses portfolios at an exogenous sequence of budget lines that pass through her fixed endowment point with -slopes (or prices) p that jump erratically between 0.2 and 5.0. The actual price sequence is always (1.06, 0.36, 1.31, 0.64, 5.00, 2.00, 0.70, 1.43, 2.81, 0.57, 0.83, 0.28, 2.33, 1.00, 0.43, 0.89, 1.57, 1.21, 0.94, 0.50, 3.57, 0.76, 0.20, 1.75, 1.13). In the next 25 periods, called Phase 1R for "revealed," each subject again faces the same price sequence.[5]

After period 50, the computer assigns each subject a "preference type" based on their Phase 1R choices and uses these types to assign subjects to markets.[6] The computer algorithm (detailed

[5]Prices are not shown explicitly and nothing special happens between periods 25 and 26. Thus, given the erratic price sequence, we do not expect subjects to realize that the sequence is repeated.

[6]We chose *ex ante* to treat Phase 1R choices as individuals' revealed preferences. Phase 1P was designed to give

in Appendix A) chooses the 2 most erratic subjects of each endowment type and assigns them as preference type E. Remaining subjects are classified as highly convex (H) preference types or less convex (L) preference types according to the non-parametric measure S_x described in Section 3.1 below. Preference type Hx consists of the 6 remaining x-type subjects with the most convex preferences (the lowest values of measure S_x), and type Lx consists of the other 6 x-type subjects. Types Hy and Ly similarly consist of the 6 subjects with the most and 6 with the least convex preferences among the 12 remaining subjects with endowment (0, 76). Subjects are never told their preference type (or even that the software is classifying them).

2.2 Phase 2: Market Exchange

Preferences elicited in Phase 1 and the sorting scheme enable us to construct four distinct laboratory economies in each session. First (in period 51) the software constructs two Phase 2M (M for main) economies: HxHy economies match the 6 Hx (most convex x-endowed) subjects with the 6 Hy (most convex y-endowed) subjects to trade in one economy, and simultaneously LxLy economies match the 6 subjects of each endowment type classified as least convex.[7] Thus there are 12 subjects in each economy.

Next (in period 52) the software re-sorts subjects into two Phase 1S (S for "shuffle") economies, HxLy and LxHy. As the labels suggest, HxLy matches the 6 revealed most convex potential suppliers of x with the 6 least convex buyers of x, and LxHy does the reverse. The two Phase 1M economies together with the two Phase 2S economies ensure that each subject engages in market interaction with each subject of the opposite endowment type.

In each Phase 2 economy, subjects trade using a version of the venerable Tatonnement adjustment process [Walras, 1877]. Under this institution, the market sets an initial price and gradually adjusts it, round by round, according to market-wide expressed excess demand. Only when excess demand reaches zero are desired trades consummated, at the final price.

subjects practice with the interface before making their "real" choices but, as explained below, it also allows us to assess the stability of elicited preferences.

[7]Subjects know they are participating in one of two equally-sized markets, but do not know the sorting procedure. Each subject of preference type E is essentially a "non-voting" participant in one of the two markets: he sees the same budget lines and makes the same sort of choices as the other participants, but his choices (and endowments) are ignored in computing excess demand Z. By classifying a handful of subjects as E types and excluding them from markets, we hoped to remove subjects who were confused and therefore making random choices. As it turns out, most E types made fairly stable choices and this design decision likely had little effect on our results.

Our tatonnement implementation is similar to the one that Guler, Lugovskyy, Puzzello, and Tucker [2012] use in an induced value experiment. Each market begins in round 0 with the initial price $p_0 = 1.57$. In each round each subject selects an allocation along the corresponding budget line (e.g., in round 0 it is the line with slope = -1.57 and, depending on the subject's endowment type, intercepts either $x = 100$ or $y = 76$), using essentially the same interface as in Figure 1. Total demand D is the sum over the 12 market participants of the x-components selected in that round. Total supply S is the sum of endowed x over the participants, so it is always $S = 6(100)+6(0) = 600$. Excess demand is $Z = D - S$.

The tatonnement algorithm selects a price increment for the next round that is a monotone and sign-preserving function of Z. For example, if $Z < 0$ in round t (as we anticipate at $p_0 = 1.57$ given the relative scarcity of security y), then the price p_{t+1} in the next round is strictly less than the current price p_t. As detailed in Appendix A, price changes in radians are proportional to Z up to rounding, with the proportionality constant decreasing in later rounds. The adjustment process is deemed complete, and trades are consummated, when the absolute value of per capita excess demand $|Z|/12$, is less than 2.0 for three consecutive rounds or, if that criterion is never met, after the 25th round. Subjects receive their actual demands in the final round of the adjustment process, with the (typically very small) difference between the sum of these demands and the aggregate endowment being supplied or consumed by the experimenter.

2.3 Implementation

At the conclusion of Phase 2 we pay subjects for their choice in one period determined randomly, as follows. Each subject draws one ball (with replacement) from a bingo cage filled with one hundred balls numbered 1-100. A draw between 1 and 50 means payment for chosen allocation in the Phase 1 period with that number. A draw between 51 and 75 means payment for post-trade allocation in the final round in Phase 2M, while a draw between 76 and 100 means payment for the post-trade allocation in Phase 2S. Thus Phases 1 and 2 have an equal probability of being selected for payment, as does each period within a given phase. After determining the relevant allocation (x, y), the subject flips a coin. Heads (Tails) yields the payment $x/3$ ($y/3$), that is, one dollar for every three tokens allocated to security X (Y) in the relevant period. The state for each subject was determined by a coin flip.

We conducted four sessions between May and September 2016 at the Subotnick Financial Services Center Lab at CUNY Baruch, each with 28 subjects recruited via e-mail through the subject pool maintained by the Department of Economics and Finance. Including instructions, practice periods[8] and payments, each session lasted about 120 minutes. Including the \$7 showup fee, the average subject earned \$24.08 with standard deviation \$14.51.

3 Theory and Hypotheses

In this section we use general equilibrium theory and an extended example to motivate a set of testable hypotheses. Section 3.1 reviews some theory relevant to Phase 1 decisions and to the sorting procedure presented in section 2. Section 3.2 shows how Phase 1 raw decisions lead directly to general equilibrium predictions, without the need to estimate (or make any assumptions about) individual subjects' preferences. Section 3.3 reprises the challenges to general equilibrium predictions in the context of our experiment. Finally, section 3.4 distills the theoretical discussion into a set of empirical hypotheses.

3.1 Revealed Preferences

What restrictions on Phase 1 behavior does rationality impose? Suppose only that a subject is *minimally rational* in that she prefers more money to less in either state and has no intrinsic attachment to the x or y label; that is, suppose that she has continuous, monotone, symmetric preferences over bundles. The following Proposition shows that from a budget set B with $p < 1$, she will always select a bundle with more x than y.

Proposition 1. *Let a minimally rational agent demand $(x^*, y^*) \in B(p, \omega)$ given price $p > 0$ and endowment $\omega \in \mathbb{R}^2_+$. Then $(1 - p)(x^* - y^*) \geq 0$. The inequality is strict if preferences are strictly monotone and $p \neq 1$.*

See Appendix B for a proof. The geometric interpretation of the inequality is that demand (x^*, y^*) always lies in L^*, the portion of budget line between the point where it crosses the diagonal and its intercept with the axis for the less expensive security.

Since the choice opportunities (x, y) are bundles of Arrow securities for two complementary and equally likely states, the Phase 1 choices have structure beyond that imposed by general general equilibrium theory. For example, we can say (see Appendix B) that the points in L^* first

[8] There were five unpaid practice periods of the individual choice task prior to Phase 1 with price sequence (1.57, 0.64, 1.00, 0.36, 2.81).

order stochastically dominate all other points in the budget set. Standard expected utility theory[9] says that minimally rational preferences can be represented by a utility function U which is the expectation of some Bernoulli function u. That is, there is some continuous strictly increasing function $u : [0, \infty) \to [0, \infty)$ such that $U(x, y) = 0.5u(x) + 0.5u(y)$ represents the given preferences. Moreover, if $p \neq 1$ and u is linear or convex then the agent will always choose the corner allocation for the less expensive Arrow security, while if u is smooth and strictly concave then she will choose that same corner allocation or a point in the interior of L^*.

Returning now to minimally rational preferences (that are montone, symmetric and continuous), the next Proposition establishes that at any given price $p \neq 1$, the less substitutable are the goods, the more balanced is the chosen bundle, i.e., the closer it is to the diagonal. "Less substitutable" is formalized in Appendix B as "more convex than," a partial ordering over preferences that is defined via a uniform inequality on marginal rates of substitution.

Proposition 2. *For given prices $p > 0$ and endowment $\omega \in \mathbb{R}^2_+$, let $(x^*_U, y^*_U), (x^*_V, y^*_V) \in B(p, \omega)$ be demands for two different minimally rational agents whose preferences represented by utility functions $U(x, y)$ and $V(x, y)$ respectively, with V more convex than U. Then $(1-p)(x^*_U - x^*_V) \geq 0$. If $p \neq 1$, both optima are interior and the more-convex-than ordering is strict, then the inequality is strict.*

Again, see Appendix B for a proof. This proposition justifies our non-parametric measure of convexity S_x (used to categorize subjects as H or L preference types as described in Section 2.1), which is defined as follows. For any subject i, let S_x be the sum of the units of security x that she chooses at prices .70, .76 and .83 in Phase 1R.[10] Since $p < 1$, Proposition 1 says that every rational subject will select bundles with $x \geq y$ at these prices, so S_x will not be too small. More to the point, Proposition 2 says that rational subjects with more convex preferences will select a bundle closer to the diagonal, hence with less x, than those with less convex preferences. Thus, according to the two Propositions, S_x gives us the desired sorting among minimally rational subjects whose preferences can be ordered by convexity. Even for subjects who are not minimally rational, S_x is

[9] Prospect theory [Kahneman and Tversky, 1979] agrees if, as seems plausible, the reference point is 0. Allocations then are all in the gain domain where the value function reduces to an ordinary Bernoulli function. Prospect theory [original or cumulative, Tversky and Kahneman, 1992] also allows for a probability distortion function, but our experiment would seem to have no scope for that function since the probabilities for the two alternative states are both constant at 0.50.

[10] Why those prices? The budget line is identical for both endowments only at price 0.76, allowing direct comparability of all subjects. To smooth out idiosyncrasies, we a priori included the two adjacent prices where budget lines are not far from identical.

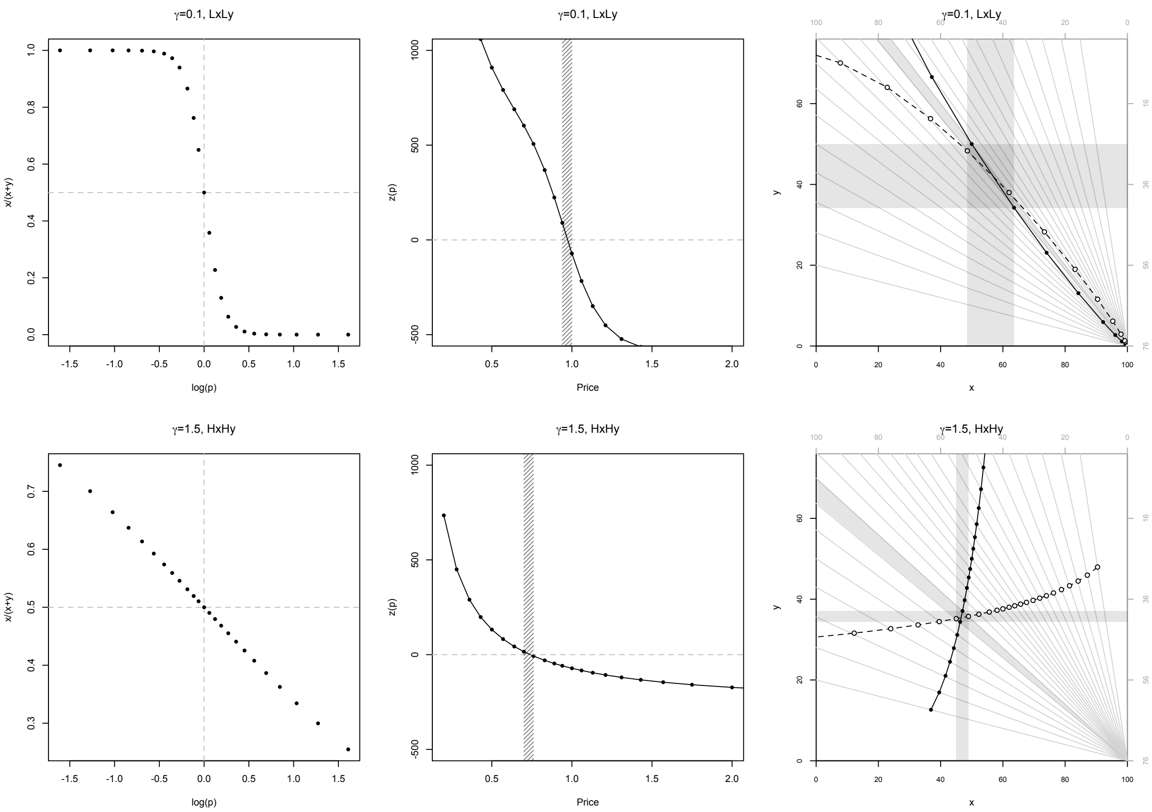


Figure 2: Choices and general equilibrium for two textbook Bernoulli functions. The first column plots the share of Arrow security X as a function of $\ln p$ for an agent with any positive endowment who maximizes the expectation of $u(c|\gamma) = \frac{1}{1-\gamma} c^{1-\gamma}$ for $\gamma = 0.1$ (top row) and for $\gamma = 1.5$ (bottom row). The second column shows excess demand Z for an economy consisting of two such agents, a y-type with endowment (0, 76) and an x-type with endowment (100, 0). The third column shows the corresponding offer curves in an Edgeworth box with the x-type's endowment in the lower right corner. Grid prices are rays emanating from that point; the x-type's offer curve (dashed lines) and the y-type's offer curve (solid lines) intersect at the competitive equilibrium allocation.

well defined and points in the right direction.

For the special case of preferences given by expected utility, Corollary 1 of Appendix B shows that more concave Bernoulli functions imply more convex preferences. Thus when conventional risk aversion measures apply, S_x sorts subjects by risk aversion: all subjects in a session classified as Hx will be more risk averse than any subject classified as Lx, and similarly for Hy and Ly.

A parametric example may help make the point. Suppose that agents have CRRA Bernoulli functions $u(c) = \frac{1}{1-\gamma} c^{1-\gamma}$ (and $= \ln c$ when $\gamma = 1$), for $\gamma > 0$. Agents solve $\max[.5u(x) + .5u(y)]$ s.t. $y + px = \omega_y + p\omega_x$, resulting in demand $x^* = \frac{\omega_y + p\omega_x}{p + p^{\frac{1}{\gamma}}}$. It is easy to verify that x^* is indeed a decreasing function of the risk aversion parameter γ when $p < 1$.

Figure 2 uses such textbook agents (with $\gamma = 0.1$ and $\gamma = 1.5$) to illustrate some consequences of our experimental design. The first column of the Figure shows how these agents would behave in Phase 1R, by plotting the share $\frac{x}{x+y}$ of x chosen as a function of $\ln(p)$. Minimally rational choices, consistent with Proposition 1, lie exclusively in the upper left and lower right quadrants in these plots. Likewise, the more risk averse agent (the agent with more convex preferences), with parameter $\gamma = 1.5$, demands less x than the $\gamma = 0.1$ agent at $p < 1$, consistent with Corollary 1.

3.2 General Equilibrium

The remaining columns of Figure 2 show how our sorting and Phase 2M procedures would play out if all subjects were CRRA agents, half of each endowment type with parameter $\gamma = 0.1$ and the other half with $\gamma = 1.5$. The middle column plots excess demand

$$Z(p) = \sum[x^* - \omega_x] = 6[\frac{76 + 100p}{p + p^{1/\gamma}} - 100]. \tag{1}$$

In the top row, the sorting algorithm puts all the $\gamma = 0.1$ agents, those with less risk aversion, into market LxLy. The competitive equilibrium (CE) price p^* is the unique solution to $Z(p) = 0$, so $p^* = (.76)^\gamma \approx 0.973$. Since p^* is not on our price grid, the Figure shows the CE price tunnel [0.94, 1.00], the interval spanning adjacent grid points where Z changes sign.

The last column of Figure 2 shows the corresponding Edgeworth boxes, including price rays (so the CE price "tunnel" is now a wedge), and the offer curves $\hat{x}(p) = \frac{w}{p+p^{1/\gamma}}$, $\hat{y}(p) = w - p\hat{x}(p)$, where w represents wealth and is equal to 76 for y-types (whose offer curve is represented by a dashed line) and $100p$ for x-types (solid line). The offer curves are close together in the top row, implying large (positive or negative) excess demand immediately outside of the CE price tunnel and thus squeezing the price towards its neutral value of 1.0. The allocation lies in the largest rectangle defined by the four x- and y-coordinates at which the offer curves enter or exit the price tunnel (if the tunnel spans prices p_i, p_{i+1}, then those coordinates are $(\hat{x}^k(p), \hat{y}^k(p)), k = S, D$ and $p = p_i, p_{i+1}$, where S and D reflect endowment types $(0, 76)$ and $(100, 0)$, respectively.

