How General Are Measures of Choice Consistency? Evidence from Experimental and Scanner Data

Mingshi Chen Tracy Xiao Liu You Shan

Shu Wang Songfa Zhong Yanju Zhou*

March 6, 2025

Abstract

Choice consistency with utility maximization is a fundamental assumption in economic analysis and is extensively measured across various contexts. Here we investigate the generalizability of consistency measures derived from purchasing decisions using supermarket scanner data and budgetary decisions from lab-in-the-field experiments. We observe a lack of correlation between consistency scores from supermarket purchasing decisions and those from risky decisions in the experiment. However, we observe moderate correlations among experimental tasks and low to moderate correlations across purchasing categories and time periods within the supermarket. These results suggest that choice consistency may be characterized as a multidimensional skill set.

Keyword: consumer choice, revealed preference, choice consistency, utility maximization, lab-in-the-field experiment

JEL: C91, D81, D91

*Chen, Liu and Wang: School of Economics and Management, Tsinghua University (Chen: cms19@mails.tsinghua.edu.cn; Liu: liuxiao@sem.tsinghua.edu.cn; Wang: shu-wang20@mails.tsinghua.edu.cn). Shan: Faculty of Business for Science and Technology, School of Management, University of Science and Technology of China (shanyou@ustc.edu.cn). Zhong: Department of Economics, National University of Singapore (zhongsongfa@gmail.com). Zhou: School of Business, Central South University (zyj4258@sina.com). We thank David Ahn, Syngjoo Choi, Federico Echenique, Shachar Kariv, Bin Miao, John Quah, Satoru Takahashi, seminar participants at Singapore Management University, Lingnan University, Nankai University, and participants at Bounded Rationality in Choice Conference (BRIC-X, 2024), Decision: Theory, Experiments, and Applications (D-TEA, 2024) and Foundations of Utility and Risk (FUR, 2024). Liu gratefully acknowledges financial support by NSFC (72222005, 72342032).

1 Introduction

Central to economic analysis is the assumption that a decision maker (DM) maximizes her utility function given her budget constraint. Revealed preference analysis characterizes the conditions under which a DM, based on a given choice dataset, indeed maximizes a well-behaved utility function (Afriat, 1967; Samuelson, 1938; Varian, 1982). It thus provides a powerful framework and serves as a textbook example of normative approaches on how decisions should be made. Going beyond normative approach, researchers have examined the descriptive validity of this framework, and used it to assess choice consistency in various settings (Chambers and Echenique, 2016; Crawford and De Rock, 2014). A growing body of literature has measured consistency scores in risky, intertemporal, and social budgetary decisions in experiments, as well as in purchasing decisions derived from expenditure surveys and from scanner data in supermarkets.[1] Furthermore, it has been suggested that consistency scores reflect an individual's capacity to make sound decisions and may serve as a significant determinant of wealth disparities among individuals in the general population (Choi et al., 2014) and of development gaps between countries (Cappelen et al., 2023).

Building on these insights into the notion of choice consistency and its applicability across settings, here we take a closer look at its empirical validity and test its generalizability—the extent to which DMs exhibiting high consistency scores in one setting also demonstrate high scores in another setting. More specifically, do DMs with high consistency scores as consumers in a supermarket environment also have higher scores as participants in a controlled experimental setting? Two related questions arise: Do DMs show correlated consistency scores across risk, social, and food-related budgetary decisions in the experiment? Do DMs exhibit correlated consistency scores across different consumption categories and time periods in the supermarket? On the one hand, it is plausible that measures of choice consistency can be generalized across settings, as they may reflect a DM's overall capacity of making good decisions. Conversely, it is also conceivable that these measures are setting-specific, as DM may possess different preference structures, face varying budget constraints, and have distinct experiences or decision rules across settings, leading to divergent scores inferred from choice data.

