A General Revealed Preference Test for Quasi-Linear Preferences: Theory and Experiments

Marco Castillo* Mikhail Freer†

May 15, 2020

Abstract

We provide a generalized revealed preference test for quasi-linear preferences. The test applies to nonlinear budget sets and nonconvex preferences and can be used in contexts such as taxation and nonlinear pricing. We use this test to evaluate the prevalence of quasi-linear preferences in a laboratory experiment on consumption with real consumption goods. We find that half of the subjects who satisfy the Generalized Axiom of Revealed Preferences also satisfy the criteria for quasi-linear preferences. A closer look at consumption patterns suggests that a majority of subjects who satisfy either of these axioms exhibit preferences that are arbitrarily close to perfect substitutes.
Keywords: revealed preferences; quasi-linear preferences; lab experiments
JEL codes: D12; D11; C91

*Department of Economics, Texas A&M University; Melbourne Institute, University of Melbourne; IZA (Bonn, Germany), e-mail: `marco.castillo@tamu.edu`

†Department of Economics, University of Essex. e-mail: `m.freer@essex.ac.uk`

1 Introduction

It is difficult to overstate the importance of the assumption of quasi-linear (QL) preferences in both theoretical and empirical economics. The assumption plays a crucial role in mechanism design, the theory of the household, and applied welfare analysis. It is, for instance, a necessary assumption for the Revenue Equivalence theorem (Myerson, 1981; Krishna, 2009), the existence of the truth-revealing dominant strategy mechanism for public goods (Green and Laffont, 1977), and the Rotten Kids theorem (Becker, 1974; Bergstrom and Cornes, 1983). It is also a frequently invoked assumption in applied welfare analysis (Domencich and McFadden, 1975; Allcott and Taubinsky, 2015). It is also a crucial assumption in the empirical literature that uses bunching at kinks and notches of budget sets (see Kleven, 2016, for a review). The first contribution of our paper is to provide a new test for QL preferences that has many empirical applications. The second contribution is to provide the first empirical test of QL preferences in a laboratory setting.

Contribution We provide criteria for a set of observed choices on nonlinear budget sets being generated by quasi-linear preferences that are not necessarily convex. Our criteria are based on the observation that when preferences are QL, they must also satisfy the property of *cyclical monotonicity* (see, e.g., Rochet, 1987). Our contribution lies in demonstrating that cyclical monotonicity holds for a larger class of choice environments than previously established. As a special case, we show that Forges and Minelli (2009)'s generalization of revealed preferences tests to nonlinear sets extends to tests of QL. This allows testing for QL preferences in some strategic environments. When the choice sets are linear, however, the condition we identify is equivalent to the definition of *cyclical monotonicity* in Brown and Calsamiglia (2007). These conditions are necessary and sufficient for choices to be rationalized by a QL utility function. We provide examples that show why it is important to consider nonlinear budget sets.

We also provide an empirical test for QL preferences in a laboratory experiment. As we discuss in the paper, the laboratory setting has advantages over consumer data in testing for QL preferences. In our experiment, we mimic different consumption groups by offering gift cards at discounted prices. We implement a five-good and a three-good treatment over 30 different budgets to test whether subjects' preferences can be rationalized by QL preferences. First, we find support for the Generalized Axiom of Revealed Preferences (GARP). For 21 out of 64 subjects in the five-good treatment and 34 out of 63 subjects in the three-good treatment we can reject the hypothesis of random behavior in favor of taking deliberate choices (i.e. GARP is satisfied). Second, regarding QL preferences, we find that the preferences of 8 out of 64 subjects in the five-good treatment and 12 out of 63 subjects in the three-good treatment are consistent with quasi-linearity. This accounts to about 40% of the subjects' preferences satisfying GARP. Remarkably, this proportion is similar for the five-good treatment and the three-good treatment. Taking a closer look at the consumption patterns we see that most of the subjects who are consistent with either GARP or QL preferences treat the goods as perfect substitutes.

Related Literature Revealed preference theory is attractive because of its robustness to functional form assumptions regarding preferences. Beginning with the work of Richter

(1966) and Afriat (1967), revealed preference theory has been used to test both individual and collective decision-making (see Chambers and Echenique, 2016, for a comprehensive overview of the results). It has also been use to test theories of consumer behavior in the lab. Applications include tests for social (e.g., Andreoni and Miller, 2002; Fisman et al., 2007; Porter and Adams, 2015), risk (e.g., Choi et al., 2007), and time preferences (e.g., Andreoni and Sprenger, 2012).

Relevant to our pursuit, some of these experiments have been conducted using real consumption goods. Cox (1997) uses the data collected by Battalio et al. (1973) to test the consistency of patients' behavior at a mental hospital. Mattei (2000) tests the consistency of college students' choices of snacks and stationery. Brocas et al. (2018) tests the consistency of older adults' and college students' choices of food items. In our study, we use gift cards that allow us to provide a set of goods that mimics the goods our subjects would choose to buy outside of the lab. Closer to our setting, Costa-Gomes et al. (2016) use real consumption goods (headphones). However, to provide added realism in our setting, we use choice bundles consisting of multiple goods rather than choices between different types of similar consumption goods.

Recently, there has been interest in revealed preference tests for QL preferences in the context of linear budgets. Brown and Calsamiglia (2007) propose a revealed preference test for the case of concave preferences, while Nocke and Schutz (2017) discuss the corresponding integrability problem without assuming concavity. Cherchye et al. (2015) extends Brown and Calsamiglia (2007)'s test to generalized QL while maintaining the concavity of the utility and the linearity of budgets. Allen and Rehbeck (2018) provide a measure for QL misspecification and use scanner data to evaluate how this misspecification varies with the level of aggregation of the data. They find that while quasi-linearity fails at the individual level, it represents the data well from the perspective of a representative agent. Chambers and Echenique (2017) show that QL preferences in the setting of combinatorial demand (with linear pricing) is equivalent to the law of demand, a condition that is simpler than cyclical monotonicity. As we will discuss, our approach provides the least restrictive test for QL preferences, and our experiment provides direct evidence that quasi-linearity may have limited empirical relevance at the *individual* level. Our experiments therefore complement these approaches nicely.

Structure The remainder of this paper is organized as follows. Section 2 presents our theoretical framework, Section 3 presents our experimental design and Section 4 presents the empirical results. Section 5 concludes the paper.

2 Theoretical Framework

In this section, we discuss the necessary and sufficient conditions for choices to be rationalized with QL preferences.

Consider a space of alternatives $Y = X \times \mathbb{R}$, where $X \subseteq \mathbb{R}^n_+$. We denote an element of this set as $(x, m) \in Y$, where $x \in X$ and $m \in \mathbb{R}$. Intuitively, x is a bundle of consumption goods and m is an amount of money. A QL utility function takes the following form:

$$v(x, m) = u(x) + m,$$

where $u : X \to \mathbb{R}$ is a real valued sub-utility function.

Let $E = ((x^t, m^t), B^t)_{t=1}^T$ be a consumption experiment, where (x^t, m^t) is the bundle chosen from budget $B^t \subseteq Y$ and T is the total number of decisions made. We assume that budgets are downward closed[1] and have compact borders and that there is a strictly decreasing function $m^t(x)$ such that $m^t(x^t) = m^t$.

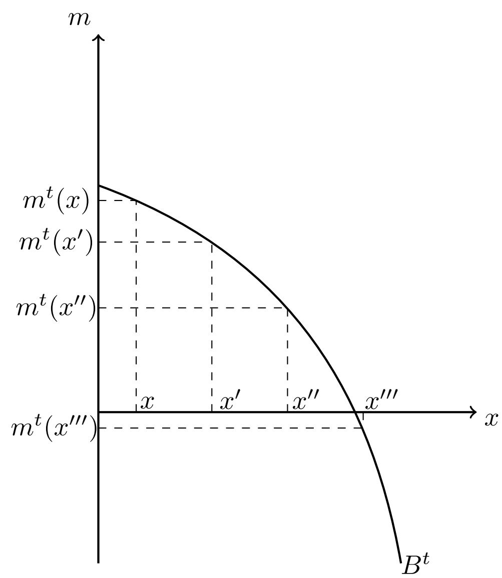


Figure 1: Obtaining $m^t(x)$.

Figure 1 illustrates how one can construct $m^t(x)$, the maximum value of m such that $(x, m^t(x)) \in B^t$. Note that $m^t(x)$ can be negative; indeed, for a high enough value of x, it *will* be negative. Assume that budget can be described by a function $g^t(x, m)$ such that $(x, m) \in B^t$ if and only if $g^t(x, m) \leq 0$.[2] We can then define $m^t(x)$ as follows:

$$m^t(x) = \underset{m \in \mathbb{R}}{\text{argmax}} \{m : g^t(x, m) \leq 0\}.$$

For the case of linear budgets, $m^t(x) = \frac{I - p_x x}{p_m}$, where I stands for income, p_x the price of the good(s) x, and p_m the price of money. We assume that $m^t = m^t(x^t)$ in order to avoid a violation of monotonicity in m. Since the budget is downward closed for every $x \in X$, $(x, m^t(x)) \in B^t$, we allow $m^t(x)$ to be negative, as this technical assumption simplifies the analysis. Intuitively one can think of $m^t(x)$ as a cost of bundle $x \in X$ given budget B^t. Note that we require $m^t(x)$ to be defined for every $x \in X$. If we consider m to be money, then a negative consumption of money can be interpreted as borrowing, and budgets being downward closed can be interpreted as the absence of a binding liquidity constraint.[3]

[1]That is, if $x \in B^t$, then $y \in B^t$ for every $y \leq x$.