The calculations for the bottom row are similar. The agents here are more convex (more specifically, in this parametric example, they are more risk averse), with identical $\gamma = 1.5$. The corresponding CE price, the unique root of equation (1), is $p^* = (.76)^{1.5} \approx 0.663$. Thus the price tunnel is at [0.64, 0.70] for this textbook HxHy economy, far below the neutral price of 1.0. The offer curves intersect at a much wider angle, implying much narrower excess demand outside the CE price tunnel so the allocation box is tighter than in the LxLy economy. We observe dominant wealth effects for an agent's endowed good directly from these offer curves; the agent's quantity demanded of her own endowed good is *increasing* in that good's relative price.

Despite its very special parametric structure (agents are expected utility maximizers with particular Bernoulli functions), this example illustrates two general points. First, everything plotted in the right two columns of Figure 2 can be calculated using only the choices plotted in the left column. It follows that once we know Phase 1 choices we can make Phase 2 market predictions without making any parametric assumptions. Second, just as the upper row of panels in Figure 2 differs from the lower row, we can expect subjects sorted by S_x to produce aggregate excess demand with different locations and shapes, and therefore to produce different predicted market outcomes.

To work out those predictions, Appendix B formalizes a "price tunnel" as *Revealed Competitive Equilibrium* (RCE) $[p_i, p_{i+1}]$, an interval spanning adjacent grid prices where excess demand changes sign. Roughly speaking, an RCE is simply a general competitive equilibrium price revealed by individual subjects' choices in Phase 1R. We generally get a price interval instead of a point prediction due to the discreteness of the Phase 1 price grid.

Comparative statics predictions arise from sorting subjects according to the convexity indicator S_x, and applying the following

Proposition 3. *Suppose that Edgeworth box economy A is more convex than Edgeworth box economy B, and that aggregate x endowment exceeds aggregate y endowment in each economy. Then* $\min\{p \in RCE(A)\} \leq \min\{p \in RCE(B)\}$ *and* $\max\{p \in RCE(A)\} \leq \max\{p \in RCE(B)\}$.

See Appendix B for a proof, and a definition of "more convex than" economies. The idea is that there is some way to pair up agents in the two economies such that every A agent has more convex preferences than his B counterpart. The conclusion is that, in an appropriate sense, RCE prices are lower in Economy A.

3.3 Existence, Uniqueness, Stability and Convergence

We designed the experiment mindful of two kinds of potential problems that might prevent GE theory from usefully predicting prices and allocations in markets. The first kind of problem is suggested by the the Sonnenschein-Mantel-Debreu theorem. RCE fails to exist if revealed excess demand Z^R remains positive (resp. remains negative) over the entire price grid, implying that p^* is above the maximum price and possibly is infinite (resp. is below the minimum price and possibly is zero). Equally problematic, Z^R may change signs more than once, implying multiple competitive equilibria. These possibilities can not be ruled out even if individual revealed preferences are quite well behaved.[11] Even under vanilla preferences (e.g., arising from CRRA Bernoulli functions), RCE can fail to exist and, using two or more different preferences (e.g., γ parameters), economies with multiple equilibria can be constructed [see also Toda and Walsh, 2017]. A related problem is that the RCE may be tatonnement-unstable: classic theory [summarized e.g. in Arrow and Hahn, 1971] predicts that mechanisms governed by tatonnement dynamics (including the tatonnement institution used in our experiment) will be dynamically unstable at an *upcrossing* RCE, where $Z^R(p_i) < 0 < Z^R(p_{i+1})$. In the neighborhood of such equilibria tatonnement dynamics tend to amplify deviations – prices increase when above and decrease when below the price where $Z^R = 0$.

The SMD literature attributes these pathologies to "backward bending" excess demand due to "wealth effects" where non-substitutability is crucial. The intuition is as follows. You own a good X from which you derive a lot of your income. Suppose that the relative price of good X drops, so you now have less income. If goods are close substitutes, you simply keep more of the endowed good and buy less of everything else; no problem. But if you really like diversified bundles, then you still want to buy other goods W, Y etc, and that's going to cost you more of your X than before. So you may keep *less* X even though its price decreased. If your trading partners also regard X as not a good substitute for the other goods, then their demand for X may not increase enough to offset the reduction from people like you. Thus aggregate excess demand for X can, over some interval of prices, decrease as own price decreases, but for higher or lower prices outside this range have the usual slope. This sort of non-monotonicity ("backward bend") is the key ingredient for the most famous examples of multiple equilibria in the GE literature [e.g., Scarf, 1960, Gale, 1963, Shapley and Shubik, 1977].

In our experiment, some subjects may regard the two Arrow securities as not very substitutable, especially if they want to ensure that their wealth never falls below some threshold value. To the extent that they demand bundles near the diagonal [$x = y$] even at price ratios far from 1.0 (e.g., they have low values of S_x), our subjects will contribute to SMD pathologies. Indeed, as shown in Corollary 1, even a textbook EU-maximizer who is a highly risk averse will want to hold more nearly equal quantities of the two securities, and so will contribute to non-monotonic aggregate excess demand Z.

[11] Sonnenschein [1973] showed that in a two-good economy such as ours, even nicely behaved preferences can produce wild excess demand functions; Mantel [1974] and Debreu [1974] soon generalized that result to economies with more goods. Indeed, Mas-Colell [1977] showed that the set of general equilibrium prices is essentially arbitrary even if each consumer i has a nice utility function $U_i(x, y)$ representing continuous, monotone and strictly convex preferences.

We shall say that an economy $\mathcal{E}$ is *nice* if (1) an RCE exists and it is unique, and (2) it is tatonnement-stable. Following the SMD literature, we shall say that the economy is *pathological* if it is not nice, i.e., if it has non-existent RCE or non-unique RCE or unstable RCE. Our sorting procedure, then, segregates agents more likely to produce pathologies into the Hx and Hy groups, and agents more likely to facilitate nice economies into the Lx and Ly groups.

The second kind of potential problem arises from preference volatility. If individual subjects' choices in Phase 2 are sufficiently different from their choices in Phase 1R, then economies might fail to reach RCE benchmarks. Section 4.3 below will propose alternative empirical measures for the degree σ_R of volatility revealed by available data (the discrepancies between Phase 1R and Phase 1P choices).

To understand the potential problem, suppose that we have a nice economy whose unique RCE tunnel $[p_i, p_{i+1}]$ is not near the top of price range. Recall that the n price grid points are ordered from low to high, so suppose that $i < n - 3$. Niceness ensures that $Z^R(p_{i+3}) < 0$, but suppose that $|Z^R(p_{i+3})|$ is very small. Then even moderate preference fluctuations can cause the excess demand observed in the market phase to reverse sign at a price near p_{i+3}, causing the tatonnement algorithm to halt or reverse course instead of moving towards the RCE. Thus volatile preferences in Phase 2 can prevent convergence to RCE when Phase 1R excess demand is (in absolute value) small. We will classify RCE that suffer from this danger as *fragile* and RCE that do not as *robust*; again, see Appendix B for formalities.

Intuitively, an RCE is robust if the excess demand curve is sufficiently steep in its neighborhood, and is fragile if the excess demand curve is flat relative to price grid increments and volatility σ_R. For example, excess demand Z is steeper near RCE in the upper center panel of Figure 2 than in the lower center panel. We would therefore expect that fragility is more likely in the lower panel, which is associated with greater risk aversion and higher values of S_x.

3.4 Testable Hypotheses

These theoretical observations lead to a set of hypotheses concerning excess demands, RCE and robustness implied by Phase 1 decisions. They also generate a second set of hypotheses concerning the behavior of markets in Phase 2.

We begin with the first set. The SMD literature provides few reasons to expect excess demands to be consistently nice in the economies we construct. That concern is compounded by the far-from-neoclassical heuristics observed in prior experiments eliciting preferences over risky assets.[12]

Thus our sorting procedure should affect the frequency of pathology — economies with more H (high convexity) agents should be more susceptible. Of course, our non-parametric measure S_x of convexity looks only at a narrow range of prices, and problems might occur elsewhere. Nevertheless, we have the following plausible, a priori prediction.

Hypothesis 1.1: *Some of our economies will be pathological. HxHy economies, which involve the most revealed-convex agents, will be pathological more often than the other three economies, and LxLy economies will be pathological less often than the other three.*

Next, we expect volatility in individual preferences to be large enough relative to excess demand to sometimes produce nice but fragile RCE. Since subjects with greater convexity (or lower S_x) should be no less volatile than other subjects, we expect fragility/robustness to depend mainly on how steep the excess demand function is. Figure 2 illustrates that more nearly risk neutral agents tend to produce steeper excess demand curves in a neighborhood of RCE. We can show that this is true for economies consisting of agents with identical CRRA Bernoulli functions, and also true over a wide range of parameter values for identical CARA Bernoulli functions. More generally, for heterogenous rational agents, more convex preferences tend to be associated with flatter Z in a neighborhood of RCE,[13] but there are counterexamples.Based on this intuition, we pose the following pattern as an empirical hypothesis.

Hypothesis 1.2: *Preferences will be volatile enough so that not all RCE will be robust. HxHy markets will have the fewest robust RCE, and LxLy markets will have the most robust RCE.*

Finally, our sorting procedure should affect the RCE price itself, as suggested in the CRRA example in Figure 2. In our experiment the aggregate endowment of x exceeds that of y, and the

[12] Choi et al. report evidence of prevalent disappointment aversion, whereby subjects choose a perfectly hedged portfolio for some neighborhood about the risk-neutral price, then "jump" to far less risk-averse portfolios just outside of this neighborhood, in some cases all the way to a corner allocation. In an earlier working paper version of the paper they also identify several other simple heuristics which we will discuss in a later section. Asparouhova et al. assume subjects have mean-variance utility but make noisy choices, and for many subjects the estimated noise component of choices is at least as important as the variance (risk aversion) component. Biais et al. report that around 30% of portfolio choices observed in their experiment are first order stochastically dominated.

[13] Here is a very rough heuristic argument. $Z(\hat{p}) < 0$ is largely independent of convexity for a sufficiently high price $\hat{p}$, at or above our starting point $p = 1.57$, since most subjects will consume very little x at such prices. Let p^* be the lower edge of the RCE tunnel for the more convex economy. By Proposition 3, the RCE tunnel for the less convex economy occurs at prices above p^*. Therefore, the *average* slope $\overline{Z'}$ of excess demand between $\hat{p}$ and p^* is less for the more convex economy, and there would seem to be no general tendency for the ordering in a neighborhood of RCE to differ from the average ordering of slopes.

revealed convexity of every Hx subject exceeds that of every Lx subject, and likewise Hy subjects are more revealed-convex than Ly subjects. To the extent that "more revealed-convex as measured by S_x" implies truly "more convex than," Proposition 3 implies

Hypothesis 1.3: *In each session, RCE prices in the HxHy economy will be lower than those in the LxLy economy.*

Turning now to hypotheses concerning Phase 2 outcomes, our premise is that RCE will organize outcomes produced by markets. First, we hypothesize that RCE comparative statics will be matched in the data: when we sort subjects into markets A and B, we expect prices to be higher in market A if its RCE is higher.

Hypothesis 2.1. *Final prices observed in Phase 2 will have the same ordering in each session as the RCE predictions for that session.*

Hypothesis 2.1 is an ex post prediction (i.e., conditioned on Phase 1 behavior). We can strengthen it by combining it with Hypothesis 1.3 to obtain the following ex ante (i.e., conditioned only on the experimental design) prediction.

Hypothesis 2.2. *In each session, final prices observed in Phase 2 will be lower in HxHy economies than in LxLy economies.*

Next, we predict that the accuracy of RCE quantitative predictions will depend on how volatile preferences are relative to the steepness of excess demand. Ex post, we expect our classifications of RCE as robust or fragile in each economy to predict how well markets converge:

Hypothesis 2.3: *The tatonnement dynamics will converge more quickly and reliably to robust RCE than to fragile RCE.*

Ex ante, we expect LxLy economies to have weaker wealth effects and steeper excess demands. Assuming no systematic difference in volatility, it follows that they will have more robust equilibria than HxHy economies. Thus

Hypothesis 2.4. *Final prices will be closer to RCE in LxLy than in HxHy markets.*

4 Results

In this section we report our empirical results. Section 4.1 characterizes the individual preferences revealed in Phase 1. Section 4.2 goes on to show that, although individual preferences are consistent with SMD-type pathologies in simple replica economies, all the actual 12-person economies in our experiment turn out to be nice. Section 4.3 assesses volatility of preferences, and uses volatility measured in Phase 1 choices to classify RCE as fragile or robust. Finally, Section 4.4 presents results from Phase 2 markets. Comparative statics predictions always hold and we get convergence to robust RCE but not always to fragile RCE.

4.1 Overview of Individual Choices

Behavior is diverse in Phase 1, but it can be classified into three broad types, as illustrated in the first row of Figure 3. Following Choi et al., the figure plots the x-share, $x/(x+y)$, of a subject's 25 choices on the vertical axis against log price, $\ln(p)$, on the horizontal axis. We classify a subject as Dominated if, as in Panel (c), she chooses more than 1/3 of the time a dominated allocation (recall Lemma 1), e.g., $y > x$ when y is more expensive so $\ln(p) < 0$. Only 5 of our 112 subjects were so classified (and fewer than 10% of choices were dominated overall in the data) suggesting that the vast majority of subjects understand the revealed preference task.

We classify a subject as Risk Neutral, as in panel (b), if (with no more than one exception) she invests entirely in the less expensive Arrow security. About 7% of our 112 subjects are so classified. The remaining subjects (almost 90%) are classified as Risk Averse.

Some of those subjects make choices consistent (up to a bit of noise) with maximizing a concave Bernoulli function, as in panel (a). However, as in Choi et al. we find widespread use of the three heuristics illustrated in the middle panels of Figure 3. We'll focus on identifying subjects who employed these heuristics more than one-third of the time in 1R.

The "Hedging" heuristic is to choose an allocation on the diagonal $x = y$ for a range of prices around $p = 1$. Panel (d) shows an extreme example consistent with infinite risk aversion; more commonly, outside of a smaller range, hedgers move towards (or even jump directly to) the less expensive corner. As noted in Lemma 2, that sort of pattern is not consistent with maximizing the expected value of a Bernoulli function, but it is consistent with maximizing a utility function $U(x, y)$ with a kink on the diagonal, as in Disappointment Aversion [Gul, 1991]. Just over 10% of subjects make hedging-consistent decisions for more than 1/3 of prices.

Panel (e) shows a weaker form of hedging that we call a "Floor" strategy. For a range of prices, the subject purchases some fixed minimum quantity (here about 30) of the more expensive Arrow security and spends the rest of her wealth on the cheaper security. Nearly 30% of our subjects set a floor that they enforce more than 1/3 of the time. Finally, panel (f) illustrates a heuristic we call "Ceiling," adopted by 6% our subjects: over a range of prices, pick a fixed maximum amount of

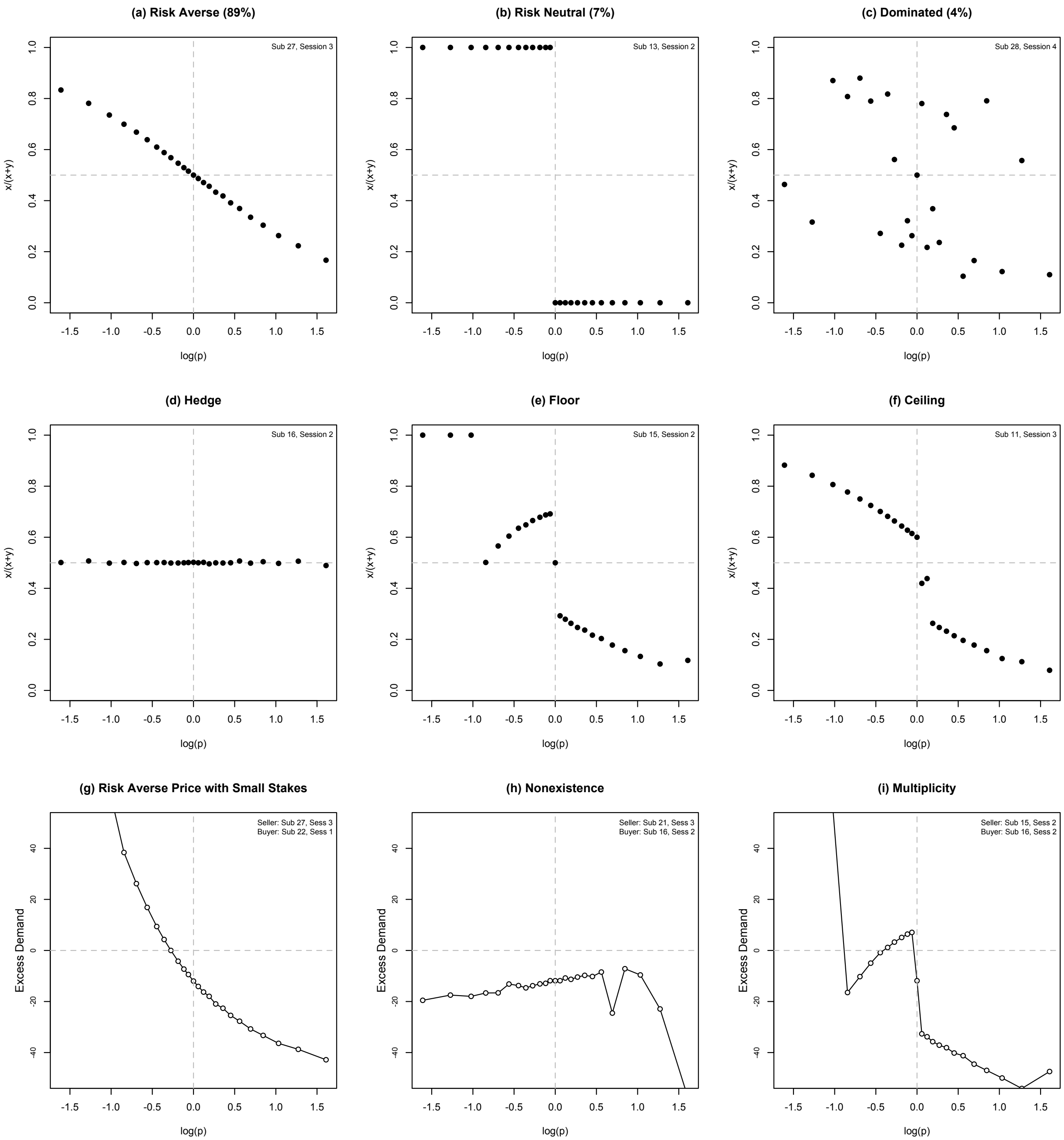


Figure 3: Phase 1R choices for six selected subjects and excess demands for pairs of subjects. Classifications (and sample share in %) are noted at the top of panels (a-c). Bottom row (g-i) plots replica economy excess demands for selected pairs of subjects.

the cheaper security and spend the rest on the more expensive security.[14] Appendix C shows the full set of choices from every subject in the dataset. Overall, choices from Phase 1 are similar to those observed in prior work.