We address these questions by combining experimental and scanner data. First, we examine consistency scores of the same individuals who make risky decisions in the lab-in-the-field exper-

[1]For studies utilizing experimental data: Ahn et al. (2014); Andreoni and Miller (2002); Choi et al. (2007); Dembo et al. (2024); Echenique et al. (2023); Ellis and Freeman (2024); Fisman et al. (2007); Halevy et al. (2018); Polisson et al. (2020), and for studies employing purchasing data: Blundell et al. (2003, 2008); Cherchye et al. (2020); Crawford (2010); Dean and Martin (2016); Echenique et al. (2011). Consistency with utility maximization is often referred to as choice consistency or rationality. For simplicity, we use choice consistency throughout this paper.

iment and make consumption decisions in the supermarket. Using the scanner data, we restrict our attention to 6,144 consumers of whom we have purchase records of meat and vegetables—the two most common consumption categories—over 24 consecutive months, so that we have sufficient power for the revealed preference analysis (see Section 3). We then successfully invite 1,055 of these 6,144 consumers to participate in a budgetary task (Experiment 1) in which participants make 22 portfolio decisions between two Arrow securities with different budget lines (Choi et al., 2007, 2014; Halevy et al., 2018; Kim et al., 2018). They need to allocate 100 experimental tokens between two accounts in which tokens are converted to cash with different exchange rates, and the amount in one of the two accounts will be paid out randomly with 50 percent probability. Therefore, by extracting these 1,055 consumers from the scanner data and aggregating the consumption at the month level, we obtain the choice dataset in the supermarket (Scanner Dataset 1).

We measure the choice consistency by evaluating the extent to which choices adhere to the Generalized Axiom of Revealed Preference (GARP), a necessary and sufficient condition whereby a dataset can be rationalized by a well-behaved utility function (in accordance with utility maximization). To assess how closely an individual's choice dataset complies with GARP, we compute the commonly used critical cost efficiency index (CCEI, Afriat, 1972), as well as the Houtman-Maks index (HMI, Houtman and Maks, 1985); money pump index (MPI, Echenique et al., 2011); and minimum cost index (MCI, Dean and Martin, 2016). Based on Experiment 1 and Scanner Dataset 1, our primary finding is that consistency scores are uncorrelated between risky decisions in the experiment and food consumption decisions in the supermarket regardless of the index used.

Second, we test whether the observed near-zero correlation is attributed to the differences in the types of decisions—risky decisions versus food consumption decisions—or due to the disparities between experimental and supermarket settings, and conduct a second experiment (Experiment 2). We invite another 302 consumers to participate in three budgetary tasks, including the risk task used in Experiment 1, social task (Andreoni and Miller, 2002; Fisman et al., 2007) and food task (Chen et al., 2023; Harbaugh et al., 2001). We find that consistency scores are moderately correlated across these three tasks, where the magnitude of correlations is the highest between risk and social tasks.

Third, we also examine the correlation of consistency scores for choices within supermarket using Scanner Dataset 2. The dataset consists of 822 consumers across different consumption categories including meat, vegetables, fruits and snacks, and 938 consumers across different time periods from 2018 to 2021. We find that the correlations of consistency scores are generally low for choices across consumption categories and over different time periods within supermarket.

We perform several robustness checks for these observed correlations. We use alternative indices of HMI, MPI, and MCI, control the power of detecting GARP violations, impose additional assumptions on preference structures using revealed preference analysis, and assess the fit of structural estimations of preferences. Our findings remain consistent across these variations with two exceptions. Namely, when we impose more assumptions on preference structures or evaluate the fit in structural estimations, the correlations within scanner data improve, both across categories and over time periods.

We investigate several factors that may influence consistency scores across different environments, including randomness in choice behaviors, budget constraints, formation of preferences, and the application of heuristic rules. Notably, our findings indicate that choices made in supermarkets tend to be more random compared to those made in experimental settings. Additionally, consistency in supermarket choices is influenced by various contextual factors, such as seasonality, time of day, promotional discounts, and individual shopping experiences. In contrast, consistency in experimental choices is shaped by learning effects across rounds of decisions, the use of heuristic rules, and participants' cognitive abilities. These distinct elements play crucial roles in how individuals navigate decision-making in the supermarket and in the experiment, and highlight the multifaceted nature of decision-making across settings.

Our study adds to the literature on measuring consistency with utility maximization based on revealed preference analysis. In addition to theoretical and experimental studies, the measures of choice consistency have been proposed as proxies for decision-making quality and widely used in applied settings, which link them to education, occupation, borrowing behavior of high-cost loan, income, and wealth (e.g., Banks et al., 2019; Cappelen et al., 2023; Carvalho et al., 2024, 2016; Choi et al., 2014; Kim et al., 2018; Li et al., 2023). In this literature, most studies focus on a specific type of decisions either in the experiment or in the field. One exception is Kim et al. (2018), in which they collect choice data on both risky and intertemporal decisions in the laboratory experiment to evaluate effects of an education program. Based on consistency scores of risky and intertemporal decisions in their dataset, we compute the Spearman's rank correlation to be 0.51, which is in line with the observed correlations across tasks in our Experiment 2. Our study is the first to examine the generalizability of consistency measures, and contributes to a better understanding of these measures.