[2]For instance, Forges and Minelli (2009) use the (Minkowski) gauge function to determine $g(x, m)$. Using the (Minkowski) gauge function requires extra assumptions we do not make since they are not crucial for our results.

[3]This assumption is technical and not crucial for our result. Imposing a liquidity constraint would require adding notation complicating the exposition without providing new insights. The assumption that m^t is well defined for every $x \in X$ is also not crucial for our result, we make these assumptions to simplify the exposition.

Definition 1. *A consumption experiment* $E = ((x^t, m^t), B^t)_{t=1}^T$ *can be* ***rationalized with QL preferences*** *if there is a function* $v(x,m) = u(x) + m$ *such that*

$$u(x^t) + m^t \geq u(x) + m \quad \text{for every } (x,m) \in B^t.$$

Definition 2. *A consumption experiment* $E = ((x^t, m^t), B^t)_{t=1}^T$ *satisfies* ***cyclical monotonicity*** *if for every sequence of observations* $k_1, \ldots, k_n \in \{1, \ldots, T\}$,

$$\sum_{k=1}^{n} m^{k_j}(x^{k_j}) - m^{k_{j+1}}(x^{k_j}) \geq 0,$$

where $k_{n+1} = k_1$.

Cyclical monotonicity was introduced by Rockafellar (1970), and its importance in characterizing QL was established by Rochet (1987), Brown and Calsamiglia (2007) and Nocke and Schutz (2017). The version of cyclical monotonicity we use is equivalent to the standard definition up to a reordering of the terms.

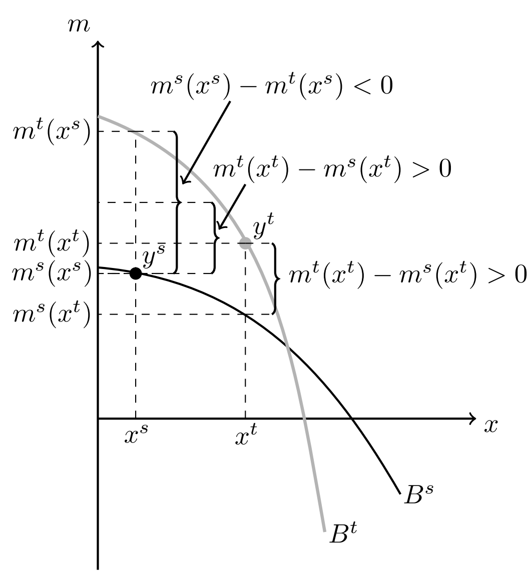


Figure 2: Data that are inconsistent with QLU rationalization.

Figure 2 shows an example of a violation of QL preferences although this set of data is rationalizable with a locally non-satiated utility function (LNU). To see this, suppose that the agent makes decisions according to QL preferences. Since y^s is chosen from B^s, it must be the case that the agent prefers y^s to $(x^t, m^s(x^t))$, and thus $(x^s, m^s(x^s) + [m^t(x^t) - m^s(x^t)])$ will be preferred to y^t (from QL). Given that $m^t(x^t) - m^s(x^t) \leq m^s(x^s) - m^t(x^s)$ we can conclude that $(x^s, m^s(x^s) + [m^t(x^t) - m^s(x^t)])$ is in the interior of B^t, y^t is revealed to be preferred instead, which is a contradiction. Such behavior is equivalent to subjects treating goods as perfect substitues. We now establish the main result for this section, whose proof is provided in Appendix A.

Theorem 1. *A consumption experiment* $E = ((x^t, m^t), B^t)_{t=1}^T$ *can be rationalized with QL preferences if and only if it satisfies cyclical monotonicity.*

We provide the intuition for the proof before we proceed further. The main step of the proof is to assign utility numbers to observed choices, that are then extended to the whole choice space using techniques developed in Afriat (1967). Since the necessity of cyclical monotonicity is clear, we concentrate on the argument of why cyclical monotonicity allows to find the proper utility values. We start by defining some graph-theoretic notions needed for the argument.

Consider a weighted directed complete graph[4] where each vertex corresponds to an observation. The weight assigned to each edge corresponds to the difference in the relative cost of bundles x^t and x^s in the budget B^t, that is $m^t(x^t) - m^t(x^s)$. Cyclical monotonicity guarantees that this graph does not contain negative cycles. Hence, we can assign a utility number to the point x^t that is equal to the weight of the shortest walk passing through x^t at the first transition. This construction guarantees that the chosen point is better than any other point in the budget. For this, we need to show that chosen values $u(x^t)$ satisfy

$$u(x^t) + m^t(x^t) \geq u(x^s) + m^t(x^s) \Leftrightarrow u(x^t) - u(x^s) \geq m^t(x^s) - m^t(x^t).$$

By construction $u(x^s)$ cannot exceed $u(x^t)+(m^t(x^s)-m^t(x^t))$, since both $u(x^s)$ and $u(x^t)$ are defined as shortest walks. This observation guarantees that QL rationalizability is satisfied for the space of chosen points. Appendix A shows that $u(x)$ can be assigned using the technique developed in Afriat (1967).

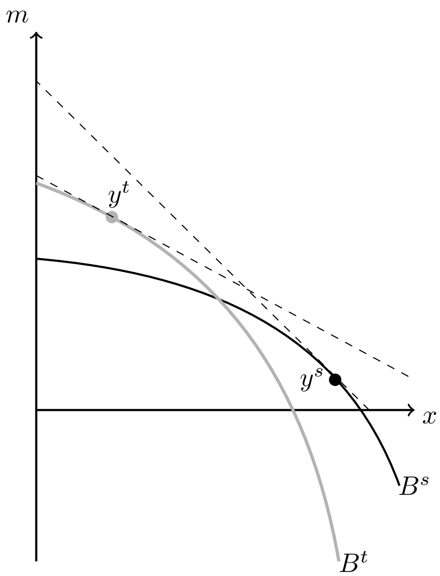
Before concluding the theoretical part, we make two remarks that illustrate the usefulness of our result. First, we will illustrate how restrictive the assumption of the concavity of the utility function is. Second, we will show that without additional assumptions regarding the structure of budgets, the requirement of cyclical monotonicity cannot be further simplified.

Figure 3 illustrates how restrictive the assumption of concavity of the utility function is when budgets are nonlinear. We use nonlinear budgets in this case because, as proven by Afriat (1967), concavity of the utility function has no empirical content when budgets are linear. Figure 3(a) shows that a consumption experiment can satisfy cyclical monotonicity, i.e., be QL rationalizable, but fail GARP for the linearized version of the budgets (see Matzkin, 1991). The dashed lines show the linearized budgets as presented in Forges and Minelli (2009), assuming that prices are constant and equal to the gradient of the gauge function of the budget set at the chosen point. Figure 3(b) shows that a consumption experiment can satisfy cyclical monotonicity, i.e., be QL rationalizable, but fail to have a *concave* QL rationalization. Cyclical monotonicity fails in the linearized budgets in the example since $m^t_\nabla(x^s)-m^s_\nabla(x^s) > m^t_\nabla(x^t)-m^s_\nabla(x^t)$, where m^t_∇ is the amount of m computed using the gradient of the border of the budget evaluated at the chosen point as the fixed price.

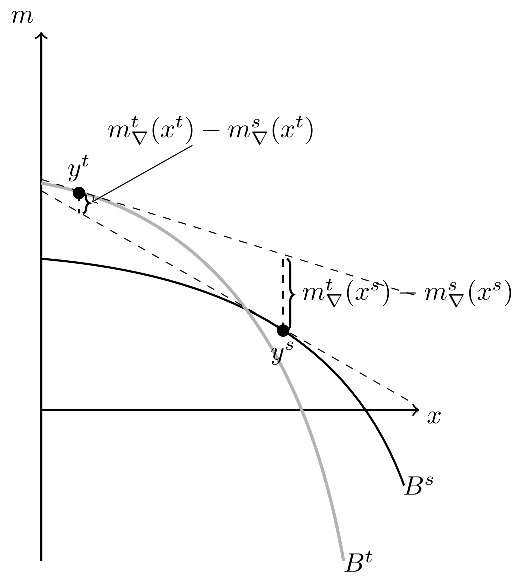
Remark 1. *A consumption experiment $E = ((x^t, m^t), B^t)_{t=1}^T$ can be rationalized by a concave QL utility function if and only if corresponding linearized experiment satisfies cyclical monotonicity.*

Finally, Figure 4 shows that transitivity plays a nontrivial role in testing QL when budgets are nonlinear. It is known that if Y is two-dimensional and the budgets are linear, then

[4] A graph is said to be complete if there is an edge between every pair of vertexes. The construction we use is similar to one in Piaw and Vohra (2003).



(a) Nonconcave utility



(b) Nonconcave QL utility

Figure 3: How restrictive is the assumption of concavity of u?

transitivity does not have testable implications (i.e., the weak axiom of revealed preferences [WARP] is equivalent to the strong axiom of revealed preferences [SARP]) (Rose, 1958). This result has been extended to the case of homothetic preferences by Heufer (2013) and to the case of QL by Chambers and Echenique (2017), who show that in the context of combinatorial demand, the law of demand is equivalent to cyclical monotonicity if the budgets are linear. The example in Figure 4 shows that, even in a two-good setting, transitivity is crucial for testing QL if the budgets are nonlinear. In the example, there are no pairwise violations of cyclical monotonicity since (1) $y^s = (x^s, m^s(x^s))$ lies on the intersection of B^r and B^s, while $y^r = (x^r, m^r(x^r))$ is not achievable with budget B^s; (2) $|m^t(x^t) - m^s(x^t)| = |m^s(x^s) - m^t(x^s)|$, since those segments of the budgets are parallel; and (3) $|m^t(x^t) - m^r(x^t)| > |m^r(x^r) - m^t(x^r)|$. However, if we consider the sequence r, t, s there is a violation of cyclical monotonicity since $m^r(x^r) - m^t(x^r) < 0$ and $|m^r(x^r) - m^t(x^r)| > |m^t(x^t) - m^s(x^t)|$, while $m^s(x^s) - m^r(x^s) = 0$. Hence, we can immediately infer the following remark.