4.2 Aggregate Excess Demands and Equilibrium Predictions

What sort of general equilibrium (RCE) predictions do we get when we aggregate these individual choices? As emphasized in section 3.2, we need not impose parametric assumptions about utility functions or error structures to answer this question. Our experiment is deliberately designed to allow us to sidestep such assumptions and to calculate aggregate excess demands directly at each of the 25 prices subjects face in Phase 1R simply by summing demands $(x - \omega_x)$ at that price across the subjects participating in that market.

The third row of Figure 3 illustrates the procedure using subjects pictured in the first two rows (and some similar subjects from elsewhere in the dataset). Panel (g) shows the excess demand function of the subject pictured in panel (a) matched with a y-type who made similar Phase 1 choices. Competitive equilibrium exists where excess demand crosses zero. The result is a nice empirical excess demand function that crosses zero only once and at a price below 1, showing this economy has a unique RCE that reflects the risk aversion of its members.

Panels (h) and (i) show that pathologies could easily arise from observed Phase 1 behavior. Panel (h) shows the y-type subject pictured in panel (d) matched with an x-type who follows a similar Hedge heuristic. The resulting excess demand function never crosses zero and so RCE fails to exist. Panel (i) matches the subjects pictured in panels (d) and (e) together, resulting in an excess demand function that crosses zero three times producing three RCE, one of which is dynamically unstable. The per-capita (x2) pictures are exactly the same if both participants are replicated m times, for an arbitrary positive integer m. In section 5 below, we will see that such pathological cases are quite common in representative agent, replica economies, occuring in nearly half of the possible pairings of our subjects. This would seem to support Hypothesis 2.1.

However, the economies we construct in our experiment (like naturally occurring economies), are not replica economies, with identical agents on each side of the market. Rather, they are built from 12 different human subjects with possibly very different preferences. Figure 4 plots excess demand functions for each session (rows 1,2,3 and 4) and economy (columns “LxLy”, “HxHy”,

[14] In a working paper version of Choi et al., the authors identify similar heuristics calling examples like panel (a) “S” (Smooth), examples like panel (d) “D” (Diagonal) and examples panel (e) a “secure level heuristic” or “B(w).”

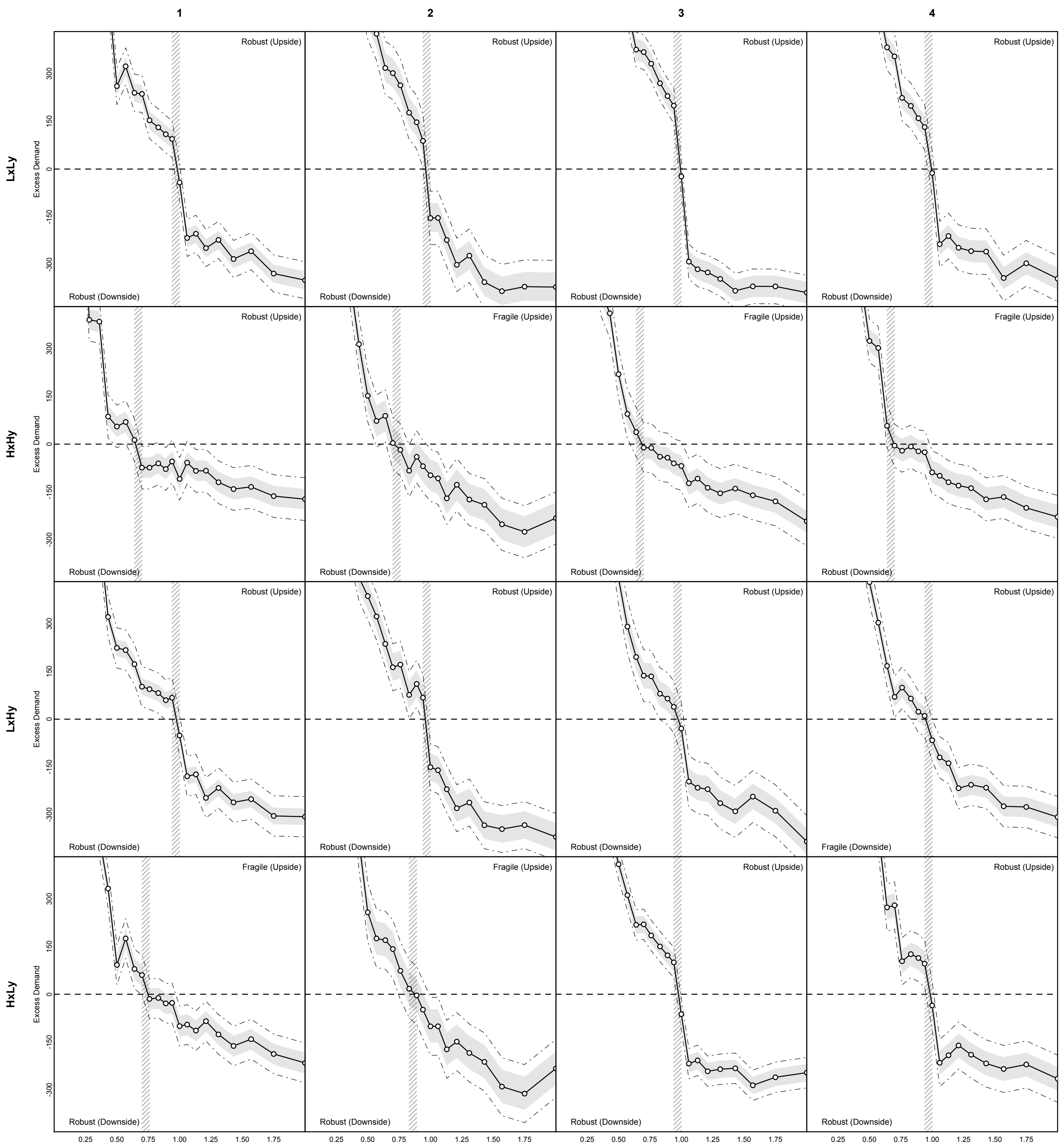


Figure 4: Excess demand functions for each economy in our dataset. Panels are organized into columns by session and row by economy type (marked on the outside of the grid). Bootstrapped interquartile bands are plotted around each function in gray, and standard deviation volatility bands as dashed-and-dotted lines.

"LxHy", "HxLy"). The results are striking: all 16 constructed economies are nice. Excess demand always crosses zero exactly once, with a negative slope. This is true even in "HxHy" economies where we deliberately sorted subjects to produce strong wealth effects and therefore vulnerability to SMD-style pathologies!

Result 1. *Contrary to Hypothesis 1.1, RCE exists, is unique, and is dynamically stable in all 16 of the economies we construct.*

The data thus allow us to reject Hypothesis 1.1. Intriguingly, this rejection is based not on the nature of individual preferences but rather on the size and structure of our experimental economies. We see frequent pathology in replica economies built using individual subjects' choices, but this seems to disappear with aggregation. Section 5 will investigate the role of aggregation and heterogeneity in attenuating pathology.

Figure 4 also plots RCE price tunnels as vertical hashed rectangles and shows evidence in support of Hypothesis 1.2. Comparing the first two rows for each column in Figure 4 reveals that, in each session, RCE occurs at much lower prices in HxHy (high risk aversion) economies than corresponding LxLy (low risk aversion) economies. While these differences in RCE were a deliberate aim of the design, we also have RCE price differences occurring by chance in the Phase 2S economies (HxLy and LxHy): RCE prices are higher in LxHy than HxLy economies in sessions 1 and 2. Thus Phase 1 choices and our sorting procedure provide a number of RCE comparative statics price predictions to test in our Phase 2 markets.

Result 2. *Revealed competitive equilibrium prices are lower, for each cohort, in HxHy economies than in LxLy economies, supporting Hypothesis 1.3. Moreover, revealed competitive equilibrium prices are higher in LxHy than HxLy in sessions 1 and 2.*

4.3 Preference Volatility and RCE Robustness

Preference volatility in our experiment is estimated by comparing a subject's Phase 1R choices to her choices in Phase 1P for exactly the same 25 budget lines. For each subject h and grid price p_i we take the discrepancy in the chosen x-component $d^{hi} = x^{hi}_{1R} - x^{hi}_{1P}$ and compile them in the set $D_{\mathcal{E}} = \{d^{hi} : h \in \mathcal{E}, i = 1, ..., 25\}$ for each of the four economies $\mathcal{E}$ = LxLy, HxHy, LxHy, HxLy in each session. Figure 5 (a) plots (in black) the cumulative distribution functions (CDFs) for each $\mathcal{E}$, combining the $D_{\mathcal{E}}$'s across sessions. For reference the panel includes (in gray) the same statistic for simulated agents making two iid uniform random choices on their 25 budget lines. The gray reference line indicates far more dispersion, while the black lines stay close to zero over most of

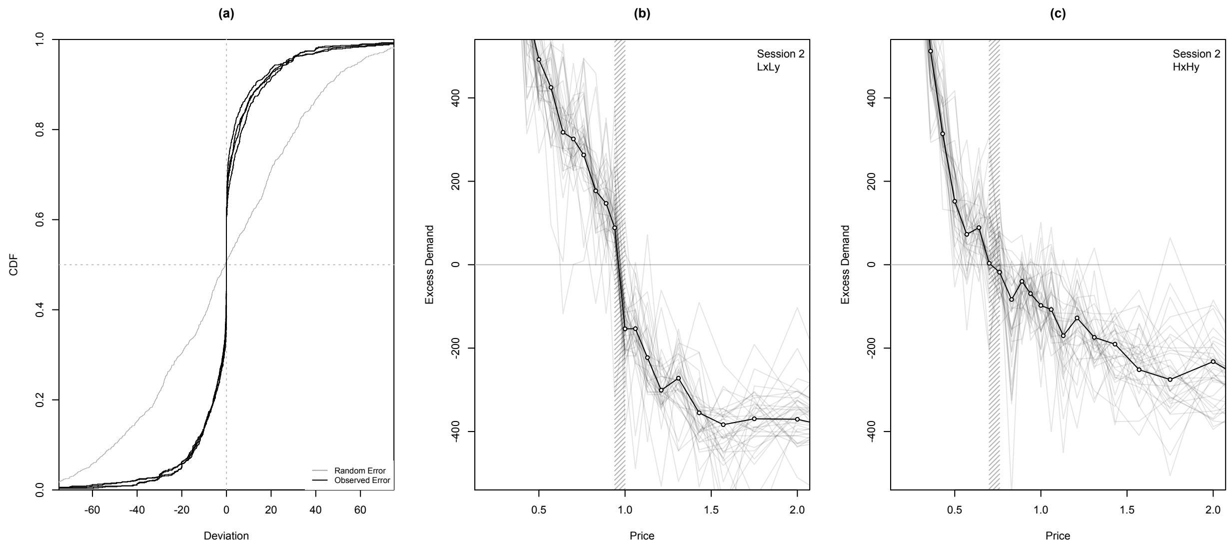


Figure 5: Choice discrepancies and sample excess demand functions. Panel (a) plots discrepancy CDFs for uniform random choices on the budget line (gray line) and for all subjects' observed discrepancies in all 4 economies (HxLy, etc, black lines). Panels (b) and (c) plot actual excess demand functions for the two extreme economies (LxLy and HxHy, in black) and for dozens of bootstrapped excess demands (in gray).

the range. This shows that our subjects' choices have considerable consistency, but the upper and lower 5% tails indicate occasional large discrepancies.

Is that volatility significant enough to interfere with convergence to RCE? To find out, we bootstrap thousands of excess demand curves as follows. For a given economy $\mathcal{E}$ in a given session and each grid price p_i, begin with aggregate excess demand $Z(p_i)$ and for each of the 12 active participants $h \in \mathcal{E}$, add or subtract (with equal probability) the absolute discrepancy $|d^{hj}|$ for a random grid point j drawn with replacement. The result is a single realization of bootstrapped demand, as illustrated by each of the dozens of gray lines plotted in panels (b) and (c) of Figure 5. In panel (c) these realizations frequently cross zero outside of the RCE tunnel, indicating a fragile RCE. By contrast, in panel (b), where the revealed excess demand is much steeper, the bootstrapped excess demands rarely cross zero outside of the RCE tunnel, indicating that the equilibria are robust.

That perspective leads to our main empirical definition of volatility as the interquartile range of the bootstrapped excess demands.[15] Our bootstrap procedure produces a symmetric distribution of aggregate discrepancies around zero, and the volatility measure σ_R for economy $\mathcal{E}$ is equal to

[15]The interquartile range is the most commonly used of the "robust measures of scale" as it is highly resistant to influence by outliers.

the 75^{th} (or, equivalently, the -25^{th}) percentile of the aggregate discrepancies for that economy (i.e., percentiles of the sum of 12 randomly signed draws from the appropriate distributions). The gray volatility bands around the excess demand functions plotted in Figure 4 are precisely that interquartile range.

Applying the earlier definitions, we classify the RCE as fragile if the volatility bands (or interquartile ranges) cross zero at grid prices outside of the RCE price tunnel and otherwise classify it as robust. Figure 4 labels an economy as "upside fragile" if these interquartile ranges overlap zero at grid prices above (to the right of) the RCE, and "upside robust" otherwise. Similarly, it is labelled "downside fragile" or "downside robust" (in the lower left corner of each panel) according to whether or not the interquartile ranges overlap at grid prices below RCE. Because the prices in our markets initiate at a price $p_0 = 1.57$ always above the RCE, we expect upside fragility or robustness to be more relevant for determining whether prices reach the RCE in our Phase 2 markets.

The Figure also shows a parametric alternative measure of volatility. Here we redefine σ_R for a given economy as the standard deviation of $D_{\mathcal{E}}$, and plot the corresponding volatility band edges as dashed-and-dotted lines. The standard deviation is less robust to outliers leading to a wider band, but fragility classifications derived from the non-parametric bootstrapping bounds and the parametric standard deviation bounds generally agree, especially for the more important upside fragility measure.

RCE in LxLy economies are universally robust − volatility bands never overlap zero outside of the RCE tunnel. By contrast in HxHy treatment where excess demand is flat, most economies (sessions 2, 3 and 4) have upside fragile equilibria, with volatility bands overlapping zero often at multiple prices. Results are intermediate in LxHy and HxLy, where only 1/4 of economies are upside fragile (sessions 1 and 2 of HxLy) and half are downside fragile. This pattern strongly supports Hypothesis 1.3.

Result 3. *Consistent with Hypothesis 1.3, most LxLy RCE are fragile and all HxHy RCE are robust, with an intermediate mix of fragility and robustness in LxHy/HxLy economies.*

Importantly, variation in fragility/robustness across economies is primarily driven by differences in the slope of excess demand functions rather than differences in the volatility of preferences. Indeed, as Figure 5 (a) shows, preferences in all four types of economies tend to be similarly noisy.[16]

[16] Indeed, comparing subject-wise standard deviations of the error using Wilcoxon tests, we cannot reject the hypothesis that subjects' preferences are equally stable in each pairwise comparison of economy-types at conventional levels of significance.

4.4 Market Outcomes

How well do the RCE and robustness/fragility classifications predict behavior in Phase 2 markets? Figure 6 plots a panel for each of our 16 economies that shows the predictions and actual market behavior. The left side of each panel shows *inverse* excess demand function and volatility band (derived from Phase 1 decisions), while the right plots round-by-round prices observed in Phase 2. The horizontal hashed rectangle extending across the panel represents the RCE price tunnel, and the fragility labels established earlier are shown at the top (bottom) of each panel.

First, we evaluate the comparative static hypotheses, 2.1 and 2.2. In every session, LxLy economies generate higher RCE price predictions than corresponding HxHy markets and in every case the final prices indeed are higher in LxLy, supporting Hypothesis 2.2. Moreover, in session 1 and 2, RCE predicted higher prices in LxHy than HxLy markets, and in each case we observe this ordering of final prices, supporting Hypothesis 2.1. A paired Wilcoxon test rejects the hypothesis that final prices are equal across economies in sessions and periods in which RCE differ ($p < 0.001$).