Our study also contributes to the literature on the generalizability of choice behaviors across settings (Camerer, 2015; Chapman et al., 2024; Levitt and List, 2007; Snowberg and Yariv, 2021). A question that has attracted great attention is the generalizability of risk preferences. For example, Weber et al. (2002) show that risk attitudes vary across domains; these include financial, health/safety, recreational, ethical, and social decisions. Dohmen et al. (2011) find that different measures of risk attitudes are imperfectly correlated across settings. Barseghyan et al. (2011) use three deductible choices made in the domain of auto and homeowner insurance to separately estimate individual risk attitude using structural approach, and reject the null of fully domain-general risk attitudes. Einav et al. (2012) provide evidence that in support of positive correlations of risk preferences within five employer-provider insurance coverage decisions, and weaker relationships between these insurance decisions and investment decisions with respect to a 401(k) plan. In addition to risk preferences, moderate or weak correlations across settings have also been reported in time preferences (Augenblick et al., 2015; Burks et al., 2012); social preferences (Bruhin et al., 2019; Fehr and Leibbrandt, 2011; Galizzi and Navarro-Martinez, 2019); and strategic sophistication (Georganas et al., 2015; Levitt et al., 2011; Rubinstein, 2016). These studies have made important contributions to the extent to which theory-guided experimental findings can be generalized and applied to real-world scenarios, which is essential for theoretical development, experimental design, and practical application.

In this study, we examine choice consistency using risky, social and food decisions in two experiments and consumption purchases in the supermarket. On the one hand, the lack of correlation between experimental and scanner data suggests that consistency scores revealed from choice behaviors in various settings may not solely reflect the general ability to make good decisions according to one's preferences. On the other hand, we do observe moderate correlations across three tasks in the experiment, and some weak to moderate correlations across categories and time in the supermarket, which indicate that consistency measures can be generalized within settings.

Apart from measurement and identification issues, these findings point to the possibility that choice consistency is multidimensional. This perspective resonates with developments in the literature on intelligence, which have evolved over decades from a one-dimensional understanding of intelligence (Spearman, 1904) to the recognition of two dimensions—fluid and crystallized intelligence (Cattell, 1943), and ultimately to the widely accepted view of multidimensional intelligence (Horn and Cattell, 1966; McGrew, 2009). In the economics literature, accumulative evidence has provided support to the notion that choice behaviors are interconnected and can be reduced to several underlying common factors (Chapman et al., 2023; Dean and Ortoleva, 2019; Stango and Zinman, 2023), and the ability to make good decisions is multidimensional (Deming, 2021). Building on these studies, our study takes an initial step to address an important yet unexplored question on the generalizability of choice consistency, and contributes to a deeper understanding of the complexities inherent in decision-making processes.

The rest of the paper is organized as follows. Section 2 presents the theoretical background of revealed preference analysis. Section 3 describes the scanner data and experimental design. Section 4 presents our main results, and Section 5 examines the underlying mechanisms. Section 6 provides some concluding remarks.

2 Theoretical Background

Consider a DM who chooses bundles $x^t \in \mathbb{R}^K_+$ from budget lines $\{x : p^t \cdot x \leq p^t \cdot x^t, p^t \in \mathbb{R}^K_{++}\}$. A *dataset* $\mathcal{O} = \{(p^t, x^t)\}_{t=1}^T$ refers to a collection of T decisions of the DM. Let $\mathcal{X} = \{x^t\}_{t=1}^T$ be the set of bundles chosen by the DM. We say that x^i is *directly revealed preferred* to x^j, denoted by $x^i \succsim^* x^j$, if a DM chooses x^i when $x^j \in \mathcal{X}$ is affordable (i.e., $p^i \cdot x^j \leq p^i \cdot x^i$). Denote the asymmetric part as $\succ^*$, which refers to the relation of *directly strictly revealed preference.* Denote $\succsim^{**}$ the transitive closure of $\succsim^*$, which refers to the *revealed preferred* relation. A dataset $\mathcal{O}$ satisfies the Generalized Axiom of Revealed Preference (GARP) if the following holds:

$$\text{for all } x^i \text{ and } x^j,\ x^i \succsim^{**} x^j \text{ implies } x^j \not\succ^* x^i. \tag{1}$$

We say that a utility function $U : \mathbb{R}^K_+ \to \mathbb{R}$ *rationalizes* the dataset $\mathcal{O}$ if for every bundle x^t :

$$U(x^t) \geq U(x) \text{ for all } x \in \mathbb{R}^K_+ \text{ s.t. } p^t \cdot x \leq p^t \cdot x^t.$$

Afriat's theorem (Afriat, 1967; Varian, 1982) states that a dataset can be rationalized by some well-behaved (continuous and strictly increasing) utility function, if and only if the dataset obeys GARP.