Remark 2. *Let $E = ((x^t, m^t), B^t)_{t=1}^T$ be a consumption experiment. Unless budgets are linear, the law of demand is not sufficient for QL rationalization of the experiment.*

In sum, our test significantly generalizes previous results. Before we proceed further we make a couple remarks on the link of our results to the existent results in revealed preference theory which do not deal directly with quasilinearity. Note that the quasilinear preferences are a special case of additive separable preferences. The seminal work of Varian (1983) considers the concave case of additive separability and exploits standard Afriat (1967) technique in order to obtain a test for it. Varian's test relies crucially on the concavity of each of separable component of utility.[5] Recent work of Polisson (2018) relaxes this assumption

[5]In this sense the quasilinear analog of the Varian (1983) is the test developed by Brown and Calsamiglia (2007).

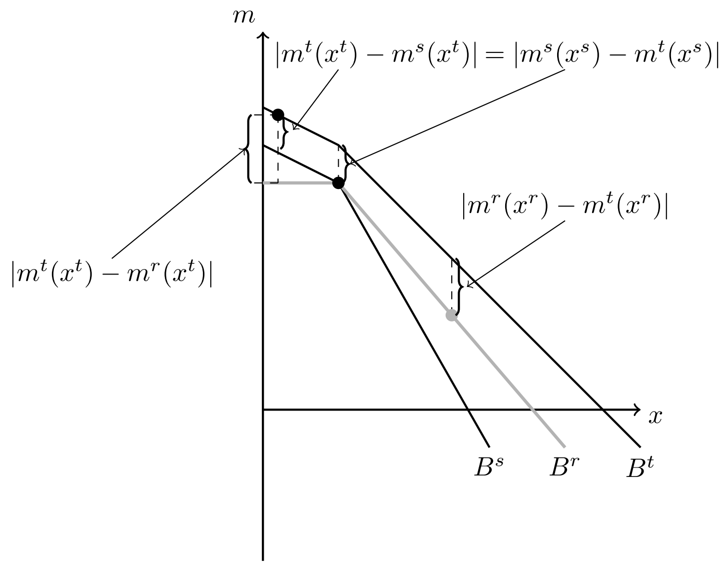


Figure 4: Violation of cyclical monotonicity in a 2-dimensional budget.

and provides a test for additive separability without concavity. However, this test requires each separable component to be one-dimensional though being unknown.[6] We allow for the first separable component to be multidimensional, while the second component has to be known and one dimensional. Hence, even when quasilinearity is a specific case of additive separability, the test we develop is not a special case of any existing tests for separable preferences.

3 Testing QL

In this section we present our experimental design for the test of QL preferences.

3.1 Experimental Design and Procedures

To test quasi-linearity, we designed an experiment in which subjects were asked to make consumption decisions that resembled day-to-day allocations of their budget. Specifically, we asked subjects to allocate 100 tokens among a fixed set of goods, including cash. We varied the cost (in tokens) of each good across the consumption categories. For instance, a subject might be able to exchange a token for 25 cents' worth of a good in one menu and for 10 cents' worth of the same good in another menu. Since these decisions can be complex, we implemented two treatments: a five-good treatment and a three-good treatment. In the five-good treatment, subjects had to allocate the 100 tokens across five goods (including cash), and in the three-good treatment, subjects had to allocate the tokens across three goods (including cash).[7] The goods used in the five-good experiment were cash, a Fandango

[6]The crucial idea behind the requirement that each component be one-dimensional is that the corresponding set is fully ordered. Hence, we treat these notions as synonyms throughout the discussion.

[7]We would like to thank Catherine Eckel for the suggestion to run the treatment with reduced amounts of goods.

gift card, a Barnes & Noble gift card, a Gap gift card, and Mason Money (described below). In the three-good treatment we included only cash, a Barnes & Noble gift card, and Mason Money. We selected these goods after consulting with students about their demand patterns.

To be more specific, the unit of measurement for each commodity was $1 (that is the subjects had to choose integer amounts) and subjects were asked to allocate 100 tokens between the goods mentioned above, whose prices were denominated as tokens per dollar value of the commodity. Each subject made choices for 30 menus, one of which was chosen at random at the end of the experiment to be implemented (i.e. subjects received the bundle they chose for this budget). To maximize the power of the test (see Choi et al. (2007)), the prices were also chosen at random for each subject. The experiment was conducted with 127 George Mason University undergraduates (64 subjects in the five-good treatment and 63 subjects in the three-good treatment).[8]

To bring the field to the lab, we chose goods that were needed by and familiar to the subjects. Fandango is a movie-theater ticket distributor and represents spending on entertainment. Barnes & Noble is a book distributor and represents spending on necessities, since textbooks are a required expenditure for students. The Gap is a clothing store and represents spending on durable goods. Mason Money is George Mason University's internal monetary system, which can be used at any on-campus restaurant, and therefore represents food spending.

We chose these goods after consulting with students about their purchase habits. The commodities were chosen to minimize consumption transaction costs, as high transaction costs might prompt subjects to trade the gift cards and therefore favor the hypothesis of rationality or QL in preferences—for instance, subjects facing high transaction costs might always choose the cards with the best trade-in value. However, money cannot be withdrawn from Mason Money cards (see Terms and Conditions), and it is not transferable since it is linked to the students' photo IDs, and, to the extent possible, we used gift cards bearing the subject's name to reduce the transfer of cards across individuals. We only have data on the usage of the Barnes & Noble cards. We found that most subjects used these cards within a month after the experiment.

Figure 5 shows the experimental interface, programmed using oTree (Chen et al., 2016). The interface allowed several allocations to be tried before progressing to the next menu and enforced the restriction that allocation cound never exceed 100 tokens. As mentioned earlier, the set of prices was randomly selected for each menu.

Table 1 presents descriptive statistics for the experimental data by treatment. The top panel shows the percentages of money allocated across the goods (dollars×tokens per dollar divided by the total expenditure of tokens subject made in current decision problem). We observe that cash was the highest-demand good and the Gap card was the lowest-demand good. These proportions are comparable across treatments conditional on the categories common to both treatments. The second panel of Table 1 shows the price ranges from which prices were chosen. The price ranges were the same for both treatments and ensured a high variation in prices across the 30 menus. One of the reasons for the heterogeneous price ranges (besides power) are the external constrains on the minimum purchase for gift cards which also make budgets to be nonlinear. In particular, the minimum purchases are $5 for Mason

[8]A complete explanation of the experimental design and procedures is provided in Appendix C.

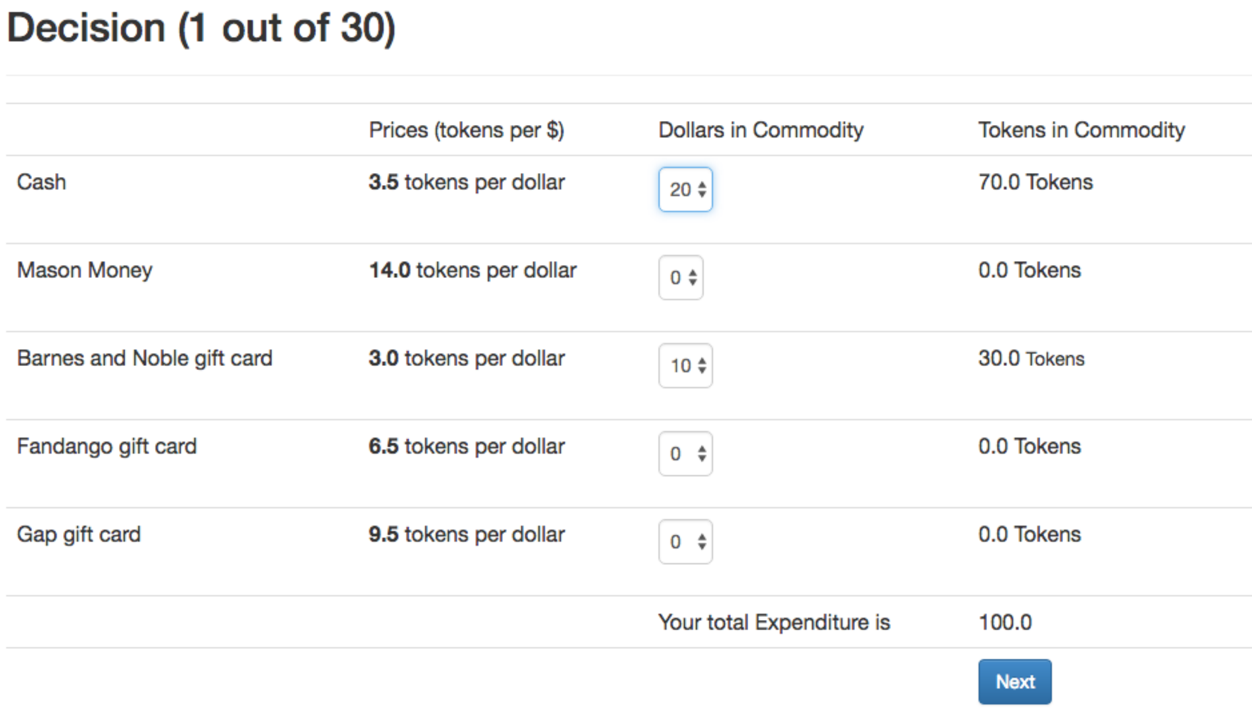


Figure 5: Experimental interface.