Result 4. *Consistent with Hypothesis 2.1, prices obey RCE comparative static predictions for every session and market period in our experiment. Consistent with Hypothesis 2.2, prices are universally lower in HxHy than LxLy economies.*

Next we evaluate the quantitative price predictions of the RCE, testing hypotheses 2.3 and 2.4. In 11 of our markets, RCE are upside robust and in every case prices reach the equilibrium bounds and converge to a point at or below the competitive equilibrium tunnels. In 10 markets, RCE are both upside and downside robust and in all but one (HxLy in session 3), prices not only reach the RCE but remain within its bounds. We conclude that when RCE are robust, markets reliably converge to the RCE.

In 5 of our markets, RCE are upside fragile and in only one of these do prices reach as far as the competitive equilibrium. In one additional case, the RCE is downside fragile and in this case prices reach the RCE but continue to fall below the RCE bounds, ending at a lower price. Thus, markets typically fail to converge to fragile RCE.

Overall, robustness and fragility classifications have strong predictive power on convergence to the RCE, supporting Hypothesis 2.3. A Chi-square test allows us to reject the hypothesis that prices are equally likely to converge in fragile and robust economies ($p = 0.016$).

The results also support Hypothesis 2.4 that our treatment interventions controlled the reliability of equilibrating dynamics: prices reach the RCE in all of our LxLy economies (our sorting algo-

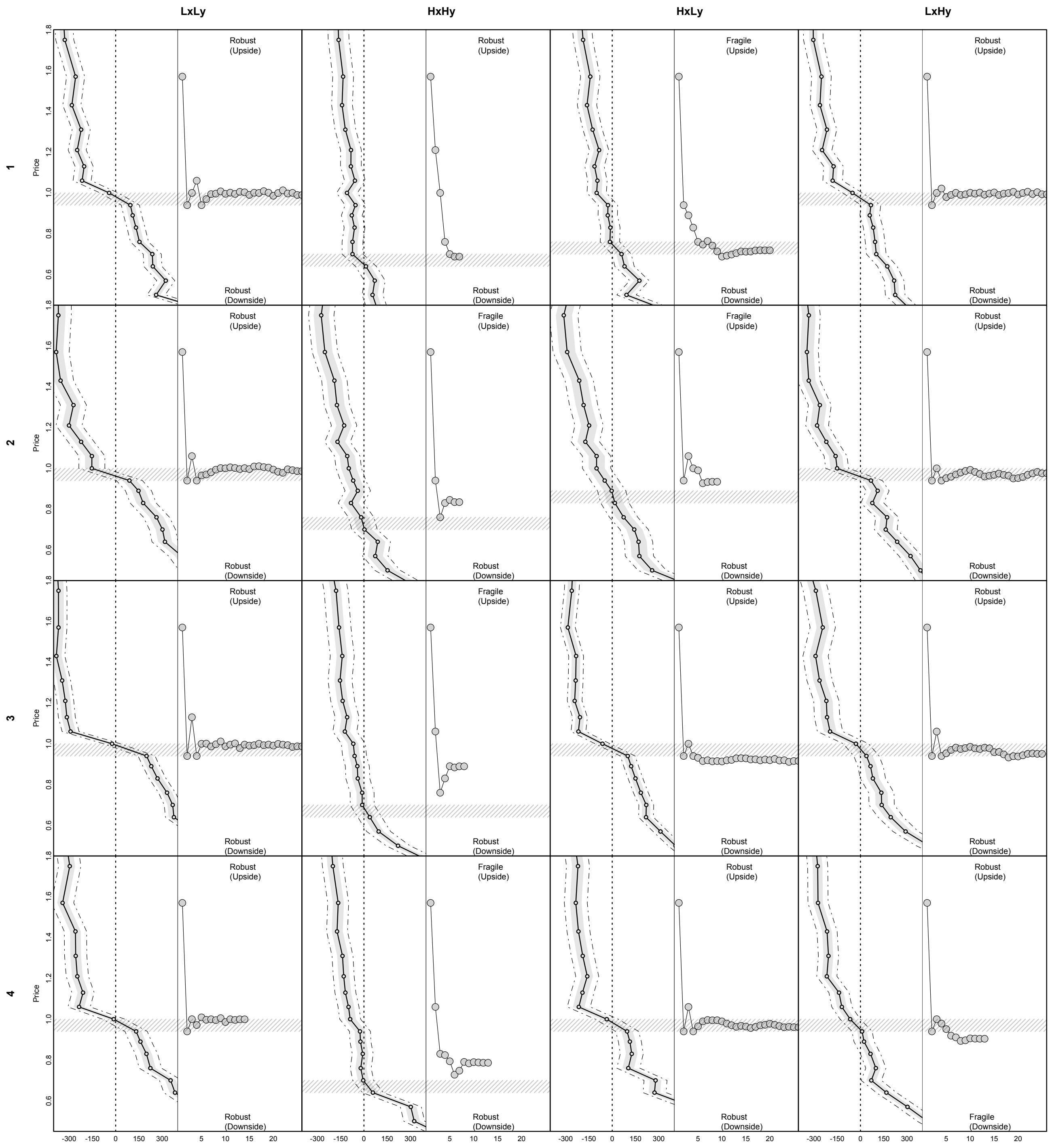


Figure 6: Inverse excess demand functions (left side of each panel) and time series of prices (right side of each panel) for each economy in the dataset. Panels are organized into columns by economy type and row by session (marked on the outside of the grid).

rithm is expected to give these the steepest excess demand functions), the majority of HxLy/LxHy economies, but only a minority of HxHy economies (expected to have the flattest excess demand functions).

Result 5. *As predicted by Hypothesis 2.3, markets reliably convergence to robust RCE, but usually fail to converge to fragile RCE. Consistent with Hypothesis 2.2, convergence occurs with greater reliability in LxLy than HxHy economies.*

Finally, we compare final allocations in Phase 2 markets to RCE predictions. Figure 6 plots Edgeworth boxes for the mean subject in each economy. In each case, boxes are drawn from the perspective of x-endowed sellers, with units of x on the x-axis and y on the y-axis. Numbers at the top and right hand side of each box show corresponding units for y-endowed buyers.The initial allocation is in the lower right hand corner of each box, and from this allocation we draw rays representing the price grid and plot offer curves generated by Phase 1 decisions. x-endowed sellers' offer curves are solid lines and y-endowed buyers are dashed lines. Where the offer curves overlap we plot a shaded wedge that visualizes the RCE price bounds. Finally, we include shaded vertical and horizontal boxes visualizing the RCE allocation bounds, as described in section 3.2, with double shading in regions consistent with RCE for both x and y.

In each Edgeworth box we include a red "X" visualizing the final allocations achieved by the corresponding market and a red dashed line corresponding to the final market price. The results show strong evidence of convergence, with allocations lying very close to the intersection of the offer curves in most economies. In all, 75% of economies generate allocations inside the bounds of the RCE plotted in the Figure, with the exceptions occurring only in economies with fragile RCE. 100% of economies with robust RCE converge within the RCE allocation bounds plotted in the Figure, providing additional support for Hypotheses 2.3 and 2.4.

Result 6. *In accord with Hypothesis 2.3, allocations tend to converge to RCE-predicted allocations in the Edgeworth Box, and universally converge in economies with robust RCE. Consistent with Hypothesis 2.4, allocations converge more reliably in LxLy than HxHy economies.*

5 Aggregation, Heterogeneity and Fragility

A striking conclusion from Section 4 is that, although replica economies built using individual preferences are often pathological, that poor behavior disappears in each of the 12-subject aggregate economies constructed in our experiment. This is true even in HxHy economies, designed delib-

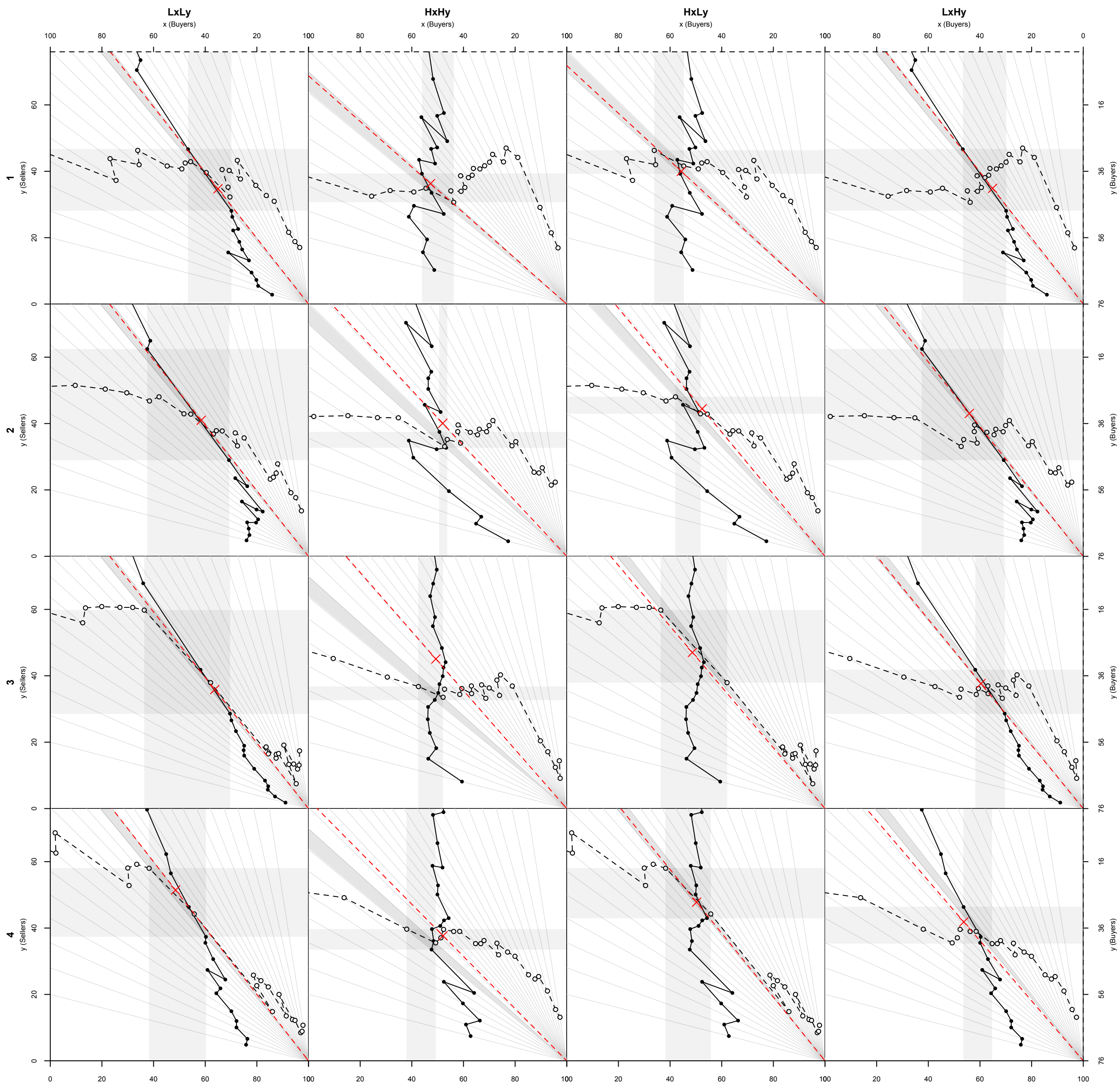


Figure 7: Empirical edgeworth boxes, plotting offer curves for the average subject of each type. Panels are organized into columns by economy type and row by session (marked on the outside of the grid). Gray rays mark the price grid, shaded wedges mark RCE price bounds and shaded gray rectangles RCE allocation bounds. Red dashed rays mark final prices from Phase 2 markets and red X's final allocations.

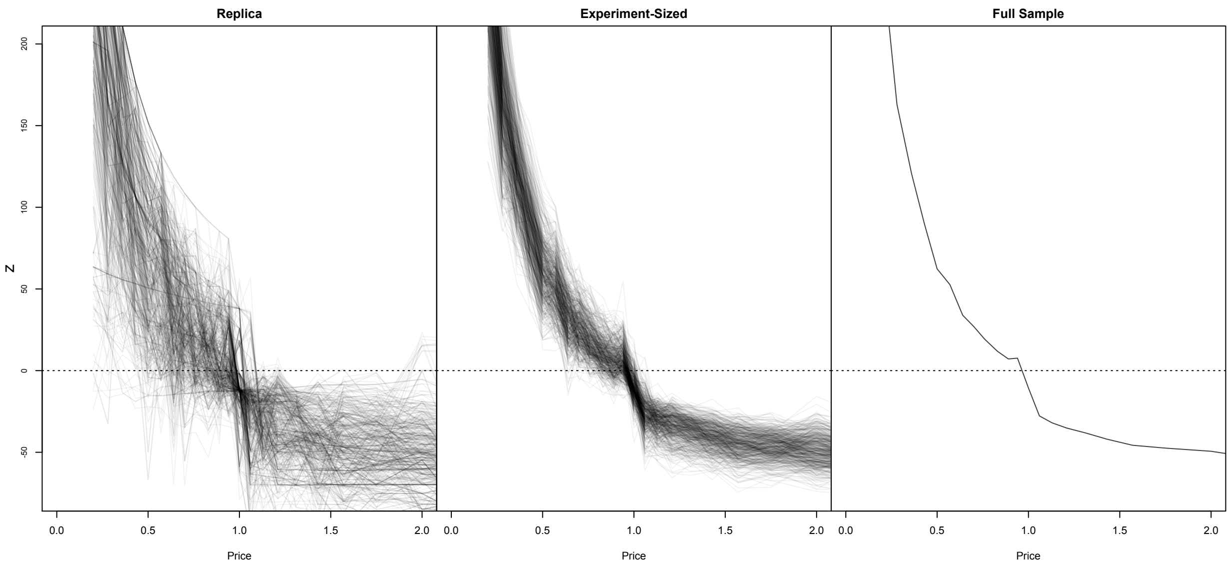


Figure 8: Counterfactual excess demand functions, sampled from subject cohorts of N=1 (Replica), N=6 (Experiment-Sized) and all subjects in the dataset (Full Sample).

erately to feature strong wealth effects and therefore particular vulnerability to SMD pathologies. The results thus provide a sort of empirical inversion of the Sonnenschein-Mantel-Debreu theorem: SMD shows that nice preferences can generate pathological excess demands in the aggregate, but in our data poorly behaved preferences (choices not rationalizable via neoclassical preferences) always aggregate into nice excess demands and rather orderly market outcomes.[17]

In this section we document this inversion more formally and provide some insight into its origins. We do so via a theoretically-motivated counterfactual exercise that constructs thousands of counterfactual economies from our data, as follows. For each counterfactual economy we sample with replacement N of the x-endowed sellers and N of the y-endowed buyers, and calculate average aggregate excess demands (dividing the total by N to retain comparability). We focus on $N = 1$ ("Replica" economies) and $N = 6$ ("Experiment-Sized", the size of the economies we study in our experiment), and also simply aggregate our entire dataset into one excess demand function ("Full Sample"). We plot a random sample of the resulting excess demand functions in Figure 8 to illustrate.

The figure shows that excess demands in REPLICA samples are quite heterogeneous and often quite poorly behaved. Conducting the exercise 10,000 times, we find that equilibria exist and are unique in only 60% of these counterfactual economies, and of the equilibria that exist 36% are

[17]Bossaerts et al. [2007] finds complementary results using a very different research strategy for CAPM markets.

unstable under classical tatonnement dynamics. Increasing the size of counterfactual economies to 12 agents (EXPERIMENT-SIZED), raises this to 85%, with 16% of equilibria being unstable. Finally, when we aggregate all of our subjects, excess demand is clearly monotonic, and equilibrium is unique and stable.

This decrease in pathology as economies grow large is not due to the size of the economy per se – after all, existence, multiplicity and instability would all be exactly the same if the REPLICA economies consisted of a billion identical agents on each side. Rather, the decrease must be due to preference heterogeneity. Indeed, the idea that heterogeneity might overcome SMD has a long history in general equilibrium theory, stretching back informally to at least Becker [1962], followed formally by Hildenbrand [1983] with respect to endowment heterogeneity, and Grandmont [1987, 1992], Hildenbrand [1994], Giraud and Quah [2003], and Hildenbrand and Kneip [2005], among others, with respect to preference heterogeneity.

The size of the economy does, however, have direct effect on the robustness of equilibria. As markets grow large, average excess demand tends to grow faster than the aggregated variance of excess demand (due to a straightforward application of the Bienaymé formula on aggregation uncorrelated individual preference noise). To see this, we repeat the sampling exercise above, this time including and aggregating volatility in preferences to calculate robustness.[18] We find that at least one equilibrium is fragile in 70% of REPLICA samples, dropping to 40% for EXPERIMENT-SIZE samples and 12% for very large $N = 60$ samples.

The set of economies that suffer from none of the SMD-type pathologies explored in the paper (that is that have existence, uniqueness, stability and robustness) is a mere 22% in REPLICA economies, 50% in EXPERIMENT-SIZED economies and 85% when $N = 60$. Calculating the largest aggregate we can, the excess demand for our entire sample, FULL SAMPLE, has only one equilibrium and it is robust.

Aggregation thus improves the predictive power of general equilibrium theory in two ways in our data. First, preferences, though often quite non-neoclassical, are sufficiently heterogeneous to create relatively monotonic excess demands, and increasingly so as the number of agents in the economy increases. Second, excess demands, though noisy enough in small economies to frequently disrupt dynamic convergence processes, become quite stable at a large aggregate scales for classical statistical reasons. Jointly, our results thus suggest that departures from neoclassical models that loom large at the individual level and in replica economies may show a tendency to yield quite

[18] For computational simplicity we conduct this exercise using standard deviation volatility bands.

neoclassical outcomes in the aggregate.