2.1 Indices of Choice Consistency

Afriat's theorem informs whether a dataset can be rationalized by some utility function. When GARP is satisfied, we know that the dataset is consistent with utility maximization; However, when GARP is violated, it does not provide information about the extent of inconsistency. To address this, several indices have been developed to quantify the degree of consistency in the dataset. Below we briefly review four of them.

Critical Cost Efficiency Index (CCEI) A popular approach for measuring the departure from utility maximization is the *critical cost efficiency index* (CCEI) proposed by Afriat (1972). A DM has a CCEI $e \in [0, 1]$ if e is the largest number with a well-behaved U that rationalizes the dataset for every $x^t \in \mathcal{X}$:

$$U(x^t) \geq U(x) \text{ for all } x \in \mathbb{R}_+^K \text{ s.t. } p^t \cdot x \leq ep^t \cdot x^t. \tag{2}$$

A CCEI of 1 indicates passing GARP perfectly. A CCEI less than 1—say, 0.95—indicates that there is a utility function for which the chosen bundle x^t is preferred to any bundle that is more than 5 percent cheaper than x^t. Put differently, the CCEI can be viewed as the amount by which a budget constraint must be relaxed in order to remove all violations of GARP, because the DM is spending more money than is required to achieve the utility targets (Polisson and Quah, 2024). Alternative, it may be linked to the behavioral notion of the just-noticeable difference, DM's innate inability to distinguish similar bundles (Dziewulski, 2020).

Houtman-Maks Index (HMI) Houtman and Maks (1985) propose an alternative approach by measuring the maximal number of observations in the observed sample that are consistent with rational choices. For example, an HMI score of 0.941 indicates that the largest proportion of subset of choices consistent with GARP is 94.1 percent.

Money Pump Index (MPI) Echenique et al. (2011) provide an index to measure the amount of money one can extract from an individual for each violation of GARP. Their index is based on the idea that an individual with a GARP violation can be exploited by an "arbitrager" as a "money pump". The "arbitrager" can choose the allocation x^1 at price p^2 and allocation x^2 at price p^1, then trade x^1 with the individual at p^1 and x^2 at p^2, which yields a profit of $p^1(x^1 - x^2) + p^2(x^2 - x^1)$. Given multiple violations of GARP, a money pump cost will be associated with each violation. Following Echenique et al. (2011), we use the mean money pump costs for cyclic sequences of allocations with the length of two. For example, an MPI score equal to 0.059 means on average 5.9 percent of expenditure can be exploited by an "arbitrager" from GARP violation.

Minimum Cost Index (MCI) Dean and Martin (2016) propose MCI to measure the minimum cost of breaking all revealed preference cycles in a dataset. MCI is defined as $\min\limits_{B \subset R_0} \sum\limits_{(i,j)\in B} p^i(q^i - q^j) / \sum\limits_{t=1}^{T} p^t q^t$ such that R_0/B is acyclic, where R_0 represents the set of preference relation $\succsim^{**}$. The index has a high value when there are a large number of cycles for which all GARP violations are based on significant monetary differences relative to total expenditure. Thus, the MCI responds to both the number and severity of revealed preference violations. For example, if the MCI is 0.004, it means that the average cost of preference relations that must be removed to render the dataset consistent with GARP is 0.4 percent.

To summarize, these indices measure the degree of GARP violations from different perspectives. More detailed theoretical discussions can be found in Apesteguia and Ballester (2015), Echenique (2021) and Polisson and Quah (2024).

3 Scanner Data and Experimental Design

In this section, we describe the scanner data and experiments. We examine purchase behaviors of consumers who shop at a leading retail company with more than 400 supermarkets spreading over 6 provinces in southern China. First, we obtained an individual transaction scanner dataset from the company to measure consistency scores in the field. A sample transaction dataset is presented in Table 1, where each observation represents the purchase of a single item, that is, one type of good bought by one consumer in a single shopping instance. The supermarket membership ID uniquely identifies each consumer. The dataset also includes the unique store ID, transaction timestamp, category of the good, the quantity purchased and the amount of expenditure. In addition, we conducted two lab-in-the-field experiments to measure consistency scores for risky, social, and food decisions. The combination of scanner and experimental datasets enables us to test the generalizability of measures of choice consistency.