Money, \$10 for Barnes and Noble gift card and \$10 for Gap gift card. The choice set for Fandango is entirely discrete, in particular subjects could only choose between \$0, \$15, \$25, and \$35 value of gift card.

Given such external constraints the natural question arises, whether subjects managed to use the entire token budget. Subjects didn't always spend exactly 100 tokens, and the maximum waste observed was 10 tokens. The maximum losses in dollar equivalent terms did not exceed \$1. However, the interface is designed in a way that it would notify a subject if she can afford a unit of any good given her current bundle. We should remark that in our experiments, subjects could acquire a significant amount of money in each category. Specifically, the maximum amount of money in cash was \$40, for Fandango it was \$35, for Barnes & Noble it was \$40, for Mason Money it was \$40, and for the Gap it was \$33. The average earning in the experiment was \$19.06 (dollar equivalent).

The bottom panel of Table 1 shows the subjects' patterns of behavior across menus. First, we report the proportion of menus that had all of the tokens allocated to the cheapest good: 19.48% in the five-good treatment and 42.70% in the three-good treatment. Second, we report the proportion of menus that had all of the tokens allocated to cash: 36.20% in the five-good treatment and 48.10% in the three-good treatment. Next, we report the proportion of subjects that had all of the tokens allocated to one good only: 58.85% in the five-good treatment and 73.33% in the three-good treatment. Finally, we report the proportion of menus that exhibited choices across more than one good. In most of the cases, tokens were allocated to one or two goods. This is partially a consequence of the fact that some accounts had a minimum balance requirement. However, we should mention that our test of QL applies to cases with nonlinear budget sets such as these and has no restrictions

Relative expenditures on the goods		
	3 Goods	5 Goods
Cash	58.66%	47.10%
Mason Money	25.36%	15.80%
Barnes & Noble	15.98%	14.14%
Fandango	–	12.22%
Gap	–	10.73%
Price ranges (tokens per $ value)		
	min	max
Cash	2.5	15
Mason Money	2.5	15
Barnes & Noble	2.5	9.5
Fandango	2.5	6.5
Gap	3	10
Patterns of choice by menu		
	3 Goods	5 Goods
Chose cheapest	42.70%	19.48%
Cash only	48.10%	36.20%
One good	73.33%	58.85%
Two goods	23.39%	36.61%
Three goods	3.28%	4.48%

Table 1: Descriptive statistics.

regarding corner choices.

3.2 Assessing the severity of violations of rationality

To assess the prevalence of QL preferences, we need to adopt a behavioral benchmark for comparison. A natural benchmark is the existence of locally nonsatiated utility (LNU) that rationalizes the data. For the choice experiment, this is equivalent to satisfying GARP, which has been tested in several domains using experimental methods (Mattei, 2000; Harbaugh et al., 2001; Andreoni and Miller, 2002; Choi et al., 2007, 2014).

Because the test for QL is binary – either the test is passed or it is not – we need to consider the possibility that people might make mistakes. To account for this possibility, we use Afriat (1973)'s CCEI as a measure of distance to rationality.[9] For any $e \in (0,1]$, define $B^t(e) = \{y \in Y : \frac{y}{e} \in B^t\} \cup \{x^t\}$. The CCEI for GARP (QL) is the maximum $e \in (0,1]$ such that a consumption experiment $(y^t, B^t(e))$ is consistent with GARP (QL). Consider the case of linear budgets, as in Afriat (1973). We are searching for the maximum $e \in (0,1]$ such that given e, the (strict) revealed preference relation R (P) defined via $y^t R y$

[9]In Appendix B we also present the robustness check using the Houtman and Maks (1985) index as a measure of the distance to rationality. We abstain from using other measures since they are poorly defined in our context.

(y^tPy) holds if and only if $p^ty(<) \leq eI^t$, where p^t denotes prices and I^t denotes income. We use an equivalent definition that works in the context of nonlinear budget sets as well: we divide both sides of the inequality by e to obtain the equivalent definition for e's (strict) revealed preference relation $p^t\frac{y}{e}(<) \leq I^t$. The idea behind CCEI is to find the maximal e such that the experiment is consistent with GARP (QL). That is minimizing the share of budget wasted due to inconsistency.

Given that we apply the tests to finite data sets, we might obtain consistency with GARP/QL by chance, although, it is not a result of maximization of the corresponding type of the utility function. In order to control for such "false positives" we use Bronars (1987) measure of power. That is, we generate 1000 random subjects for each of the budget compositions (recall that each subject faced 30 random budgets) and compute the corresponding CCEI. These simulations generate the distribution of the CCEI levels under the null hypothesis of uniform random choices. Hence, we consider a subject to pass the test if her CCEI level is above the α percentile, where $1-\alpha$ is the corresponding confidence level. Finally, note that QL preferences is a theory *nested* within LNU. That is, a subject can only be consistent with QL preferences if she is consistent with LNU. Therefore, to control for "false positives" for QL preferences we use random subjects which are consistent with LNU. That is, we select a subsample of random subjects for which the CCEI for LNU is above α and conduct similar test for this subsample but using QL as criteria for rationality.[10]

4 Experimental Results

This section presents the results of the experiment. We consider two theories listed above, which are LNU and QL preferences. In addition, we test for a behavioral theory which we refer to as cash dominant rational (CD). CD is not a formal theory, but aims to capture rules-of-thumb observed in the data. Recall that one of the goods is Cash and other goods can be roughly classified as gift cards. That is, the rest of the goods, are restricted uses of cash sometimes offered at a discount. Hence, if there is (almost) no transaction costs of buying the gift cards, then cash is the dominant good over gift cards if prices are similar. That is, we assume a choice of a gift card instead of cash if the price of the gift card is higher to be a mistake absent transaction costs. We correct these 'mistakes' by reallocating the corresponding amount to the cash account. Note that if the price of the gift card was strictly higher than the price of cash, it would immediately generate the violation of GARP, because the total expenditure in the same budget would be strictly greater, hence, the subject made a dominated choice. In general, this correction could generate more violations of GARP if they cannot be represented by a LNU, i.e. if this was not an obvious mistake, but a deliberate consumption decision due to, for instance, non-trivial transaction costs of buying a gift card.

Table 2 presents the consistency results for the experimental data. The left panel presents the results for the 5 goods treatment and right panel presents results for the 3 goods treatment. The first column in each panel shows the exact pass rate, that is the share of subjects who are consistent with the test exactly. The second column in each panel shows the power corrected pass rate at $\alpha = .95$ that is the share of subjects for whom we can with 5%

[10]Exactly the same method has been used in Cherchye et al. (2019), although, the measure of rationality they used was different.

	5 Goods Treatment		3 Goods Treatment	
	Pass Rate	Pass Rate $\alpha = .95$	Pass Rate	Pass Rate $\alpha = .95$
LNU (unconditional)	25% (17/64)	33% (21/64)	33% (21/63)	54% (34/63)
95% conf. interval	(.16–.39)	(.22–.46)	(.22–.46)	(.41–.67)
CD (conditional on LNU)	41% (7/17)	100% (21/21)	100% (21/21)	100% (34/34)
95% conf. interval	(.18–.67)	–	(.22–.46)	–
QL (conditional on LNU)	29% (5/17)	38% (8/21)	24% (5/21)	35% (12/34)
95% conf. interval	(.10–.56)	(.18–.62)	(.08–.47)	(.20–.54)

Table 2: Consistency results for experimental data

confidence say that they are not taking decisions at random.[11] Each row presents the pass rate together with the confidence interval.[12] The first row contains results for LNU, the second and third rows contain the results for CD and QL. The results for CD and QL are conditional on passing LNU since both theories are nested within LNU. Therefore, for both nested theories we provide the analysis on the subsample of subjects consistent with LNU. In the case of pass rate, we select only subjects who pass GARP (exactly consistent with LNU) and for power corrected pass rates we select only subjects for whom we can reject the hypothesis that subjects are taking random choice in favor of LNU with $p \leq .05$.

Seventeen out of 64 subjects (25%) are consistent with LNU in the five-good treatment. In the three-good treatment, 21 out of 63 subjects (33%) are consistent with LNU. Seven out of 64 subjects (11%) are consistent with CD in the five-good treatment. In the three-good treatment, 21 out of 63 subjects (33%) are consistent with CD. Five out of 64 subjects (8%) are consistent with QL in the five-good treatment. In the three-good treatment, 5 out of 63 subjects (8%) are consistent with QL preferences. Considering the absolute numbers, we can make a claim that complexity of the task matters. In particular, it is particularly prominent for CD given that 10 out the 17 subjects who are consistent with LNU in the five-good treatment fail CD while in the three-good treatment all subjects who are consistent with LNU are consistent with CD.

Power corrected pass rates for LNU are 33% and 54% for the five-goods and three-gods treatments. Conditional power corrected pass rates for CD are 100% for both treatments. Conditional power corrected pass rates for QL are 38% for the five-goods treatments and 35% for the three-goods treatment. Hence, complexity appears to affect the levels of rationality especially for LNU, although, these differences are not statistically significant. To summarize the findings, we confirm the hypothesis of behavior consistent with (almost) no transaction costs of buying gift cards given the absolute conditional pass rates for CD. We also observe that the pass rates for LNU and QL preferences are not significantly different across treatments. Hence, the empirical validity of QL preferences is comparable to one of LNU.

[11] Appendix B contains the robustness checks for different confidence levels.

[12] Confidence intervals for the pass rates are obtained using the Clopper-Pearson procedure

4.1 Who is consistent?

We now take a closer look at the subjects who are consistent with LNU or QL preferences. First, we provide an illustrative example of the demand function of the experimental subjects. Second, we look at the expenditure shares of the subjects who are consistent with each of theories. In this section, we exclude CD since it produces results similar to LNU.