Result 7. *Pathologies — non-existence, instability, and multiplicity of equilibria — disappear as markets grow large with the heterogeneous preferences measured in our experiment. Fragility also disappears in larger economies due to attenuation of volatility.*

6 Conclusion

We introduce new experimental methods for studying general equilibrium economies composed of naturally occurring (rather than artificial, induced) preferences. In our experiment we illustrate these methods by applying them to preferences for risk, an important sub-class of preferences. Specifically, subjects reveal their preferences over a pair of Arrow securities, and we use these preferences to sort subjects into different economies with a range of predicted properties. We compare market outcomes in these economies to general equilibrium predictions generated from participants' revealed preferences. Doing so allows us to empirically study open questions in general equilibrium theory that are impossible to study using prior methods.

What do we learn from this approach? First, although individual subjects' revealed preferences are quite prone to the types of pathologies highlighted by the Sonnenschein-Mantel-Debreu theorem (multiplicity, non-existence, instability), preference heterogeneity across subjects has a powerful taming effect on aggregate excess demand as markets get even moderately large. At the individual level (and as in prior experimental work on revealed preferences), subjects reveal preferences that are often non-neoclassical, and replica economies formed of pairs of subjects suffer SMD-like problems (especially multiplicity and instability) almost half the time. These preferences are, however, highly heterogeneous and aggregating only 12 subjects' preferences produces excess demand functions that are not far from monotone, with a unique (and stable) competitive equilibria in every economy in our dataset. These empirical results reinforce the suggestion of some theorists, including Grandmont [1987] and Hildenbrand [1994], that human preferences indeed may be heterogeneous enough to overcome SMD-style concerns in the aggregate.

Second, general equilibrium theory, applied to revealed preferences, predicts behavior in our markets quite well even though the preferences we study are not perfectly stable across decisions and our markets are small. Simply by partitioning subjects into economies according to their revealed preferences, we are able to influence the observed level of market prices in ways consistent with GE comparative statics. Moreover, quantitative GE benchmarks predict final prices in most markets and these benchmarks are especially good at predicting final allocations in our markets.

Third, volatility in individual preferences has a predictable influence over the ability of market dynamics to guide prices to equilibrium states. In economies with strong wealth effects, fluctuations in individual preferences can be large relative to excess demands, halting (or even reversing) convergence towards predicted equilibria. By measuring inconsistencies in individual choices in our revealed preference task, we can pre-classify economies as "robust" (likely to converge) or "fragile" to these fluctuations before ever observing subjects in markets. These classifications strongly predict convergence in our markets. What's more, because we can deliberatively construct economies with strong or weak wealth effects simply by sorting subjects into markets based on the convexity of their revealed preferences, we can experimentally control the likelihood of robustness and thereby the reliability of equilibrium convergence.

Our results illustrate some of the ways that studying economies built using real preferences can enrich our understanding of competitive behavior, and they highlight connections between markets and revealed preferences that are inaccessible to experiments using the usual sort of induced preference procedures. To follow up, we see two important new empirical directions.

First, we have so far studied a particular class of preferences, those for risk. Do our largely positive results on aggregation and convergence carry over to other sorts of preferences measurable in the lab (e.g. time preferences)? We had expected homegrown risk preferences to provide a particularly strong challenge to GE, given the frequent violations of neoclassical choice behavior in prior experimental risk-elicitation research. This leads us to be conjecture similarly positive results for GE economies constructed using other sorts preferences, but new empirical surprises surely are possible.

Second, our experiment was conducted using a particular market institution with features particularly well-suited to the questions motivating our research. The tatonnement institution is implemented using an interface very similar to our revealed preference task, and it is initially adapted to a pre-specified price grid matching the grid used in the Phase 1 elicitation. Perhaps most importantly, tatonnement allows trade only once, after prices have equilibrated; endowments remain constant thoughout market exchange, allowing us to use elicited excess demands to make sharp predictions about dynamics without imposing parametric assumptions on preferences. Studying other, more naturalistic institutions such as the double auction (in which endowments change over the course of the period) will require more elicitation or stronger structural assumptions, but will also allow us to study richer questions about dynamics in settings in which the properties of excess demand are changing over the course of market interaction.

References

David Ahn, Syngjoo Choi, Douglas Gale, and Shachar Kariv. Estimating ambiguity aversion in a portfolio choice experiment. *Quantitative Economics*, 5(2):195–223, 2007.

Christopher M. Anderson, Charles R. Plott, Ken-Ichi Shimomura, and Sander Granat. Global instability in experimental general equilibrium: The scarf example. *Journal of Economic Theory*, 115(2):209–249, 2004.

James Andreoni and John Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.

Kenneth Joseph Arrow and Frank Hahn, editors. *General competitive analysis*. North-Holland, 1971.

Elena Asparouhova, Peter Bossaerts, and Charles Plott. Excess demand and equilibration in multi-security financial markets: The empirical evidence. *Journal of Financial Markets*, 6(1):1–21, 2003.

Elena Asparouhova, Peter Bossaerts, and John O. Ledyard. Price formation in continuous double auctions; with implications for finance. 2011.

Elena Asparouhova, Peter Bossaerts, Nilanjan Roy, and William Zame. "lucas" in the laboratory. *Journal of Finance*, 71(6):2727–2780, 2016.

Gary S. Becker. Irrational behavior and economic theory. *Journal of Political Economy*, 70(1): 1–13, 1962.

Bruno Biais, Thomas Mariotti, Sophie Moinas, and Sébastien Pouget. Asset pricing and risksharing in a complete market: An experimental investigation. 2017.

Peter Bossaerts and Charles Plott. Basic principles of asset pricing theory: Evidence from large-scale experimental financial markets. *Review of Finance*, 8:135–169, 2004.

Peter Bossaerts, Charles Plott, and William Zame. Prices and allocations in financial markets: Theory, econometrics, and experiments. *Econometrica*, 75(4):993–1038, 2007.

Adriana Breaban and Charles N. Noussair. Trader characteristics and fundamental value trajectories in an asset market experiment. *Journal of Behavioural and Experimental Finance*, 8:1–17, 2015.

Timothy N. Cason and Daniel Friedman. Price formation in single call markets. *Econometrica*, 65 (2):311–345, 1997.

Edward H. Chamberlin. An experimental imperfect market. *Journal of Political Economy*, 56(2): 95–108, 1948.

Gary Charness, Uri Gneezy, and Alex Imas. Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior and Oranganization*, 87:43–51, 2013.

Syngjoo Choi, Raymond Fisman, Douglas Gale, and Shachar Kariv. Consistency and heterogeneity of individual behavior under uncertainty. *The American Economic Review*, 97(5):1921–1938, 2007.

James C. Cox and Glenn W. Harrison, editors. *Research in Experimental Economics: Volume 12*. JAI Press, 2008.

Sean Crockett, Ryan Oprea, and Charles R. Plott. Extreme walrasian dynamics: The gale example in the lab. *The American Economic Review*, 101(7):3196–3220, 2011.

Sean Crockett, John Duffy, and Yehuda Izhakian. An experimental test of the lucas asset pricing model. 2017.

Gerard Debreu. Excess demand functions. *Journal of Mathematical Economics*, 1(1):15–21, 1974.

David Easley and John O. Ledyard. Theories of price formation and exchange in double oral auctions. In Daniel Friedman and John Rust, editors, *The Double Auction Market: Institutions, Theories, and Evidence*, pages 63–97. Addison-Wesley, 1993.

Raymond Fisman, Shachar Kariv, and Daniel Markovits. Individual preferences for giving. *The American Economic Review*, 97(5):1858–1876, 2007.

Daniel Friedman. On the efficiency of experimental double auction markets. *The American Economic Review*, 74(1):60–72, 1984.

Daniel Friedman. A simple testable model of double auction markets. *Journal of Economic Behavior and Organization*, 15:47–70, 1991.

Daniel Friedman, R. Mark Isaac, Duncan James, and Shyam Sunder. *Risky Curves: On the Empirical Failure of Expected Utility*. Routledge, 2014.

David Gale. A note on global instability of competitive equilibrium. *Naval Research Logistics Quarterly*, 10:81–87, 1963.

Gael Giraud and John K.-H. Quah. Homothetic or cobb-douglas behavior through aggregation. *The B.E. Journal of Theoretical Economics*, 3, 2003.

Steven Gjerstad. Price dynamics in an exchange economy. *Economic Theory*, 52(2):461–500, 2013.

Steven Gjerstad and John Dickhaut. Price formation in double auctions. *Games and Economic Behavior*, 22(1):1–29, 1998.

Dharanjay K. Gode and Shyam Sunder. Allocative efficiency of markets with zero intelligence (zj) traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101:119–137, 1993.

Jacob Goeree and Luke Lindsay. Market design and the stability of general equilibrium. *Journal of Economic Theory*, 165:37–68, 2016.

Jean-Michel Grandmont. Distributions of preferences and the "law of demand". *Econometrica*, 55: 155–161, 1987.

Jean-Michel Grandmont. Transformation of the commodity space, behavioral heterogeneity, and the aggregation problem. *Journal of Economic Theory*, 57(1):1–35, 1992.

Faruk Gul. A theory of disappointment aversion. *Econometrica*, 59(3):667–686, 1991.

Bulent Guler, Volodymyr Lugovskyy, Daniela Puzzello, and Steven Tucker. An experimental study of bubble formation in asset markets using the tatonnement pricing mechanism. 2012.

Yoram Halevy, Dotan Persitz, and Lanny Zrill. Parametric recoverability of preferences. *Journal of Political Economy*, forthcoming.

Peter Hammond and Stefan Traub. A three-stage experimental test of revealed preference. Technical Report 72, Competitive Advantage in the Global Economy (CAGE) Working Paper Series, 2012.

Werner Hildenbrand. On the law of demand. *Econometrica*, 51:997–1019, 1983.

Werner Hildenbrand. *Market Demand: Theory and empirical evidence*. Princeton University Press, 1994.

Werner Hildenbrand and Alois Kneip. On behavioral heterogeneity. *Economic Theory*, 25(1): 155–169, 2005.

Masayoshi Hirota, Ming Hsu, Charles R. Plott, and Brian W. Rogers. Divergence, closed cycles and convergence in scarf environments: Experiments in the dynamics of general equilibrium systems. 2005.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.

Haim Levy. Risk and return: An experimental analysis. *International Economic Review*, 38(1): 119–149, 1997.

Rolf R. Mantel. On the characterization of aggregate excess demand. *Journal of Economic Theory*, 7(3):348–353, 1974.

Andreu Mas-Colell. On the equilibrium price set of an exchange economy. *Journal of Mathematical Economics*, 4(2):117–126, 1977.

Charles R. Plott. Market stability: Backward-bending supply in a laboratory experimental market. *Economic Enquiry*, 38(1):1–18, 2000.

Charles R. Plott and Glen George. Marshallian vs. Walrasian stability in an experimental market. *The Economic Journal*, 102:437–460, 1992.

Herbert Scarf. Some examples of global instability of the competitive equilibrium. *International Economic Review*, 1(3):157–172, 1960.

Lloyd S. Shapley and Martin Shubik. An example of a trading economy with three competitive equilibria. *Journal of Political Economy*, 85(4):873–875, 1977.

Vernon L. Smith. An experimental study of competitive market behavior. *Journal of Political Economy*, 70(2):111–137, 1962.

Vernon L. Smith, Gerry L. Suchanek, and Arlington W. Williams. Bubbles, crashes, and endogenous expectations in experimental spot asset market. *Econometrica*, 56(5):1119–1151, 1988.

Hugo Sonnenschein. Do Walras' identity and continuity characterize the class of community excess demand functions? *Journal of Economic Theory*, 6(4):345–353, 1973.

Alex Akira Toda and Keieran J. Walsh. Edgeworth box economies with multiple equilibria. *Economic Theory Bulletin*, 5:65–80, 2017.

Amos Tversky and Daniel Kahneman. Prospect theory: An analysis of decision under risk. *Journal of Risk and Uncertainty*, 5(4):297323, 1992.

Marie-Esprit-Léon Walras. *Elements of Pure Economics*. Translated, Harvard University Press (1954), 1877.

Arlington Williams, Vernon L. Smith, John Ledyard, and Steven Gjerstad. Concurrent trading in two experimental markets with demand interdependence. *Economic Theory*, 16:511–528, 2000.

A Supplemental Online Appendix: Details on the Design

A.1 Preference Typing

Following Phase 1, the 28 subjects in the experiment are partitioned into three groups: H (12 subjects), L (12 subjects), and E (4 subjects). Subjects assigned to group E made the most "erratic" choices in Phase 1R. Subjects assigned to group H reveal preferences in Phase 1R that are "more-convex-than" subjects assigned to group L. The subjects in H comprise one market as Phase 2 of the experiment commences, the subjects in L comprise another. Subjects in E are assigned to a market, but their choices and endowments are ignored for the purpose of calculating excess demand, and thus their decisions have no impact on prices. Subjects are not aware that they have been partitioned into these groups, though they are aware there are two equal-sized markets simultaneously in operation during Phase 2.

Subjects are sorted into groups on the basis of their 25 choices in Phase 1R. First, we sort these choices by price, in ascending order. Let $x_{i,t}$ be the quantity of x chosen by subject i in Phase 1R at the t^{th}-lowest price. Thus $x_{i,1}$ is i's choice at $p = .2$ (the lowest price), $x_{i,2}$ is i's choice at $p = .28$ (the second-lowest price). And so on. We define $y_{i,t}$ similarly. Let $c_{i,t} \in \mathbf{R}^2_+ = (x_{i,t}, y_{i,t})$ be the portfolio chosen by i at the t^{th}-lowest price.

Assignment of erratic subjects to group E.

Following Phase 1, we create three 24×1 vectors for each subject: $d_{i,x}$, $d_{i,y}$, and d_i. Let $d_{i,x}(t) = x_{i,t+1} - x_{i,t}$, $d_{iy}(t) = y_{i,t+1} - y_{i,t}$, and $d_{i,t} = \|c_{i,t+1}, c_{i,t}\|$, for $t \in \{1, 2, \ldots, 24\}$. Then, for each subject i we create a "noise" variable N_i where

$$N_i = \sum_{t=1}^{23} \eta_t d_{i,t}$$

where η is an indicator variable equal to 1 if $\text{sign}\left(d_{i,x}(t+1)\right) \neq \text{sign}\left(d_{i,x}(t)\right)$ <u>or</u> $\text{sign}\left(d_{i,y}(t+1)\right) \neq \text{sign}\left(d_{i,y}(t)\right)$. The two subjects of each endowment type with the largest value of N_i were assigned to group E.

Thus N_i is the length of the subject's linearly interpolated offer curve, but only including the line segments in which there is a sign change in one of the two goods quantities. A subject viewing the goods as perfect complements would generate $N_i = 0$. The maximum value of N_i for a subject who views the goods as perfect substitutes is roughly 146 for a subject with the corner X endowment (note that the subject's choice would be unrestricted for $p = 1$), and about 104 for a subject with the corner Y endowment. All symmetric, smooth, concave utility functions that we simulated produce N_i within these bounds. Disappointment aversion simulations produced slightly larger values than the perfect substitutes benchmark (by about 5-10%).

Subjects in the experiment assigned to group E have mean values of N_i equal to 673 and 522 for corner X and Y endowments, respectively (the medians were similar). By way of comparison, subjects assigned to H or L have mean values of N_i equal to 302 and 196 for corner X and Y endowments, respectively. Visual inspection of subject offer curves confirm unambiguously erratic choices for all subjects placed in E. Two subjects from E (one from each endowment type) are placed in each market during Phase 2, but their demands do not impact price changes.

Assignment of subjects to groups H and L. When $p = .76$, subjects with opposite endowment types ($(100, 0)$ and $(0, 76)$) face an identical budget line. We define a non-parametric measure of convexity in the neighborhood of this budget line to sort subjects into groups H and L. Specifically, we compute $S_x^i = x_{i,8} + x_{i,9} + x_{i,10}$ for each subject i at price gridpoints 8, 9, and 10 (prices 0.7, 0.76, and 0.83, respectively). We include these neighboring prices to smooth the measure, but we chose not to smooth across a wide range of prices because disappointment-averse subjects can change from viewing the goods as perfect complements to perfect substitutes in one grid point; thus S_x^i is a local non-parametric measure of convexity near the price of 0.76.

Regardless of endowment, S_x^i is monotonically decreasing in i's convexity of preferences by Proposition 2 . Of the 24 subjects not already assigned to group E, there are 12 with a corner X endowment (we refer to them as "Sellers") and 12 with corner Y endowment (to whom we refer as "Buyers"). We place the six sellers and six buyers with the lowest values of S_x^i in group H. The remaining twelve unassigned subjects are assigned to group L.

A.2 Tatonnement Algorithm

Our algorithm is designed to implement large price changes in early rounds of the period, and for the magnitude of price changes to be reduced by a factor of one half after each successive sign change of Z.

Let M be the number of subjects in a given market ($M = 12$ throughout our experiment). Let $Z_t = \sum_{i=1}^{M} x_{i,t} - 100(M/2)$ be realized aggregate excess demand given choices $x_{i,t}$ of security X in round t. Let $z_t = Z_t/M$ be per capita excess demand in round t. Let K =(.1745, .08725, .043625, .021813, .010906) be a scaling vector we use to reduce the magnitude of price changes after sign changes in Z_t. We refer to element n of K as $K(n)$. Let p_t be the price at the beginning of round t. Let $\bar{z}$ be expected per capita excess demand in the first round of the market at the exogenous price; we set $\bar{z} = 13.5$.