Table 1: Sample Records for Scanner Dataset

Membership ID	Store ID	Transaction timestamp	Category	Quantity (kg)	Expenditure (RMB)
a1	A1	2019-04-13 12:07:10	Meat	1.5	45.0
a1	A1	2019-09-21 16:35:45	Meat	2.0	40.0
a1	A1	2019-10-19 10:10:10	Vegetable	4.2	8.4
a1	A1	2019-11-10 19:05:05	Vegetable	1.8	3.6

3.1 Scanner Dataset

We focus on two categories of consumption, meat and vegetables, in the baseline analysis. The choice of these two categories is due to the following two reasons. First, because these two are most frequently purchased goods in the supermarkets, we are able to garner a sufficient number of observations. Second, because the unit for meat and vegetables is kilogram, it is convenient for us to have a refined measure of the quantity. We measure the quantity and price for meat and vegetables by averaging the quantity and prices of items in each category. We denote q^t_{in} as the quantity of item i in consumption category I for consumer n in month t, and Q^t_{In} as the quantity of consumption category I for consumer n in month t, which is obtained by the sum of q^t_{in}.

$$Q^t_{In} = \sum_{i \in I} q^t_{in}.$$

In a similar vein, we denote P^t_{In} as the price of the consumption category I for consumer n in month t, which is obtained by averaging prices weighted by quantity:

$$P^t_{In} = \frac{\sum_{i \in I} e^t_{in}}{Q^t_{In}},$$

where e^t_{in} is the expenditure for item i in category I. Note that we use shelf price instead of the final price with discount (Echenique et al., 2011). With quantity Q^t_{In} and price P^t_{In}, we can construct budget lines for consumer n in month t and perform revealed preference analysis for each consumer n.[2]

It is worth noting that it is often challenging to construct individualized price using consumption data. For example, Echenique et al. (2011) assume the same price for each item for all consumers and Dean and Martin (2016) assume the same price for all items under the same good category for all consumers. Since we are able to access high frequency data at the individual level, we can construct individualized price index for each consumer. This helps to reduce measurement errors and increase the power of GARP tests.

We restrict our attention to 6,144 highly frequent consumers who have purchase records of the two categories for 24 consecutive months from January 2019 to December 2020. After matching the scanner dataset with participants in Experiment 1 introduced in the next subsection, we have a dataset of 1055 consumers, which we refer to as Scanner Dataset 1.

Furthermore, to measure correlations of consistency scores across different categories in the supermarket, we include two additional frequently purchased categories: fruits and snacks, yielding 822 consumers who have purchase records of all four categories for 24 consecutive months in 2019 and 2020. To measure correlations of consistency scores over time periods, we focus on 938 consumers with purchase records of both meat and vegetables for 48 consecutive months from 2018 to 2021. We refer to this dataset as Scanner Dataset 2.

3.2 Experiments

In Experiment 1, we measure consistency scores for risky decisions using a standard budgetary design (Choi et al., 2007). In each round of the task, participants allocate 100 tokens between two contingent assets (accounts) and receive the money in one asset with 50 percent probability. The exchange rate between token and cash differs for the two assets and varies across decisions. If

[2]To perform the revealed preference analysis, we assume that preferences over meat and vegetables are weakly separable from other goods and services (Dean and Martin, 2016; Echenique et al., 2011).

participants are risk neutral or risk seeking, they would allocate all tokens to the cheaper asset; if they are risk averse, they would allocate some tokens to each asset depending on their risk preferences.