Figure 6 presents examples of the demand functions for different types of subjects for both the the three-good and the five-good treatments. The first two graphs show the demand function for cash for two subjects whose preferences are not consistent with GARP (CCEI $\leq .85$).[13] The second two graphs show the demand function for cash for two subjects whose preferences are consistent with GARP but not with quasi-linearity (CCEI for GARP $\geq .95$ and for QL $\leq .8$). The last two graphs show the demand function for cash for two subjects whose preferences are consistent with quasi-linearity (CCEI for QL $\geq .95$). Among the subjects whose preferences are not consistent with GARP, we can see that there is no clear demand function for cash. For those whose preferences are consistent with GARP but not with quasi-linearity, we can see that there are several possible demand functions, presumably with different slopes. Finally, for the subjects whose preferences are consistent with quasi-linearity, we see a unique demand function for cash. We should note that it is not the case that subjects whose preferences are consistent with quasi-linearity do not consume any goods except cash; rather, we observe that those subjects spend a larger proportion of their budget on cash than on any other commodity. In sum, we find a robust empirical regularity in that both GARP and QL preferences are well represented in the subjects' behavior. Half of the subjects whose preferences satisfy GARP also satisfy quasi-linearity.

	5 goods treatment			
	LNU		QL	
	Exact Pass	Power Corrected Pass	Exact Pass	Power Corrected Pass
Cash	61.8% (35.9%)	62.5% (34.1%)	100.0% (0.0%)	90.9% (16.6%)
Mason Money	9.4% (10.5%)	11.6% (12.1%)	0.0% (0.0%)	9.0% (16.7%)
Barnes & Noble	11.2% (14.4%)	10.7% (13.8%)	0.0% (0.0%)	0.5% (0.4%)
Fandango	7.2% (10.4%)	6.8% (10.1%)	0.0% (0.0%)	0% (0%)
Gap	10.3% (15.7%)	8.4% (14.6%)	0.0% (0.0%)	0% (0%)
One good	88.6%	86.7%	100.0%	99.2%
	3 goods treatment			
	LNU		QL	
	Exact Pass	Power Corrected Pass	Exact Pass	Power Corrected Pass
Cash	59.6% (25.9%)	59.6% (23.8%)	79.7% (20.7%)	72.3% (22.4%)
Mason Money	26.6% (17.8%)	26.5% (16.6%)	20.3% (20.7%)	26.7% (22.5%)
Barnes & Noble	13.8% (16.6%)	13.9% (14.7%)	0.0% (0.0%)	1.0% (1.9%)
One good	97.9%	87.8%	99.3%	90.8%

Table 3: Expenditure shares for consistent subjects

Table 3 presents the relative expenditure shares for different goods for subjects consistent

[13]Since GARP is equivalent to the existence of locally nonsatiated utility fuction (LNU), we use these terms interchangeably.

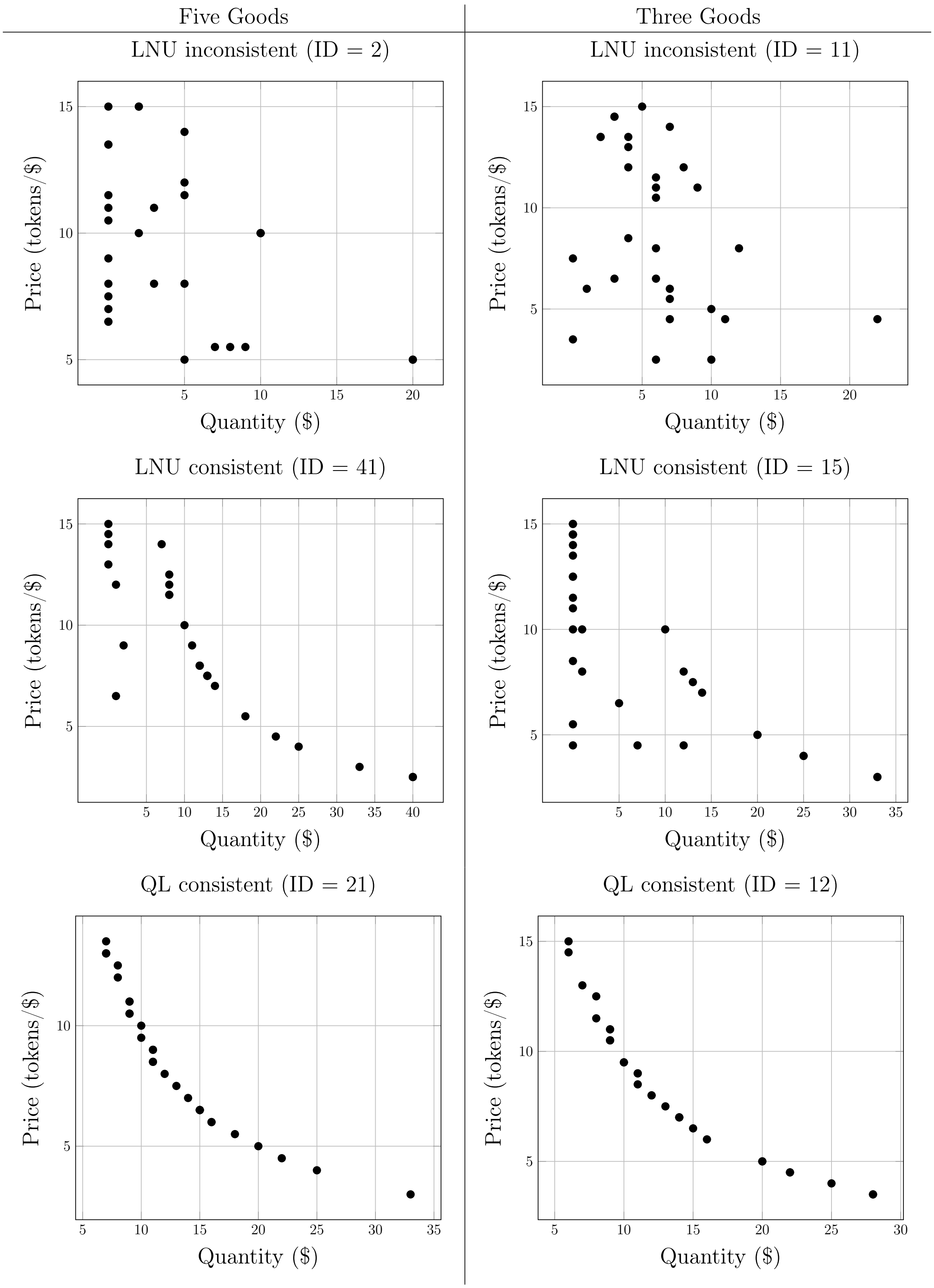


Figure 6: Examples of demand functions for cash.

with QL preferences or LNU, where each row presents the relative expenditure share for each of the corresponding goods. The last row of every panel presents the share of budgets in which subjects selected one good per budget, that is spent the entire endowment on the single good. The relative expenditure shares are computed the same way as in the top panel of Table 1, hence it makes sense to compare these numbers in addition to just studying the observed patterns in the data. The top panel corresponds to the five-goods treatment and the bottom panel corresponds to the three-goods treatment. The left panel presents results for the subjects who are consistent with LNU and the right panel presents the results for subjects who are consistent with QL preferences. We present results for subjects who pass the exact test and for subjects who pass the power corrected test at the 5% level.

In the five-goods treatment, subjects consistent with LNU spend more on Cash than in the general sample ($\approx 62\%$ for consistent vs $\approx 47\%$ in general). For instance, the expenditure on the Fandango gift card is higher than on the Gap gift card in the entire sample, while this relation is reversed for subjects consistent with LN preferences. The frequency of choosing a single good per budget is higher for LNU than for in the entire sample ($\approx 87\%$ for LNU consistent vs $\approx 59\%$ for the general sample). The choice pattern of subjects consistent with QL preferences is clearly concentrated on choosing Cash only ($90-100\%$) and almost always choosing only one good per budget (at least 99% of the cases).

Next we analyze the results for the three-goods treatment. Expenditure shares for subjects consistent with LNU are similar to those in the general sample. However, the frequency of allocating all the token endowment to a unique good is much higher ($\approx 87-97\%$ for LNU consistent vs $\approx 73\%$ for the general sample). Expenditure shares for subjects consistent with QL preferences are higher for cash ($\approx 72-79\%$ for QL~~U~~ consistent vs. 59% overall). Their frequency of assigning all the endowment to a unique good is significantly higher than in the general sample ($\approx 90-99\%$ for QL consistent vs. $\approx 73\%$ overall).

To sum up, in the five-goods treatment, subjects who are consistent QL preferences are not only radically different from the entire sample but also from subjects consistent with LNU. They almost always allocate the entire endowment to Cash and almost always only purchase only one good. Subjects consistent with QL preferences are less different from LNU consistent subjects in the three-goods treatment. While subjects consistent with QL preferences prefer cash more than other goods, the frequency with which they allocate the entire endowment to a unique good is comparable to that of subjects consistent with LNU. The common feature among consistent (both LNU and QL) subjects in both treatments is that they are prone to allocate the entire endowment to the single good than the general sample.

5 Concluding Remarks

We provide necessary and sufficient condition for a set of observed choices to be rationalizable with QL preferences. This condition applies to choices over compact and downward closed budget sets and does not require the utility function to be concave. We conduct the first laboratory experiment to test whether preferences are QL. We find that about half of the subjects whose preferences satisfy GARP are also consistent with QL preferences. This result is robust to the number of goods among which subjects have to choose.