As previously noted, $p_1 = 1.57$. Let $k_1 = \frac{1}{\bar{z}}K(1)$, and for $t \geq 1$ let $k_{t+1} = k_t$ if $\operatorname{sign}(Z_{t+1} * Z_t) = 1$, otherwise $k_{t+1} = \frac{1}{\bar{z}}K(1+r)$, where r is the minimum of 4 and the number of times the sign of excess demand has changed in the current period, including the current round.

Let $A_t = k_t z_t$. Let $B_t = (.26175)\operatorname{sign}(z_t)$. Let $C_t = \operatorname{sign}(z_t)\min\{|A_t|, |B_t|\}$. B_t serves to guarantee that price changes are capped at a maximum of .26175 radians (about 15 degrees), so that prices don't swing too wildly in the face of large excess demand. If $C_t < 0$, let $D_t = \max\{\arctan(p_t) + C_t, .01\}$ (this guarantees that price is always greater than .01). Otherwise, if $C_t > 0$, then $D_t = \min\{\arctan(p_t) + C_t, 1.5608\}$ (which guarantees price is always less than 100). The bounds reflected by D_t are never close to binding in our experimental markets.

Finally, let $\rho_{t+1} = \tan(D_t)$, and let $\hat{\rho}_{t+1}$ equal ρ_{t+1} rounded to the price grid point used in the individual choice experiment. If $\hat{\rho}_{t+1} = p_t$, then $p_{s+1} = \rho_{s+1}$ for all $s \geq t$ (that is, if rounding price to the nearest gridpoint would cause a repeat of p_t in period $t+1$, then beginning in $t+1$ prices are no longer rounded to the nearest gridpoint). Otherwise, $p_{t+1} = \hat{\rho}_{t+1}$. Rounding to the nearest gridpoint in early rounds of the period facilitate a tighter comparison to choices in Phase 1 of the experiment.

If $\|z_t\| < 2$ for three consecutive rounds t, $t+1$, and $t+2$, then p_{t+2} is the market price in the current period, and each subject's portfolio choice $(x_{i,t+2}, y_{i,t+2})$ is the final allocation for the period. Note we implicitly assume that our tatonnement mechanism is a market maker who can absorb small excess demand imbalances. Our software also provides an alternative rationing option, but we did not use it in this experiment. If the convergence criterion on $\|z_t\|$ is not met through 25 rounds, then p_{25} is the market price for the current period, and each subject's portfolio choice $(x_{i,25}, y_{i,25})$ is the final allocation for the period.

B Supplemental Online Appendix: Theoretical Notes

Recall that a *preference* relation $\succeq$ is a complete, reflexive and transitive binary relation. For convenience we shall also assume continuity, which (together with monotonicity) is well known to ensure representation by a continuous *utility function*. Recall also that preferences are *convex* if every upper contour set (all points $\succeq$-preferred to a given point) is convex.

A preference relation $\succeq$ over $\mathbb{R}^2_+$ is monotone if $(x, y) \geq (x', y') \in \mathbb{R}^2_+ \implies (x, y) \succeq (x', y')$ and it is symmetric if $(x, y) \sim (y, x) \;\; \forall (x, y) \in \mathbb{R}^2_+$. We say that an agent in our economy is *minimally rational* if she has preferences over $\mathbb{R}^2_+$ that are continuous, monotone and symmetric.

The *budget set* $B(p, \omega)$ for given price $p > 0$ and endowment $(\omega_x, \omega_y) \neq (0, 0) \in \mathbb{R}^2_+$ consists of all bundles $(x, y) \in \mathbb{R}^2_+$ such that $y + px \leq \omega_y + p\omega_x$. For given preferences $\succeq$ over $\mathbb{R}^2_+$, a bundle (x^*, y^*) is *demanded* at price $p > 0$ and endowment ω if $(x^*, y^*) \in B(p, \omega)$ and $(x^*, y^*) \succeq (x, y) \; \forall (x, y) \in B(p, \omega)$.

Proposition 1. *Let a minimally rational agent demand $(x^*, y^*) \in B(p, \omega)$ given price $p > 0$ and endowment $\omega \in \mathbb{R}^2_+$. Then $(1 - p)(x^* - y^*) \geq 0$. The inequality is strict if preferences are strictly monotone and $p \neq 1$.*

Proof. Since preferences are monotone and continuous, they are represented by a continuous utility function $U(x, y)$. The budget set is compact, so U is maximized somewhere on $B(p, \omega)$, i.e., demand (x^*, y^*) exists. To prove the result, take a point $(x, y) \in B(p, \omega)$ where the inequality fails, so $(1 - p)(x - y) < 0$. It suffices to find a point in $B(p, \omega)$ that is strictly preferred to (x, y).

Suppose that $p < 1$ and $x < y$ for choice $(x, y) \in B(p, \omega)$. Consider bundle (y, x), for which $U(x, y) = U(y, x)$ by symmetry. Bundle (y, x) is cheaper than (x, y), since $px + y - (py + x) = (1 - p)(y - x) > 0$. By strict monotonicity of preferences, bundle $(y, x + (1 - p)(y - x)) \in B(p, \omega)$ is preferred to (x, y). Similarly, when $p > 1$, any affordable choice (x, y) with $x > y$ is strictly less preferred than a range of affordable choices including $(y, x + (p - 1)(x - y))$. $\square$

When (x, y) is a bundle of complementary equi-probable Arrow securities (i.e, the only states are X and Y, and each occurs with probability 0.5), it is a straightforward to show directly from the definitions that the condition $(1-p)(x-y) < 0$ implies that (x, y) is strictly first-order stochastically dominated by other affordable allocations, e.g., by $(y, x + (1 - p)(y - x))$ when $p < 1$ and $x < y$.

Suppose that $u : \mathbb{R} \to \mathbb{R}$ is a Bernoulli function, i.e., it is strictly increasing and continuous. In the case of equi-probable Arrow securities, let $U(x, y)$ be its expected value, i.e., $U(x, y) =$

$0.5[u(x)+u(y)]$. Then it is straightforward to show that U is monotone, symmetric and continuous, so Proposition 1 applies. Thus Proposition 1 generalizes the well known result that expected utility maximizers do not choose lotteries that are first-order stochastically dominated.

Disappointment aversion Gul [1991] in the present context implies choice on the diagonal $x = y$ for all prices in some neighborhood of $p = 1$. The underlying preferences are continuous, monotone, symmetric and convex, but not smooth at the diagonal and not representable as the expectation of a Bernoulli function.

Recall that the marginal rate of substitution (MRS) for given preferences at a bundle (x, y) is the (negative) slope of its indifference curve through that point, i.e., $MRS_U = -\frac{dy}{dx}|_{U=const}$ where $U(x, y)$ represents the given preferences. MRS is not defined at kinks but it is not very restrictive to assume, as we do below, that it is defined almost everywhere (a.e.), e.g., everywhere except along the diagonal for Disappointment Averse preferences. We will say that a demand $(x^*, y^*) \in B(p, \omega)$ is *interior* if $x^*y^* > 0$ and $MRS(x^*, y^*)$ is defined. The following partial ordering on the space of preferences will be useful.

Definition 1. *The preferences that $V(x, y)$ represent on $\mathbb{R}^2_+$ are* more convex than *those that $U(x, y)$ represent if*

1. *U and V both represent continuous, monotone, symmetric and convex preferences with MRS_U and MRS_V defined a.e., and*
2. *$(x - y)(MRS_U(x, y) - MRS_V(x, y)) \geq 0 \;\; \forall (x, y) \in \mathbb{R}^2_+$ for which the left hand side is defined.*

Proposition 2. *For given prices $p > 0$ and endowment $\omega \in \mathbb{R}^2_+$, let $(x^*_U, y^*_U), (x^*_V, y^*_V) \in B(p, \omega)$ be demands for two different minimally rational agents whose preferences represented by utility functions $U(x, y)$ and $V(x, y)$ respectively, with V more convex than U. Then $(1-p)(x^*_U - x^*_V) \geq 0$. If $p \neq 1$, both optima are interior and the more-convex-than ordering is strict, then the inequality is strict.*

Proof. Note that $(x^*_U, y^*_U), (x^*_V, y^*_V)$ are unique, since preferences are strictly convex and the budget set is convex. Standard arguments from intermediate microeconomics now ensure that, if interior, the demands are characterized by $p = MRS_U(x^*_U, y^*_U) = MRS_V(x^*_V, y^*_V)$.

First consider the case $p < 1$, strict ordering, and interior demands. Then $x^*_U > y^*_U$ by Proposition 1, and $p < MRS_V(x^*_U, y^*_U)$ since V-preferences are strictly more convex than the U-preferences. Now strict convexity implies that MRS_V decreases as we move along the budget line $\{(x, y) : y + px = \omega_y + p\omega_x\}$ from (x^*_U, y^*_U) towards the diagonal point $(\frac{\omega_y+p\omega_x}{1+p}, \frac{\omega_y+p\omega_x}{1+p})$.

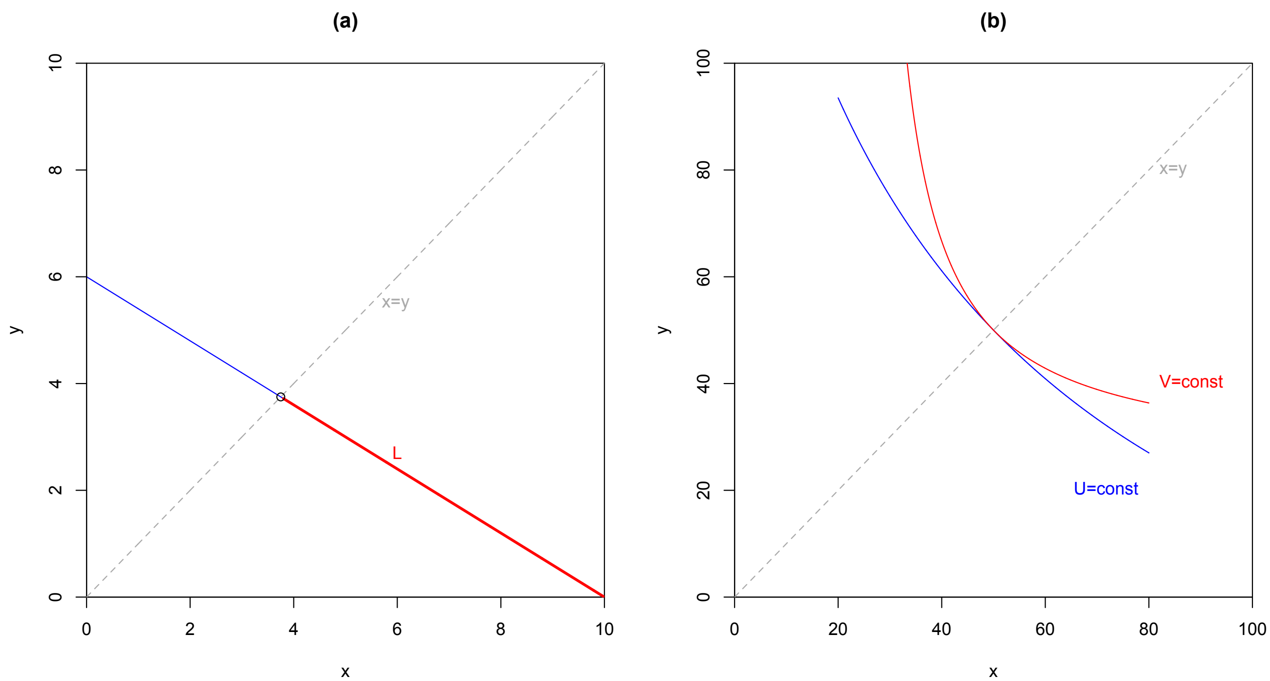


Figure 9: Demand and Convexity. Panel A: For $p < 1$, Proposition 1 implies that demand must be on the line segment L; the portion of the budget line above the diagonal is dominated. Panel B: Preferences represented by V are more convex than those represented by U. If it has a kink on the diagonal $x = y$ then V represents disappointment averse preferences, which are not the expected value of any Bernoulli function.

Proposition 1 implies that $x_V^* \geq \frac{\omega_y + p\omega_x}{1+p}$, so we conclude that $x_V^* \in [\frac{\omega_y+p\omega_x}{1+p}, x_U^*)$. Thus the desired inequality $(1-p)(x_U^* - x_U^*) > 0$ holds in this case. The argument for the interior case when $p > 1$ is similar. The weak inequality can be established in non-interior cases by taking appropriate limits. $\square$

Recall that $u : \mathbb{R} \to \mathbb{R}$ is a *Bernoulli function* if it is strictly increasing and continuous, and that Bernoulli function v is *more concave than* u if $v(c) = h(u(c)) \quad \forall c \geq 0$ where h is a positive increasing concave function.

Lemma 1. *Let $U(x, y) = 0.5(u(x) + u(y))$ and $V(x, y) = 0.5(v(x) + v(y))$ be expected utility for differentiable concave Bernoulli functions u, v where v is more concave than u. Then V represents more convex preferences than U.*

Proof. Condition 1 in the more-convex-than definition holds because, as expected values (with equiprobable states) of Bernoulli functions, U and V automatically represent continuous, monotone, symmetric and convex preferences on $\mathbb{R}^2_+$. Since the Bernoulli functions are differentiable, the MRS's are defined everywhere on $\mathbb{R}^2_+$, with $MRS_U(x, y) = -\frac{dy}{dx}|_{U=const} = \frac{\partial U}{\partial x} / \frac{\partial U}{\partial y} = \frac{u'(x)}{u'(y)}$. Similarly, $MRS_V(x, y) = \frac{v'(x)}{v'(y)}$. By definition, v more concave than u means that $v(c) = h(u(c))$ for some concave funtion h, which we can assume to be differentiable. Since $v'(c) = h'(u(c))u'(c)$, we have $MRS_V(x, y) = \frac{h'(u(x))}{h'(u(y))} MRS_U(x, y)$. Thus $MRS_V(x, y) \leq MRS_U(x, y) \iff \frac{h'(u(x))}{h'(u(y))} \leq 1 \iff x \geq y$, since h is concave and u is increasing. Condition 2 follows directly. $\square$

Corollary 1. *If v is more concave than u, then on any budget line with $p \neq 1$, an expected v-maximizer will demand an allocation closer to the diagonal $x = y$ than will an expected u-maximizer.*

Proof. By Lemma 1, an expected v-maximizer's preferences on $\mathbb{R}^2_+$ represented by V are more convex than an expected u-maximizer's preferences represented by U. Proposition 1 tells us that both demands fall on the same side of the diagonal, and Proposition 2 tells us that $(1-p)(x_U^* - x_V^*) \geq 0$. The conclusion follows directly. $\square$

Revealed Competitive Equilibrium (RCE). Let $\mathcal{E} \in$ {LxLy, HxHy, HxLy, LxHy}, or more generally let $\mathcal{E}$ be any Edgeworth box economy, that is, an economy with a finite set of agents $\mathbb{H}_{\mathcal{E}}$ who each have a fixed endowment $\omega^h = (\omega_x^h, \omega_y^h)$ in $\mathbb{R}^2_+$ and preferences represented by a continuous monotone symmetric utility function $U_h(x, y)$. Suppose that we observe all agents' choices $(x^h(p_i), y^h(p_i))$ on budget lines defined by those endowments and a fixed price grid $p_0 \leq p_1 \leq ... \leq p_n$. To avoid trivialities, assume that $p_i = 1$ for some i with $0 < i < n$. Let $Z^R(p_i) = \sum_{h \in \mathbb{H}} z^h(p_i) \equiv \sum_{h \in \mathbb{H}} x^h(p_i) - \omega_x^h$ be revealed aggregate excess demand.

Definition 2. *A RCE is a price interval* $[p_i, p_{i+1}]$ *spanning adjacent grid prices such that* $Z^R(p_i)Z^R(p_{i+1}) < 0$. *If* $Z^R(p_i) = 0$ *at some grid point* p_i *then* $[p_i, p_i]$ *is also a RCE.*

For two endowment types S and D (that is, for each $h \in \mathcal{E}$, either $\omega^h = \omega^S$ or $\omega^h = \omega^D$), we predict that the allocation corresponding to an RCE price interval $[p_i, p_{i+1}]$ will lie in a particular rectangle $[\check{x}, \bar{x}] \times [\check{y}, \bar{y}]$ in the usual Edgeworth box diagram. The corners are defined by
$\check{w} = \min\{S_w(p_i), S_w(p_{i+1}), D_w(p_i), D_w(p_{i+1})\}$ and
$\bar{w} = \max\{S_w(p_i), S_w(p_{i+1}), D_w(p_i), D_w(p_{i+1})\}$, for $w = x, y$, where in turn
$S(p) = (S_x(p), S_y(p)) = \sum_{h\in\mathcal{E}} \omega^h - \sum_{h\in S} \left(\hat{x}^h(p), \hat{y}^h(p)\right)$ and
$D(p) = (D_x(p), D_y(p)) = \sum_{h\in D} \left(\hat{x}^h(p), \hat{y}^h(p)\right)$ are the total allocations chosen at price p by the set of traders with endowment type S (from the perspective of type D's origin), and by the complementary set of traders with endowment type D.

Definition 3. *Edgeworth box economy A is* more convex than *Edgeworth box economy B if there is a bijection* $\rho : \mathbb{H}_A \to \mathbb{H}_B$ *such that for all* $h \in \mathbb{H}_A$

1. $\omega^h = \omega^{\rho(h)}$, *and*
2. U_h *is more convex than* $U_{\rho(h)}$.