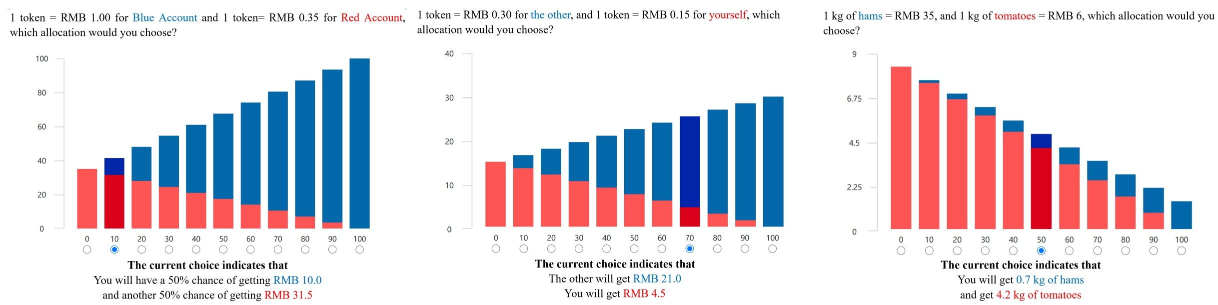


Figure 1: Screenshots for the Risk, Social and Food Tasks

Moreover, to help consumers in the supermarket understand the experimental tasks, we discretize the choice set as in Kim et al. (2018), by which they choose among 11 allocation options.[3] Figure 1 presents a screenshot of the task. In total, participants face 22 decision tasks presented in random order. Different from Choi et al. (2007, 2014), we do not use randomly generated budget lines for each participant. Instead, every individual is presented with the same set of budget lines to ensure that the power for revealed preference tests is the same for all participants and to facilitate comparison across participants (Halevy et al., 2018). At the end of the experiment, we include a 10-question version of Big Five personality questions, seven questions from Raven's progressive matrices, and collect demographic information in the post-experiment questionnaire.

To further investigate correlations of consistency scores across different tasks in the experiment, we conduct Experiment 2. In addition to risky decisions as in Experiment 1, we also measure consistency scores for social and food decisions. We use a modified dictator game to measure consistency scores for social decisions (Andreoni and Miller, 2002; Fisman et al., 2007). Participants allocate 100 tokens between themselves and another randomly matched participant across 22 rounds. In every round, each token has a different cash value for the individual and her matched participant.

To measure consistency scores for food decisions in the experiment, we construct a shopping environment (Chen et al., 2023; Harbaugh et al., 2001). In our design, participants allocate the expenditure of RMB 50 (USD 1 $\approx$ RMB 6.5) between a specific type of meat (ham) and a specific type of vegetables (tomato) in 22 rounds. The price level for each product is chosen from the range of actual prices from 2020 to 2021. We present the risk, social and food tasks in random

[3] Polisson and Quah (2013) show that Afriat's theorem is applicable in discrete settings.

order to participants and end the experiment with the post-experiment questionnaire identical to that in Experiment 1. Screenshots for the two experiments are presented in Figure 1, and detailed instructions are provided in Appendix C.

The two experiments were conducted from July to September 2021 and in December 2021, respectively. In Experiment 1, of the 6,144 frequent consumers identified in our baseline consumption analysis, we have contact information for around half of them. We invited these consumers to participate in the incentivized experiment in nearby supermarkets. In total, 1,055 highly frequent consumers participated in the experiment across 96 supermarkets. In Experiment 2, another 302 consumers participated in the experiment.[4] All participants in both experiments received RMB 50 for completing the experiment. Additionally, we randomly paid 10 percent of participants based on their choices. Note that if the decision in food task was chosen for reward, we deposited redemption vouchers into the consumer's membership account based on her choice. Each individual received an average payment of RMB 57.3 in Experiment 1 and RMB 55.1 in Experiment 2. Table A1 reports summary statistics for our participants in both experiments. For example, among the 1,055 participants in Experiment 1, the average monthly expenditure on meat and vegetables is RMB 482.2, with an average purchase frequency of 11.8 days per month.

4 Results

4.1 Experiment 1 vs. Scanner Dataset 1

To examine the correlation of choice consistency between experimental and scanner data, we utilize choices made by 1,055 individuals both in Experiment 1 and Scanner Dataset 1. Figure 2 presents the cumulative distribution of consistency indices for risky decisions (solid line) and the summary statistics are in Table A2 (Panel A). Taking CCEI as the example, we find that 28.2% participants have no violations of GARP, and the average CCEI score is 0.941, which implies that the budget must be reduced by 5.9 percent to remove all GARP violations.[5] To examine the power of the test, we calculate simulated scores by generating observations using uniformly random allocations over 11 options across 22 rounds for 1,000 times (Bronars, 1987). In contrast

[4]To avoid learning spillover between experiments, we only invite consumers who did not participate in Experiment 1. Therefore, most of them are not highly frequent consumers and we do not have enough power to examine the consistency scores of their choices in the supermarket.