An important advantage of the test of QL that we present here is that it applies to a large class of problems. First, it applies to consumer problems in the presence of distortions due to taxes, subsidies, or nonlinear pricing. Second, it can also be extended to a test for QL preferences in strategic situations where an assumption of QL preferences might be invoked (e.g., auction theory). Regarding our empirical evidence, we show that under tight experimental conditions, QL preferences are empirically relevant in individual decision-making. Our field-in-the-lab experiment hints at behavioral rules used by subjects that produce behavior close to rational. These rules might have empirical content outside the lab and deserve further research.

References

Afriat, S.N., 1967. The construction of utility functions from expenditure data. International Economic Review 8, 67–77.

Afriat, S.N., 1973. On a system of inequalities in demand analysis: an extension of the classical method. International Economic Review , 460–472.

Allcott, H., Taubinsky, D., 2015. Evaluating behaviorally motivated policy: experimental evidence from the lightbulb market. American Economic Review 105, 2501–2538.

Allen, R., Rehbeck, J., 2018. Assessing misspecification and aggregation for structured preferences .

Andreoni, J., Miller, J., 2002. Giving according to garp: An experimental test of the consistency of preferences for altruism. Econometrica 70, 737–753.

Andreoni, J., Sprenger, C., 2012. Estimating time preferences from convex budgets. American Economic Review 102, 3333–56.

Battalio, R.C., Kagel, J.H., Winkler, R.C., Fisher Jr, E.B., Basmann, R.L., Krasner, L., 1973. A test of consumer demand theory using observations of individual consumer purchases. Economic Inquiry 11, 411–428.

Becker, G.S., 1974. A theory of social interactions. Journal of Political Economy 82, 1063–1093.

Bergstrom, T.C., Cornes, R.C., 1983. Independence of allocative efficiency from distribution in the theory of public goods. Econometrica: Journal of the Econometric Society , 1753–1765.

Brocas, I., Carrillo, J.D., Combs, T.D., Kodaverdian, N., 2018. Consistency in simple vs. complex choices by younger and older adults. Journal of Economic Behavior & Organization .

Bronars, S.G., 1987. The power of nonparametric tests of preference maximization. Econometrica: Journal of the Econometric Society , 693–698.

Brown, D.J., Calsamiglia, C., 2007. The nonparametric approach to applied welfare analysis. Economic Theory 31, 183–188.

Chambers, C.P., Echenique, F., 2016. Revealed preference theory. volume 56. Cambridge University Press.

Chambers, C.P., Echenique, F., 2017. A characterization of combinatorial demand. Mathematics of Operations Research .

Chen, D.L., Schonger, M., Wickens, C., 2016. otree—an open-source platform for laboratory, online, and field experiments. Journal of Behavioral and Experimental Finance .

Cherchye, L., Demuynck, T., De Rock, B., 2015. Is utility transferable? a revealed preference analysis. Theoretical Economics 10, 51–65.

Cherchye, L., Demuynck, T., De Rock, B., Freer, M., 2019. Revealed preference analysis of expected utility maximization under prize-probability trade-offs. FEB Research Report Department of Economics DPS19. 13 .

Choi, S., Fisman, R., Gale, D., Kariv, S., 2007. Consistency and heterogeneity of individual behavior under uncertainty. American economic review 97, 1921–1938.

Choi, S., Kariv, S., Mueller, W., Silverman, D., 2014. Who Is (More) Rational? American Economic Review 104, 1518–1550.

Costa-Gomes, M., Cueva, C., Gerasimou, G., Tejiscak, M., 2016. Choice, deferral and consistency .

Cox, J.C., 1997. On testing the utility hypothesis. The Economic Journal 107, 1054–1078.

Domencich, T.A., McFadden, D., 1975. Urban travel demand-a behavioral analysis. Technical Report.

Fisman, R., Kariv, S., Markovits, D., 2007. Individual preferences for giving. The American Economic Review 97, 1858–1876.

Forges, F., Minelli, E., 2009. Afriat's theorem for general budget sets. Journal of Economic Theory 144, 135–145.

Green, J., Laffont, J.J., 1977. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. Econometrica: Journal of the Econometric Society , 427–438.

Harbaugh, W.T., Krause, K., Berry, T.R., 2001. GARP for kids: On the development of rational choice behavior. American Economic Review 91, 1539–1545.

Heufer, J., 2013. Testing revealed preferences for homotheticity with two-good experiments. Experimental Economics 16, 114–124.

Houtman, M., Maks, J., 1985. Determining all maximal data subsets consistent with revealed preference. Kwantitatieve methoden 19, 89–104.

Kleven, H.J., 2016. Bunching. Annual Review of Economics 8, 435–464.

Krishna, V., 2009. Auction theory. Academic press.

Mattei, A., 2000. Full-scale real tests of consumer behavior using experimental data. Journal of Economic Behavior & Organization 43, 487–497.

Matzkin, R.L., 1991. Axioms of revealed preference for nonlinear choice sets. Econometrica: Journal of the Econometric Society , 1779–1786.

Myerson, R.B., 1981. Optimal auction design. Mathematics of Operations Research 6, 58–73.

Nocke, V., Schutz, N., 2017. Quasi-linear integrability. Journal of Economic Theory 169, 603–628.

Piaw, T.C., Vohra, R.V., 2003. Afriat's theorem and negative cycles. preprint .

Polisson, M., 2018. A lattice test for additive separability .

Porter, M., Adams, A., 2015. For love or reward? characterising preferences for giving to parents in an experimental setting. The Economic Journal 126, 2424–2445.

Richter, M.K., 1966. Revealed preference theory. Econometrica: Journal of the Econometric Society , 635–645.

Rochet, J.C., 1987. A necessary and sufficient condition for rationalizability in a quasi-linear context. Journal of mathematical Economics 16, 191–200.

Rockafellar, R.T., 1970. Convex Analysis. Princeton University Press, Princeton, NJ.

Rose, H., 1958. Consistency of preference: the two-commodity case. The Review of Economic Studies 25, 124–125.

Varian, H.R., 1983. Nonparametric tests of models of investor behavior. Journal of Financial and Quantitative Analysis 18, 269–278.

A Proof of Theorem 1

Proof. (⇒) Consider a function $v(x,m) = u(x) + m$ that rationalizes the consumption experiment. For clarity, we maintain the notation $m^t(x^s)$, $m^t(x^t)$ and $m^t(x)$ for the rest of the proof. Then, for every sequence of observations $k_1, k_2, \ldots, k_n \in \{1, \ldots, T\}$, the following is true:

$$u(x^{k_{j+1}}) + m^{k_{j+1}}(x^{k_{j+1}}) \geq u(x^{k_j}) + m^{k_{j+1}}(x^{k_j}),$$

where $k_{n+1} = k_1$. We can simplify this and obtain the following inequalities:

$$\begin{cases} m^{k_1}(x^{k_1}) - m^{k_1}(x^{k_2}) \geq u(x^{k_2}) - u(x^{k_1}), \\ m^{k_2}(x^{k_2}) - m^{k_2}(x^{k_3}) \geq u(x^{k_3}) - u(x^{k_2}), \\ \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \cdots \\ m^{k_n}(x^{k_n}) - m^{k_n}(x^{k_1}) \geq (u(x^{k_1}) - u(x^{k_n}). \end{cases}$$

Summing up these inequalities, we obtain the following:

$$m^{k_1}(x^{k_1}) - m^{k_1}(x^{k_2}) + m^{k_2}(x^{k_2}) - m^{k_2}(x^{k_3}) + \ldots + m^{k_n}(x^{k_n}) - m^{k_n}(x^{k_1}) \geq 0.$$

This is exactly the cyclical monotonicity condition.

(⇐) This part of the proof is split in two parts. First, we construct the subutility numbers corresponding to the observed choices and show that the observed chosen point (x^t, m^t) is at least as good as any other point $(x^s, m^t(x^s))$. Next, we extend the subutility function to the entire $\mathbb{R}^n_+$ and show that corresponding QLU rationalizes the data.

Let

$$u^t = \min\{m^{k_1}(x^{k_1}) - m^{k_1}(x^t) + \ldots + m^{k_n}(x^{k_n}) - m^{k_n}(x^{k_{n-1}})\}$$

over all sequences in the data, including those with repeating elements.[14] We will show that the minimum is well-defined. It will be enough to show that there will be no cycles in the minimal sequence, since the rest will follow from the fact that the minimum is taken over finite sums of finite numbers. So assume, to the contrary, that the minimum sequence contains a cycle; then we have

$$\ldots + \left(m^{k_n}(x^{k_n}) - m^{k_n}(x^{k_{n-1}}) + \ldots + m^{k_1}(x^{k_1}) - m^{k_1}(x^{k_n})\right) + \ldots.$$

However, cyclical monotonicity implies that this term is greater or equal than zero, and hence, excluding it would make the sequence even smaller. That contradicts the original assumption that the sequence was the smallest.

Further, we show that such a construction of u^t will guarantee that the following system of inequalities is satisfied. For any $t, s \in \{1, \ldots, T\}$ we want to show that

$$u^t - u^s \geq m^t(x^s) - m^t(x^t).$$

[14] In terms of graph theory, this is equivalent to searching for the shortest walk (on a weighted directed graph), rather than a shortest path (on a weighted directed graph). However, the shortest walk is not always well-defined. A sufficient condition for it to be well-defined is the absence of negative cycles, which is guaranteed by cyclical monotonicity if we define the complete directed graph with vertexes as observations and weights as $w_{s\to t} = m^t(x^t) - m^t(x^s)$.