That is, we can pair up agents with the same endowments across economies, in such a way that each agent in A has more convex preferences than his counterpart in economy B. If our empirical convexity measure S_x captures the "more convex than" ordering of individual preferences, then *any* 1:1 mapping ρ between agents in Hx and Lx and between agents in Hy and Ly will satisfy the definition just given, and allow us to conclude that the HxHy economy is more convex than the LxLy economy.

Proposition 3. *Suppose that Edgeworth box economy A is more convex than Edgeworth box economy B, and aggregate x endowment is greater than aggregate y endowment in each economy. Then* $\min\{p \in RCE(A)\} \le \min\{p \in RCE(B)\}$ *and* $\max\{p \in RCE(A)\} \le \max\{p \in RCE(B)\} \le 1$.

Proof. $Z_A(p) = \sum_{h\in\mathbb{H}_A} z^h(p) \equiv \sum_{h\in\mathbb{H}_A}[x^h(p) - \omega_x^h]$ is excess demand in economy A. Applying the bijection ρ and Definition 3 part 1 we see that $\sum_{h\in\mathbb{H}_B} \omega_x^h = \sum_{h\in\mathbb{H}_A} \omega_x^{\rho(h)} = \sum_{h\in\mathbb{H}_A} \omega_x^h$, so aggregate endowments are the same in both economies. By hypothesis, in each economy $\omega_Y \equiv \sum_{h\in\mathbb{H}_A} \omega_y^h < \sum_{h\in\mathbb{H}_A} \omega_x^h \equiv \omega_X$. Writing aggregate demand as $X(p) = \sum_{h\in\mathbb{H}_A} x^h(p)$ and $Y(p) = \sum_{h\in\mathbb{H}_A} y^h(p)$, we see (by summing the budget constraints) that

$$Y(p) + pX(p) = \omega_Y + p\omega_X. \tag{2}$$

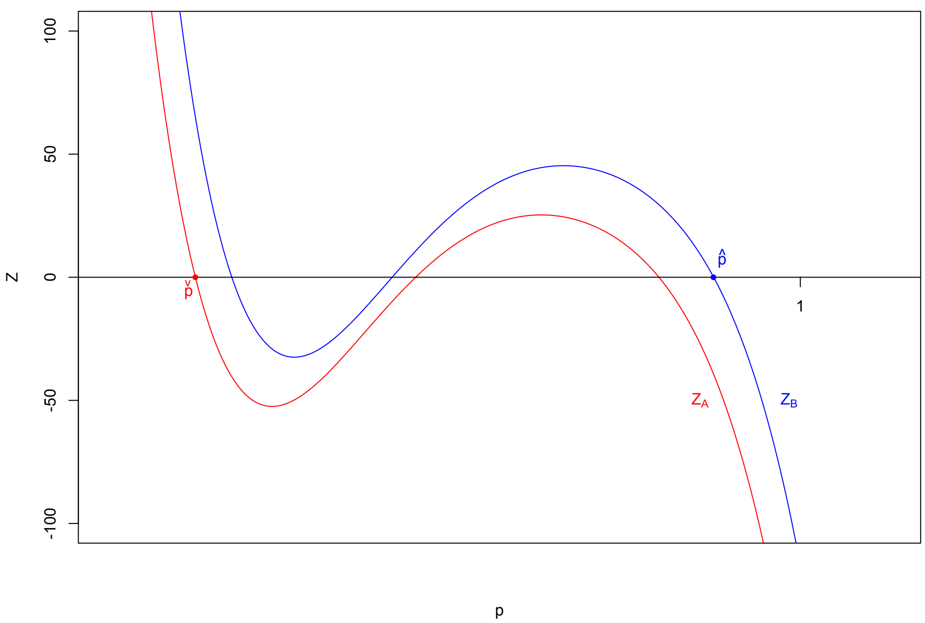


Figure 10: Excess demands for two economies. Economy A is more convex than economy B, and therefore $Z_A(p) \leq Z_B(p) \;\; \forall p \leq 1$.

Suppose that $p > 1$. Then applying Proposition 1 and summing yields $Y(p) > X(p)$, so the left hand side of equation (2) exceeds $(1+p)X(p)$. We already noted that $\omega_Y < \omega_X$ by hypothesis, so the right hand side of equation (2) is strictly less than $(1+p)\omega_X$. Therefore $X(p) < \omega_X$ so $Z(p) < 0$ for any $p > 1$. This proves the last inequality of the Proposition, that no RCE price exceeds 1.0 in either economy.

Now suppose that $p \leq 1$. Again applying Proposition 1 and summing yields $Y_k(p) \leq X_k(p)$ for each economy $k = A, B$. Applying Definition 3 part 2 and Proposition 2, and then summing, tells us that $X_A(p) \leq X_B(p)$. (That is, when $p \leq 1$, both aggregate demands are below the diagonal, but demand in the more convex economy is closer to the diagonal.) Since aggregate endowment is the same in both economies, we see that excess demand satisfies $Z_A(p) \leq Z_B(p) \;\; \forall p \geq 1$. Let $\hat{p} = \max\{p \in RCE(B)\}$. Recalling that Z is continuous and is negative for $p > 1$, we see that $p \geq \hat{p} \implies Z_B(p) < 0$ and a fortiori $\implies Z_A(p) < 0$. Thus all roots of Z_A are indeed bounded above by $\hat{p}$.

It remains only to prove that $\min\{p \in RCE(A)\} \leq \min\{p \in RCE(B)\}$. This follows from a parallel argument for $\check{p} = \min\{p \in RCE(A)\}$, recalling that, by monotonicity, Z is positive for all $p < \check{p}$. $\square$

Figure 10 illustrates why the "natural" ordering of RCE applies to the largest and smallest equilibrium prices (generically index +1), but not to all multiple equilibria, which can include upcrossings (index -1).

We shall say that an economy $\mathcal{E}$ is *nice* if its excess demand function Z^R

(1) has a unique RCE tunnel $[p_i, p_{i+1}]$, and

(2a) either $Z^R(p_i) > 0 > Z^R(p_{i+1})$ or

(2b) $Z^R(p_i) = 0$ and $Z^R(p_{i-1}) > 0 > Z^R(p_{i+1})$.

That is, $\mathcal{E}$ is *nice* if (1) an RCE exists and it is unique, and (2) it is tatonnement-stable. Following the SMD literature, we shall say that the economy is *pathological* if it is not nice, i.e., if it has non-existent RCE or non-unique RCE or unstable RCE.

In defining the classification formally, we must deal with the possibility that the economy is pathological, and that there are multiple RCE. For a particular RCE $[p_i, p_{i+1}]$, let $N(i) \subset \{p_0, ..., p_n\}$ be the subset of grid prices (excluding p_i and p_{i+1}) that are closer to the given RCE tunnel than to any other RCE; if the given RCE is unique, then $N(i)$ is the entire price grid other than p_i, p_{i+1}. Given the empirical value $\sigma_R > 0$ of a volatility measure, we shall say that the given RCE is *fragile* if, for any $j \in N(i)$, we have $(Z(p_j) - \sigma_R)(Z(p_j) + \sigma_R) < 0$. That is, for some grid price outside the price tunnel, volatility can change the sign of excess demand, in which case tatonnement dynamics may converge to some non-RCE price. Otherwise, we classify the RCE as *robust*.

C Supplemental Online Appendix: Revealed Preferences

In this section we provide a plot, for each session, of all subjects' revealed preferences from Phase 1R of the experiment. Each panel plots the log price on the x-axis and the corresponding share of x, $x/(x+y)$ chosen by the subject at that price. Vertical lines mark a price of 1 ($log(p) = 0$) and horizontal lines an equal share $x/(x+y) = 0.5$. The preference classification made by the software for the subject ("H", "L" or "E") and the endowment type assigned to the subject ("x" or "y") are shown in the left margin in each case.

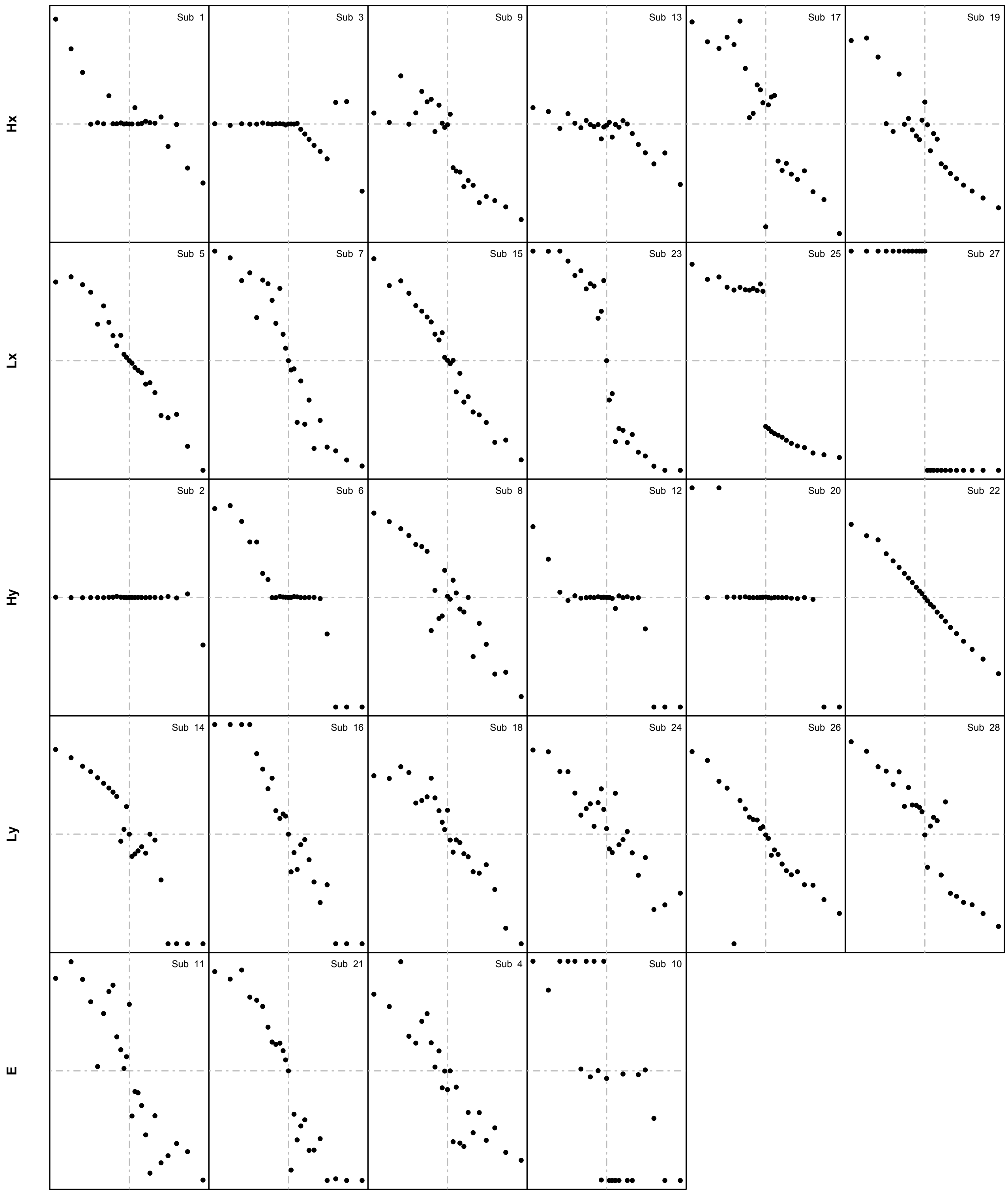


Figure 11: Phase 1R choices for all subjects in session 1.

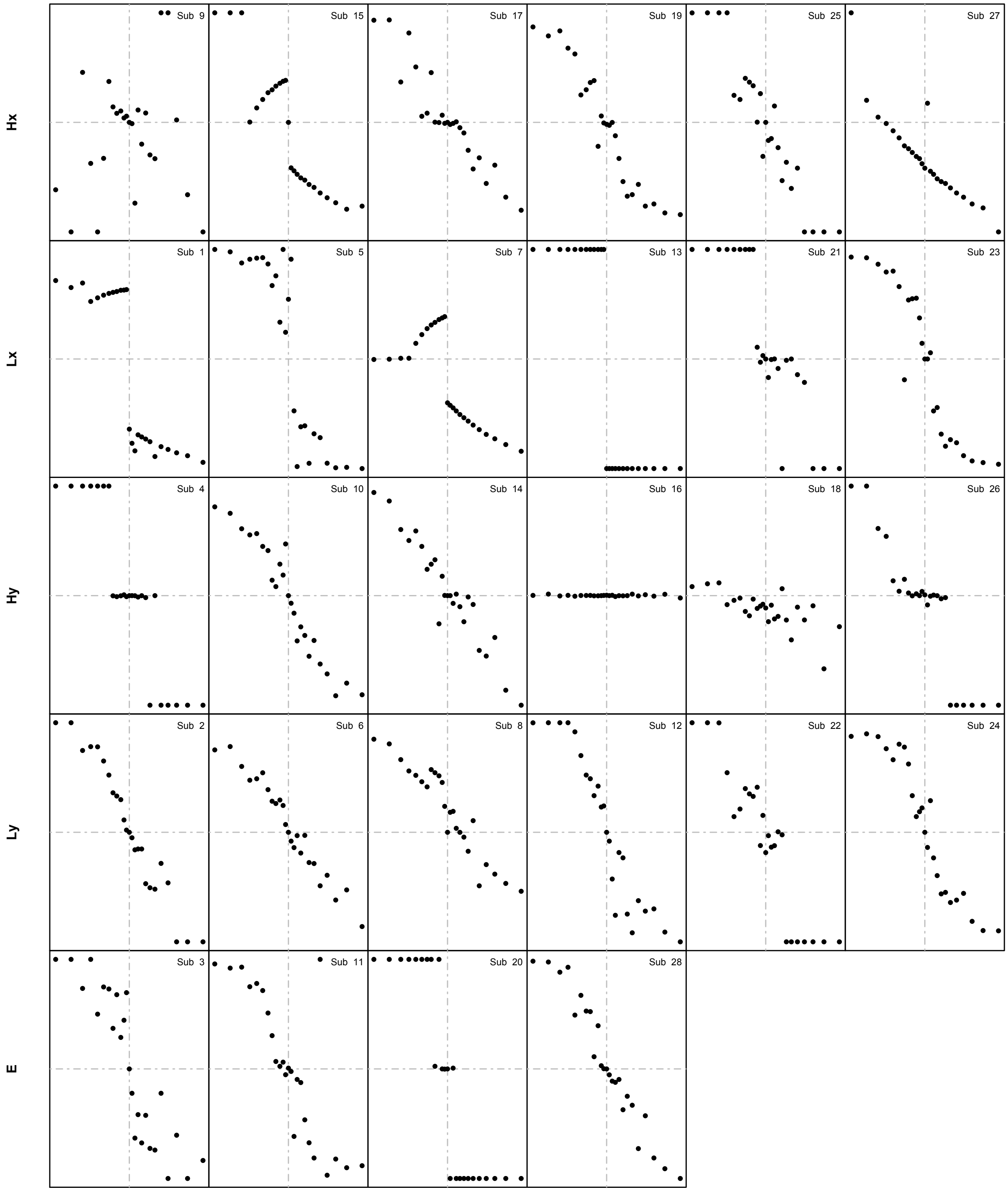


Figure 12: Phase 1R choices for all subjects in session 2.

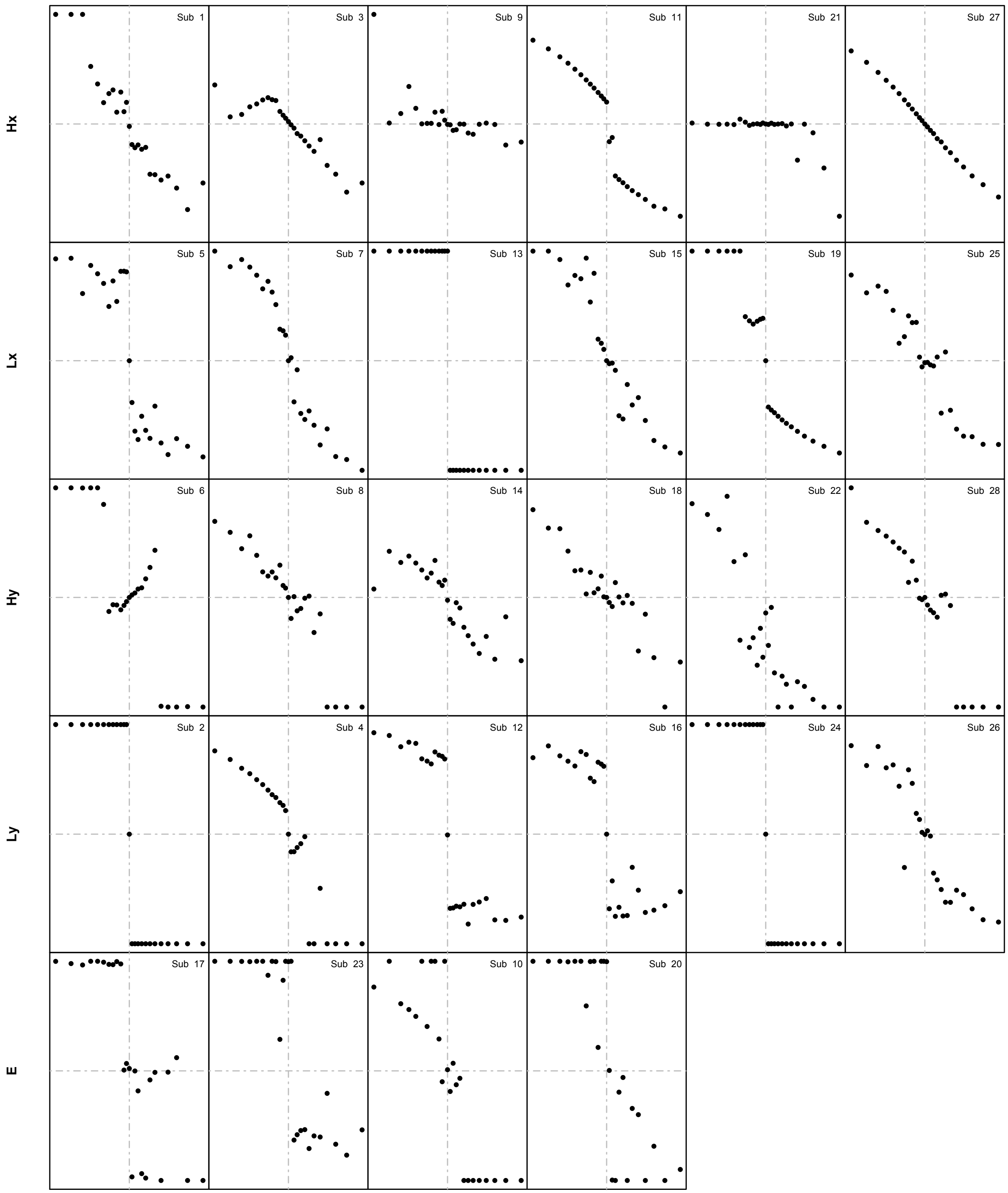


Figure 13: Phase 1R choices for all subjects in session 3.