[5]The observed CCEI distribution is comparable to those in other experimental studies on risky decisions. For example, the average CCEI scores are 0.881 in Choi et al. (2014), 0.937 in Choi et al. (2007), and 0.979 in Halevy et al. (2018).

to the CCEI scores based on choice data, data from simulated random allocations (dashed line in Figure 2) have substantially lower scores ($p < 0.01$, two-sided two-sample t-tests).[6] For example, only 1.5 percent of their CCEI scores are over 0.9. This suggests that our choice data have sufficient power to detect GARP violations.

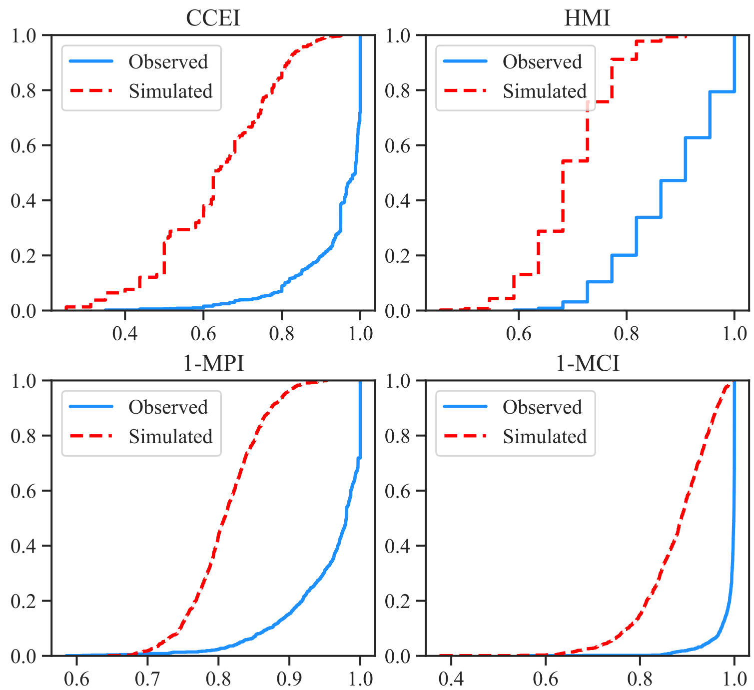


Figure 2: Distribution of Consistency Indices for Risky Decisions in Experiment 1

Figure 3 presents the cumulative distribution of observed (solid line) and simulated (dashed line) scores of four consistency indices for consumption decisions in Scanner Dataset 1 (see Panel B in Table A2 for summary statistics). We find that 15.5% consumers have no violations of GARP, and the average CCEI score is 0.946.[7] As supermarket consumers face different budget lines, we calculate simulated CCEI score for each consumer. Each simulated CCEI score is the average score generated by repeating the uniformly random allocations over budget shares along one consumer's 24 budget lines for 100 times (Dean and Martin, 2016). The results show that simulated CCEI scores are significantly smaller than CCEI scores for actual choices ($p < 0.01$, two-sample t-tests), and only 6.2 percent of them are over 0.9.[8] This confirms the sufficient power of budget lines in

[6]Without specific clarification, we use two-sided tests throughout the study.

[7]Using scanner data, the average CCEI scores are 0.976 in Echenique et al. (2011), and 0.99 in Dean and Martin (2016).

[8]Using two-sample t-tests, all observed scores are significantly higher than simulated scores ($p < 0.01$) in both

the scanner dataset.