By the construction of $u(x)$, we can guarantee that

$$u^s \leq m^t(x^t) - m^t(x^s) + u^t,$$

since

$$u^t = m^{k_1}(x^{k_1}) - m^{k_1}(x^t) + \ldots + m^{k_n}(x^{k_n}) - m^{k_n}(x^{k_{n-1}})$$

for some (minimal) sequence, and we construct $u(x^s)$ using that minimal sequence. Recall that we allowed taking every sequence, including those with repeating elements, so we can add any element to the existing sequence and the utility level $u(x^s)$ will not exceed the value of the new extended sequence. Therefore,

$$u^t - u^s \geq u^t - \big(m^t(x^t) - m^t(x^s) + u(x^t)\big) = m^t(x^s) - m^t(x^t).$$

This concludes the first part of the proof. Next, we use the numbers constructed above to recover the entire utility function.

For every $x \in X$, let

$$u(x) = \min_{t \in \{1,\ldots,T\}} \{u^t + m^t(x^t) - m^t(x)\}.$$

Note that since m^t is continuous and monotone, so is $u(x)$.

First, we will show that for every $t \in \{1, \ldots, T\}$, $u(x^t) = u^t$. As we have shown above for every $s \in \{1, \ldots, T\}$ for which $m^s(x^t)$ is defined,

$$u^t \leq u^s + m^s(x^s) - m^s(x^t).$$

Therefore, $u(x^t) = u^t$. Next, we show that the constructed utility rationalizes the data, that is $v(x, m) \leq v(x^t, m^t(x^t))$ for every $(x, m) \in B^t$, where $v(x, m) = u(x) + m$. By the construction of $u(x)$, we know that $u(x) = \min_{t \in \{1,\ldots,T\}} \{u^t + m^t(x^t) - m^t(x)\} \leq u^t + m^t(x^t) - m^t(x)$. Therefore, $u(x) + m \leq u(x) + m^t(x) \leq u(x^t) + m^t(x^t)$. □

Given the proof of the Theorem 1 one can immediately infer the following remark.

Remark 3. *A consumption experiment $E = ((x^t, m^t), B^t)_{t=1}^T$ can be rationalized by a QLU function if and only if there is a monotone function $u : C \to \mathbb{R}$ such that*

$$u(x^t) - u(x^s) \geq m^t(x^s) - m^t(x^t) \text{ for every } t, s \in \{1, \ldots, T\}.$$

where $C = \bigcup_{t \in \{1,\ldots,T\}} \{x^t\}$ is the set of chosen points.

Remark 3 crucially important for the following reason. Cyclical monotonicity is an elegant condition, it may be less computationally tractable especially once mixed with measures of distance to rationality. However, the linear programming approach provides the definitely computationally tractable way of testing QLU rationalizability.

B Empirical Robustness Checks

This section provides additional information and robustness checks of the empirical analysis in the main body of the paper. As an alternative to the Critical Cost Efficiency Index (CCEI) (introduced by Afriat, 1973), we use the Houtman and Maks (1985) index (HMI). The HMI is the largest subset of the data that is consistent with a given axiom. Hence, the HMI is computed as the proportion of budgets over which a subject satisfies an axiom. In addition, we use power-corrected pass rates corresponding to the HMI index. Power adjusted pass rates are computed as follows. First, we compute the distribution of Bronars (1987) subjects. Next, we define the threshold based on the empirical distribution of the measure of rationality for these random subjects. That is we take the corresponding α cut off at the empirical distribution. This allows us to reject the hypothesis that the given subject randomly passed the test with a 90%, 95% and 99% chance correspondingly.

B.1 CCEI

	5 Goods Treatment			
	Exact pass rate	$p \leq .10$	$p \leq .05$	$p \leq .01$
LNU (unconditional)	25% (17/64)	34% (22/64)	33% (21/64)	27% (17/64)
95% confidence interval	(.16 – .39)	(.23 – .47)	(.22 – .46)	(.16 – .39)
CD (conditional)	41% (7/17)	100% (22/22)	100% (21/21)	100% (17/17)
95% confidence interval	(.18 – .67)	–	–	–
QL (conditional)	29% (5/17)	36% (8/22)	38% (8/21)	41% (7/17)
95% confidence interval	(.10 – .56)	(.17 – .59)	(.18 – .62)	(.18 – .67)
	3 Goods Treatment			
	Exact pass rate	$p \leq .10$	$p \leq .05$	$p \leq .01$
LNU (unconditional)	33% (21/63)	64% (40/63)	54% (34/63)	40% (25/63)
95% confidence interval	(.22 – .46)	(.50 – .75)	(.41 – .67)	(.28 – .53)
CD (conditional)	100% (21/21)	100% (40/40)	100% (34/34)	96% (24/25)
95% confidence interval	–	–	–	(.80 – .99)
QL (conditional)	24% (5/21)	38% (15/40)	35% (12/34)	48% (12/25)
95% confidence interval	(.08 – .47)	(.23 – .54)	(.20 – .54)	(.28 – .69)

Table 4: Pass rates according to CCEI

Table 4 presents the exact and the power adjusted pass rates. The top panel presents the result for the five-goods treatment and the bottom panel presents the results for the three-goods treatment. Within each panel the top row presents the results for LNU (locally non-satiated utility), CD (cash dominant preferences) and QL (quasilinear preferences). The columns present the exact and power adjusted pass rates. These results confirm those presented in the main body of the paper. The first result is that cash dominance is a feasible assumption and is valid for every subject who is consistent with LNU. Second, we confirm that LNU is an empirically relevant theory. Third, we show that the QL is as empirically relevant as LNU. Finally, we confirm behavior akin to choice overload as we increase the number of goods. That is, subjects adhere to rationality less frequently in the five-good treatment than in the three-good treatment.

	5 Goods Treatment			
	Exact pass rate	$p \leq .90$	$p \leq .95$	$p \leq .99$
LNU	88.6%	86.2%	86.7%	88.6%
QL	100%	99.2%	99.2%	99.5%
	3 Goods Treatment			
	Exact Pass rate	$p \leq .90$	$p \leq .95$	$p \leq .99$
LNU	97.9%	85.3%	87.8%	90.9%
QL	99.3%	88.4%	90.8%	90.8%

Table 5: Percent of budgets with purchases in only one good for CCEI consistent subjects

Table 5 presents expenditure shares for consistent subjects at the different significance levels. The top panel presents results for the five-goods treatment and the bottom panel presents results for the three-goods treatment. The first row of each panel presents results for LNU and the second row presents results for QL. We omit results for CD since they are similar to LNU. Given the results obtained in the main body of the paper we concentrate on the proportion of budgets in which subjects buy only one good. Such behavior is equivalent to subjects treating goods as perfect substitutes. The analysis of expenditure shares also confirm the findings in the main body of the paper. First, both LNU and QL consistent subjects frequently (over 80% of the time) allocate the entire budget to a single good. In the five-goods treatment we see that, regardless of the consistency criteria used, QL consistent subjects almost always spend the entire budget on a single good (99-100%), while the corresponding shares are smaller for LNU consistent subjects (<90%). In the three-goods treatment the shares of LNU and QL consistent subjects are similar (85-91%).

B.2 HMI

Table 6 presents the exact and power adjusted pass rates using the HMI index. The top panel corresponds to the five-goods treatment and the bottom panel corresponds to the three-goods treatment. The columns present the exact and the power adjusted pass rates. Note that using a different index of distance to rational behavior does not affect the results qualitatively, but quantitatively. The power adjusted pass rates for QL and LNU are significantly higher. However, we still observe similar patterns in the data. For instance, the pass rates in the three-goods treatment are higher than those in the five-goods treatment albeit marginally so. We also observe that every subject consistent with LNU is also consistent with CD. Our results are therefore robust to the measure of distance to rationality we use.

Table 7 presents the expenditure shares for consistent subjects using the HMI index. Given the results obtained in the main body of the paper, we concentrate on the share of budgets in which subjects spend only on one good. Such behavior is equivalent to treating all goods as perfect substitutes.

Unlike the analysis using the CCEI, we observe that in the five-goods treatment the frequency of budgets with purchases in only one good is smaller. This is expected given that

	5 Goods Treatment			
	Exact pass rate	$p \leq .10$	$p \leq .05$	$p \leq .01$
LNU (unconditional)	25% (17/64)	61% (39/64)	55% (35/64)	42% (27/64)
95% confidence interval	(.16 – .39)	(.48 – .73)	(.42 – .67)	(.30 – .55)
CD (conditional)	41% (7/17)	100% (39/39)	97% (34/35)	100% (27/27)
95% confidence interval	(.18 – .67)	–	–	–
QL (conditional)	29% (5/17)	74% (29/39)	74% (26/35)	78% (21/27)
95% confidence interval	(.10 – .56)	(.58 – .87)	(.57 – .88)	(.58 – .92)
	3 Goods Treatment			
	Exact pass rate	$p \leq .10$	$p \leq .05$	$p \leq .01$
LNU (unconditional)	33% (21/63)	70% (44/63)	67% (42/63)	62% (39/63)
95% confidence interval	(.22 – .46)	(.60 – .81)	(.54 – .78)	(.49 – .74)
CD (conditional)	100% (21/21)	100% (44/44)	100% (42/42)	100% (39/39)
95% confidence interval	–	–	–	–
QL (conditional)	24% (5/21)	84% (37/44)	83% (35/42)	82% (32/39)
95% confidence interval	(.08 – .47)	(.70 – .99)	(.69 – .93)	(.67 – .93)

Table 6: Pass rates according to HMI

	5 Goods Treatment			
	Exact pass rate	$p \leq .90$	$p \leq .95$	$p \leq .99$
LNU	88.6%	74.8%	77.5%	77.5%
QL	100%	62.1%	62.1%	62.1%
	3 Goods Treatment			
	Exact Pass rate	$p \leq .90$	$p \leq .95$	$p \leq .99$
LNU	97.9%	85.8%	86.7%	88.4%
QL	99.3%	84.0%	84.0%	84.0%

Table 7: Percent of budgets with purchases in only one good for HMI consistent subjects

more subjects are classified as rational using the HMI index The same applies to subjects consistent with QL who have about 62% of their budgets with purchases in one good only. Interestingly, these shares are lower than the corresponding numbers for LNU (62% for QL and 75-77% for LNU). The results for the three-goods treatment are comparable to those obtained before. In particular, we observe that the frequency with which subjects spend the entire budget in a single good is about 84-88% for both QL and LNU consistent subjects according to the HMI, while the corresponding numbers according to the CCEI are 85-91%. Hence, we can conclude that while the result that the behavioral pattern is the same for QL and LNU consistent subjects is robust to the change of index.