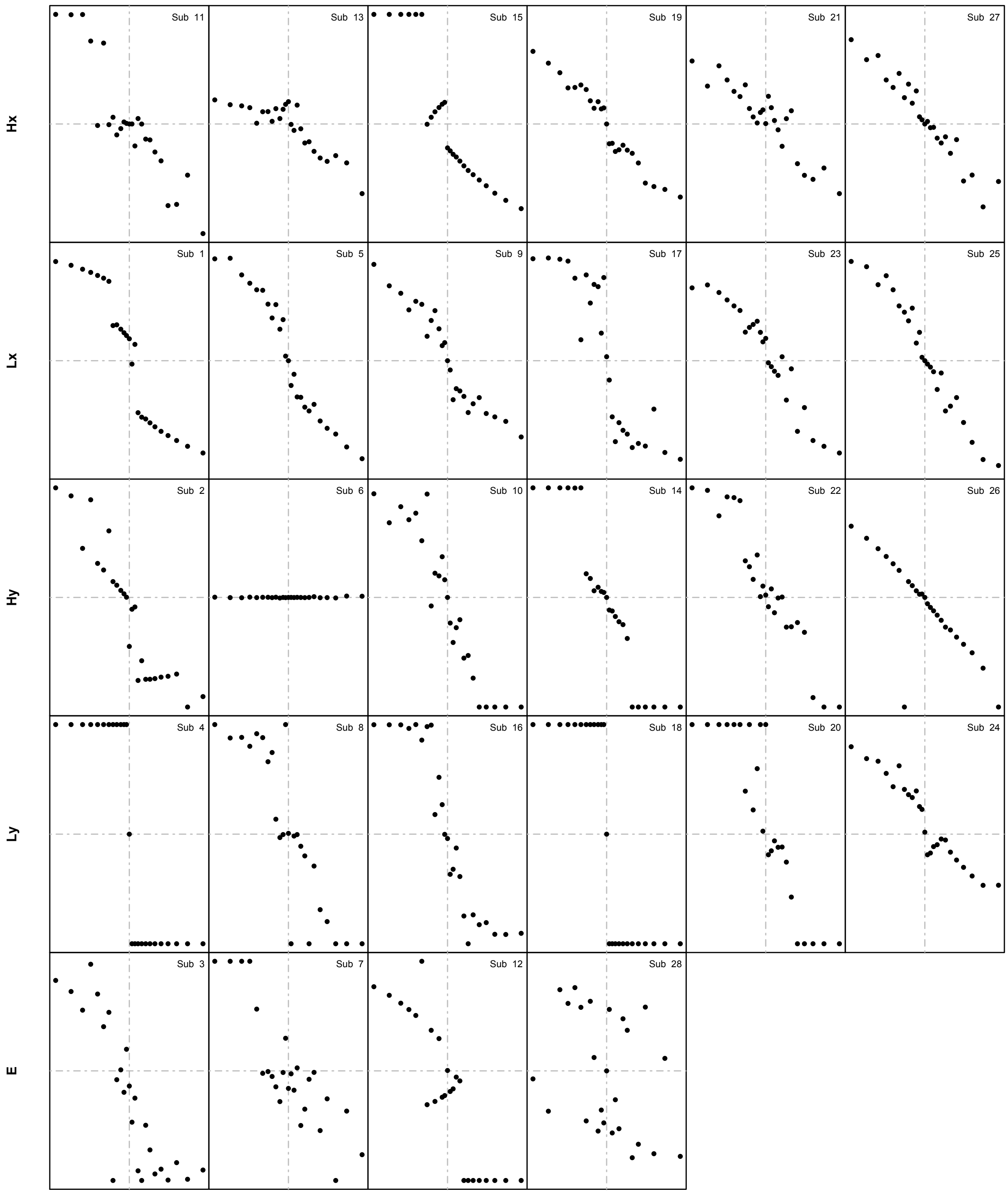


Figure 14: Phase 1R choices for all subjects in session 4.

D Supplemental Appendix: Instructions to Subjects

The following, first part of the instructions were handed out at the very beginning of the experiment.

Instructions

This is an experiment in decision-making funded by the National Science Foundation. Your payment will depend partly on your decisions and partly on chance. Please pay careful attention to the instructions; you are likely to make more money if you understand them completely.

The entire experiment should be completed within an hour and a half. At the end of the experiment you will be paid privately, in cash. You will receive \$7 as a participation fee (simply for showing up on time and participating), plus an additional amount that depends on your decisions in the experiment, as explained below.

During the experiment we will speak in terms of experimental tokens instead of dollars. Your payment will be calculated in terms of tokens and then translated at the end of the experiment into dollars at the following rate: 3 Tokens = 1 Dollar

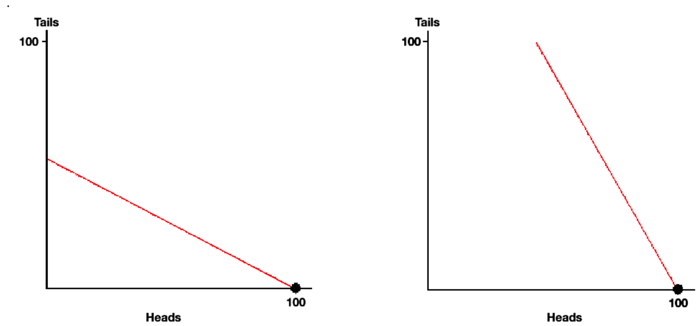
Experiment Outline

This experiment will consist of two separate phases: Phase 1 and Phase 2. In these instructions we will explain Phase 1, and when Phase 1 is complete we will hand out instructions concerning Phase 2. After both phases are complete, we will randomly select one of your decisions to determine your cash payment for the experiment. Which decision will be determined by the draw of a ball from a bingo cage filled with numbered balls.

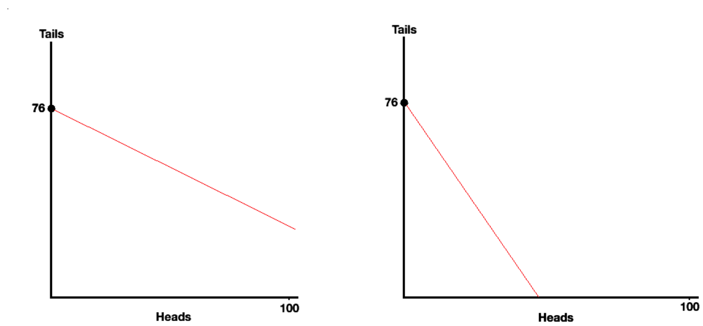
Phase 1: Independent decision problems

In Phase 1 you will face a sequence of periods and make one choice in each. You will begin each period with an endowment: some quantity of tokens in each of two separate accounts, Heads and Tails. Your decision in each period will be whether to keep your endowment or select an alternative allocation of tokens into these accounts. Your endowment will be the same in each period of Phase 1.

Your computer screen will show your endowment as a dot on a simple, 2-dimensional graph with the number of units in your Heads account shown on the horizontal (x) axis and your Tails account on the vertical (y) axis. You are welcome to keep your endowment allocation by pressing the "Confirm Selection" button. However, the screen will also show you a line (called the budget line) containing



Instructions Figure 1:

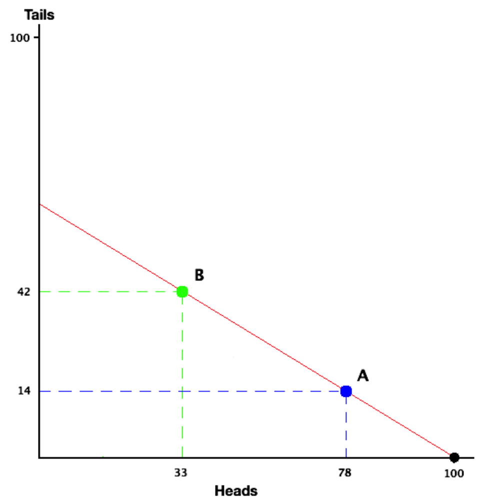


Instructions Figure 2:

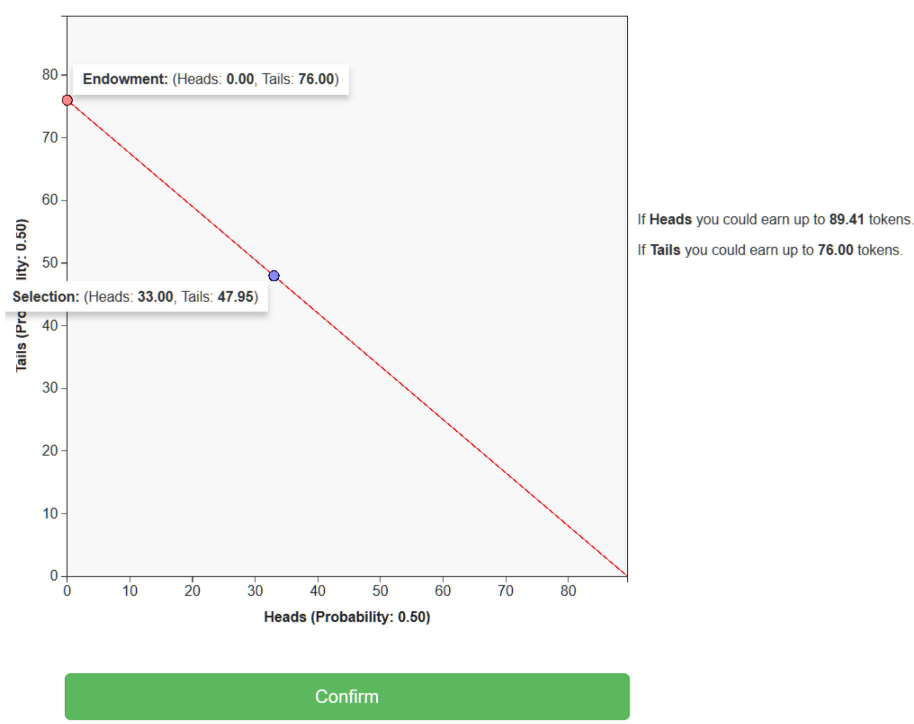
a range of other available allocation choices. Simply by left-clicking anywhere on the line, you can select a new allocation: after clicking you will see a blue dot appear, showing your new selection.

When you are satisfied with your choice, you must press the “Confirm Selection” button, making the selection your official allocation choice for the period. You will have up to two minutes to make each decision (time remaining will be displayed at the top of your screen); if you have not confirmed your selection by the time the counter reaches zero, your current selection will be automatically selected.

Examples of budget lines you may face appear in Figures 1 and 2. In Figure 1, the endowment is 100 tokens in the Heads account and 0 tokens in the Tails account. In Figure 2, the endowment is 0 tokens in the Heads account and 76 tokens in the Tails account. When all participants have confirmed their selections, a new period will begin with a new budget line (but the same endowment). For example, as illustrated in Figure 3, choice A represents a decision to allocate 78 tokens in the Heads account and 14 tokens in the Tails account. Another possible allocation is B, in which you allocate 33 tokens in the Heads account and 42 tokens in the Tails account. Notice that your



Instructions Figure 3:



Instructions Figure 4:

endowment is always an available choice.

After all participants have submitted their decisions, you will move on to the next decision problem. The actual computer program dialog window is shown in Figure 4.

If one of the periods in Phase 1 is (randomly) selected to determine your payment for the experiment, the amount of cash you earn will be determined by the number of tokens in your Heads and Tails accounts for that period. Specifically, we will flip a coin: if it comes up heads, you will receive the number of tokens in your Heads account, and if it comes up tails, you will receive the number of tokens in your Tails account (this is why "(Probability: 0.5)" appears beside Heads and Tails in Figure 4 and on your screen). Thus, in the end you will either receive tokens from your Heads account or from your Tails account, but not both.

For example, if the period shown in Figure 4 was randomly selected to determine your payment, you would have a 50% chance of receiving 33.00 tokens and a 50% chance of receiving 47.95 tokens. Recall that tokens convert to dollars at a rate of 3 tokens to 1 dollar. After all participants have completed 5 practice problems and all of the "official" decision problems in Phase 1, we will tell you that Phase 1 has ended and provide instructions for Phase 2. When both phases are completed we will determine payments for the experiment.

Privacy Policy

Your participation in the experiment and any information about your payoffs will be kept strictly confidential. Your payment-receipt and participant consent form are the only places in which your name is recorded. In order to keep your decisions private, please do not reveal your choices to any other participant. Please do not talk with anyone during the experiment, please turn off any phones or other electronic devices, and please refrain from using headphones or earbuds during the experiment. We ask everyone to remain silent until the end of the last round. If there are no further questions, you are ready to start.

The following, second part of the instructions were handed out after the conclusion of Phase 1.

Phase 2: Market decision problems

In each period of Phase 2 you will you will begin with the same endowment of tokens as in Phase 1, and you will make decisions by clicking on budget lines running through your endowment point, just as in Phase 1. However, unlike Phase 1, you will make several budget line decisions in each period, and the budget lines you face will be a consequence of the interaction of participant choices in a market. Here is how it works.

1. At the beginning of the period you will be placed into one of two markets with other participants. Participants in the experiment will be equally divided between the two markets. In each market some participants will have an endowment of tokens only in their Heads account, others only in their Tails account. This endowment of tokens will match each individual's endowment in Phase 1. As in Phase 1, you will have the opportunity to keep your endowment or express a preference to trade it for an alternative allocation by choosing a point on your budget line. However, for each decision the slope (that is, the steepness) of the budget line will always be identical for everyone in your market. This means that the number of tokens in your Tails account that you must give up in order to receive an additional token in your Heads account (or the number of tokens you can receive in your Tails account for each token you are willing to give up in your Heads account) is the same for everyone in the room. Let's call the number of Tails tokens you can give/receive per Heads token on the current budget line, the price of Heads.

2. As in Phase 1, you can choose to either keep your endowment or click on a new preferred point on the budget line and click "Confirm Selection." Unlike Phase 1, this choice is not "final." Instead, this choice is your demand for Heads and Tails tokens at the current price.

3. The computer will add up all of the demands for Heads and Tails tokens and add up all of the endowments of Heads and Tails tokens across all of the participants in your market. If the summed demands for Heads is significantly greater than the summed endowments of Heads (that is, if participants in your market request more Heads tokens than are available in the market), the computer will increase the price of Heads (that is, it will make the budget line steeper) and let everyone choose again. If instead the summed demands for Heads is significantly smaller than the total endowments of Heads, the computer will instead decrease the price of Heads (that is it will make the budget line flatter) and let everyone choose again. In either case, your endowment will not change when facing your new choice.

4. If the summed demands for Heads and Tails are "close enough" to the summed endowments, the computer will end the period and whatever your last choice was in the period will be your final allocation for the period (the computer may also end the period if it lasts too long). Thus, each decision you make may be your final choice for the period, depending on the choices of everyone else in the market. Since there will be many participants in your market, your choice will generally represent a relatively small fraction of total demand, and thus will

have a negligible impact on the price.

After the period ends we will have one additional period, with participants potentially reassigned between the two markets. Then Phase 2 will be complete. Afterward, each participant will draw a ball from a bingo cage with balls numbered 1-100. If the ball comes up between 1 and 50, you will be paid for the corresponding decision in Phase 1 (recall you made 50 non-practice decisions in Phase 1). If the number on the ball is between 51 and 75, you will be paid for the final allocation in the first period of Phase 2, and if the number on the ball is between 76 and 100, you will be paid for the final allocation in the second period of Phase 2. Thus there is a 50% chance your payment will be determined by Phase 1 and a 50% chance your payment will be determined by Phase 2, where each of the two periods in Phase 2 has an equal chance of being the payment period.