C Experimental Instructions (Online Only)

The experiment consisted of 30 independent decision problems. In each of these, subjects were asked to allocate tokens among five accounts (commodities): “Cash," “Mason Money," “Barnes & Noble gift card," “Fandango gift card," and “Gap gift card." There were additional restrictions due to the companies' restrictions:

- The minimum positive balance for Mason Money was $5.
- The minimum positive balance for the Barnes & Noble gift card was $10.
- The minimum positive balance for the Gap gift card was $10.
- For the Fandango gift card, the balance could be one of: $0, $15, $25, $35.

The resolution for each commodity (except the Fandango gift card) was $1.

Before starting the experiment, instructions were read out aloud after each subject was given a paper copy of the instructions. All subjects took a quiz that tested their understanding of the decision-making task. At the end of the experiment, one of the decision problems was randomly chosen (from the discrete uniform distribution) and subjects were paid according to the decisions they made for that problem. We sent e-gift cards that could be used immediately to the subjects' official GMU e-mail addresses. Cash choices were also paid at the end of the experiment, and Mason Money was sent to their Mason Money accounts.

The experiment was implemented using oTree (Chen et al. (2016)), and the demo version of the experiment is available at `https://qlt.herokuapp.com`. Next, we present the instructions for the five-goods treatment and, after that, the instructions for the three-goods treatment.

INSTRUCTIONS

Thank you for participating in today's experiment. Please remain silent during the experiment. If you have any questions, please raise your hand and the experimenter will assist you in private.

This is an experiment in individual decision-making. Your earnings from the experiment will depend in part on your decisions and on chance. Your earnings will **not** depend on the decisions of other participants. Please pay careful attention to the instructions as a considerable amount of money is at stake.

Your payment in today's experiment will not be lower than $15 equivalent. This will be paid to you at the end of the experiment in private.

You will face 30 decision problems. In each decision problem, you will be given 100 tokens to be divided among 5 commodities. The five commodities are:

- **Cash**

  You can choose any (integer) dollar amount of this commodity. The amount of your choice will be paid to you at the end of experiment.

- **Mason Money**

  This is a prepaid debit program that provides a fast, safe and convenient way to make purchases on and off campus. Mason Money is accepted at all cafes and restaurants on campus and is linked to your Mason ID.

- **Barnes and Noble gift card**

  Barnes and Noble is the largest retail bookseller in the United States, and a leading retailer of content, digital media and educational products in the country.

- **Fandango gift card**

  Fandango allows to buy tickets to more than 26,000 theaters nationwide. It is available online, and through their mobile and connected television apps.

- **Gap gift card**

  Gap is a US-based multinational clothing and accessories retailer. The card can be used in any store or the online store.

You will be asked to allocate your 100 tokens to each of these commodities for 30 different sets of prices for each commodity. Prices will range from 2.5 tokens per dollar to 15 tokens per dollar.

Prices are set up in the way that you can always buy an equivalent of $15. For instance, suppose that each commodity price is 5 tokens per dollar, then you can purchase $20 in cash, or $4 in each commodity.

Due to the restrictions set by companies on gift cards **additional restrictions** apply**:**

- If you purchase any positive amount of Mason Money, you should purchase at least $5.
- If you purchase any positive amount of Barnes and Noble gift card, you should purchase at least $10
- If you purchase any positive amount of Gap gift card, you should purchase at least $10.
- For Fandango gift card, you should purchase one of the following amounts: $0, $15, $25, $35.

Sample Screenshot:

Decision (1 out of 30)

	Prices (tokens per $)	Dollars in Commodity	Tokens in Comodity
Cash	**6.0** tokens per dollar	0	0.0 Tokens
Mason Money	**11.5** tokens per dollar	0	0.0 Tokens
Barnes and Noble gift card	**5.0** tokens per dollar	0	0.0 Tokens
Fandango gift card	**4.0** tokens per dollar	0	0.0 Tokens
Gap gift card	**7.5** tokens per dollar	0	0.0 Tokens
		Your total Expenditure is	0.0
			Next

Figure 1. Decision Screen

Figure 1 is an example of the decision screen. The first column contains the list of commodities. The second column contains the list of prices for each commodity (in tokens per dollar). The third column contains the quantity of commodity you will be asked to choose. This is the column you will use to make your decisions. The fourth column shows the allocation of your tokens among commodities ($Quantity\ of\ commodity * Price\ of\ commodity$).

Once you have decided what allocation of tokens you prefer, press the button **“Next”**. The computer will present you with the next decision problem or a screen asking you to wait for your final payments if you already went through 30 decision problems.

Your earnings:

Your earnings from the experiment are determined as follows. At the end of experiment, the computer will randomly select one of 30 decision problems. You will be paid according to choices you made. This means that each decision problem has the same 1 in 30 chance of being randomly selected. Note that each decision problem is independent from each other. Therefore, please pay attention to each one of them.

Once the computer randomly selects a decision problem, it will privately show you your choices (in the selected decision problem) and your payment (in terms of dollars of each commodity). As part of the check out process, you will have an opportunity to observe the experimenters imputing the amount of Mason Money and other commodities on the web. Note that e-gift cards purchased today can **only** be sent to your Masonlive e-mail account. You therefore need to provide a valid masonlive account to implement your payments.

Note that we will start conducting payments only after everyone in the room has finished the experiment. If you finish early, please remain silent and wait until everyone is done with the experiment.

If you have any questions, please raise your hand and and an experimenter will assist you in private.

Thank You and Good Luck!

INSTRUCTIONS

Thank you for participating in today's experiment. Please remain silent during the experiment. If you have any questions, please raise your hand and the experimenter will assist you in private.

This is an experiment in individual decision-making. Your earnings from the experiment will depend in part on your decisions and on chance. Your earnings will **not** depend on the decisions of other participants. Please pay careful attention to the instructions as a considerable amount of money is at stake.

Your payment in today's experiment will not be lower than $15 equivalent. This will be paid to you at the end of the experiment in private.

You will face 30 decision problems. In each decision problem, you will be given 100 tokens to be divided among 3 commodities. The three commodities are:

- **Cash**

  You can choose any (integer) dollar amount of this commodity. The amount of your choice will be paid to you at the end of experiment.

- **Mason Money**

  This is a prepaid debit program that provides a fast, safe and convenient way to make purchases on and off campus. Mason Money is accepted at all cafes and restaurants on campus and is linked to your Mason ID.

- **Barnes and Noble gift card**

  Barnes and Noble is the largest retail bookseller in the United States, and a leading retailer of content, digital media and educational products in the country.

You will be asked to allocate your 100 tokens to each of these commodities for 30 different sets of prices for each commodity. Prices will range from 2.5 tokens per dollar to 15 tokens per dollar. Prices are set up in the way that you can always buy an equivalent of $15. For instance, suppose that each commodity price is 5 tokens per dollar, then you can purchase $20 in cash, or $10 in Mason Money and $10 in Barnes and Noble gift card.

Due to the restrictions set by companies on gift cards **additional restrictions** apply:

- If you purchase any positive amount of Mason Money, you should purchase at least $5.

- If you purchase any positive amount of Barnes and Noble gift card, you should purchase at least $10

Sample Screenshot:

Decision (1 out of 30)

	Prices (tokens per $)	Dollars in Commodity	Tokens in Commodity
Cash	**11.0** tokens per dollar	0	0.0 Tokens
Mason Money	**6.5** tokens per dollar	0	0.0 Tokens
Barnes and Noble gift card	**7.0** tokens per dollar	0	0.0 Tokens
		Your total Expenditure is	0.0
			Next

Figure 1. Decision Screen

Figure 1 is an example of the decision screen. The first column contains the list of commodities. The second column contains the list of prices for each commodity (in tokens per dollar). The third column contains the quantity of commodity you will be asked to choose. This is the column you will use to make your decisions. The fourth column shows the allocation of your tokens among commodities ($Quantity\ of\ commodity * Price\ of\ commodity$).

Once you have decided what allocation of tokens you prefer, press the button **"Next"**. The computer will present you with the next decision problem or a screen asking you to wait for your final payments if you already went through 30 decision problems.

Your earnings:

Your earnings from the experiment are determined as follows. At the end of experiment, the computer will randomly select one of 30 decision problems. You will be paid according to choices you made. This means that each decision problem has the same 1 in 30 chance of being randomly selected. Note that each decision problem is independent from each other. Therefore, please pay attention to each one of them.

Once the computer randomly selects a decision problem, it will privately show you your choices (in the selected decision problem) and your payment (in terms of dollars of each commodity). As part of the check out process, you will have an opportunity to observe the experimenters imputing the amount of Mason Money and other commodities on the web. Note that e-gift cards purchased today can **only** be sent to your masonlive account. You therefore need to provide a valid masonlive account to implement your payments.

Note that we will start conducting payments only after everyone in the room has finished the experiment. If you finish early, please remain silent and wait until everyone is done with the experiment.

If you have any questions, please raise your hand and and an experimenter will assist you in private.

Thank You and Good Luck!