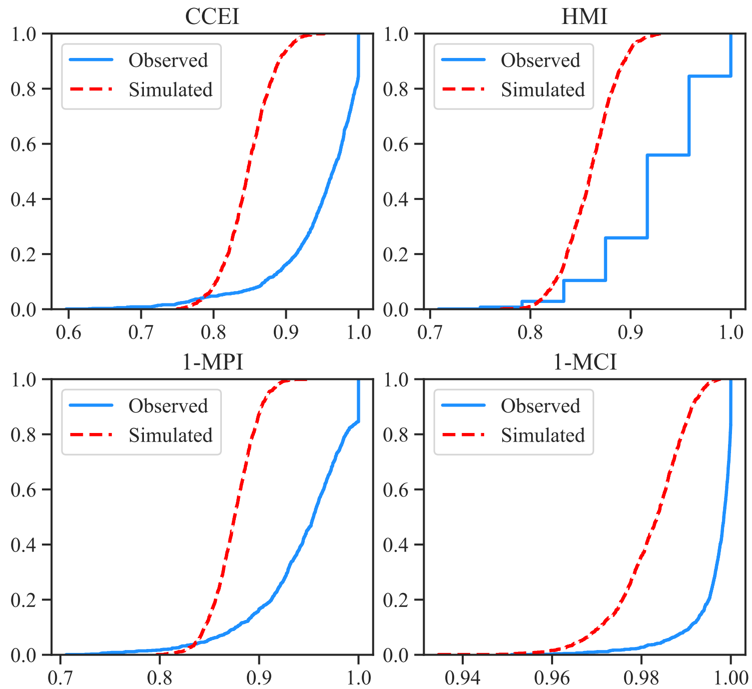


Figure 3: Distribution of Consistency Indices for Consumption Decisions in Scanner Dataset 1

We now turn to the main research question regarding the generalizability of consistency measures between risky decisions in the experiment and consumption decisions in the supermarket. We check their association using Spearman's rank correlation (r), which is the Pearson correlation between the rank values of those two variables. Moreover, for the correlation coefficients reported in the study, we refer to $[0, 0.1)$ as very low or no correlation, and, following the convention proposed by Cohen (1988), we refer to $[0.1, 0.3)$ as low, $[0.3, 0.5)$ as moderate, and $[0.5, 1]$ as high.

Figure 4 presents scatter plots using four indices, CCEI, HMI, MPI and MCI, and shows that the rank correlation coefficients are very low and statistically insignificant ($p > 0.1$ for all indices). That is, the consumer with a high level of consistency score for risky decisions in the experiment does not necessarily exhibit a high level of consistency for consumption decisions in the supermarket. We summarize this finding as follows.

Result 1. *The consistency scores are uncorrelated between risky decisions in Experiment 1 and consumption decisions in Scanner Dataset 1.*

Experiment 2 and Scanner Dataset 2. To avoid redundancy, we omit these repeated expressions in the following analyses.

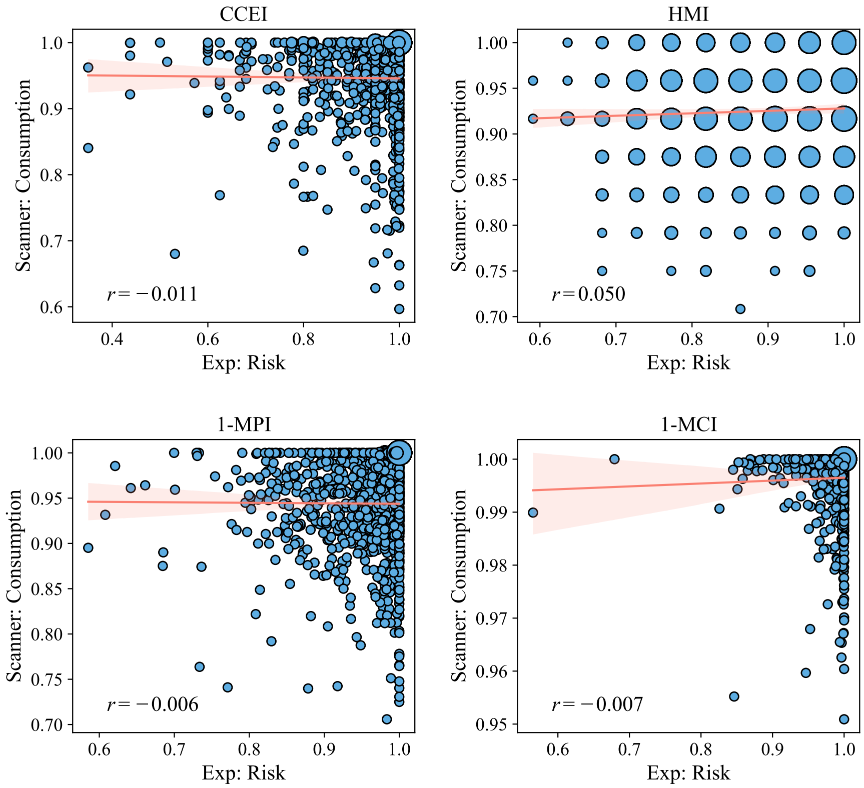


Figure 4: Correlation of Consistency Indices Between Experiment 1 and Scanner Dataset 1

Notes: $N = 1,055$. Spearman's rank correlations (r) are reported. The size of each circle represents the number of observations at the point. The lines represent the linear fit, and 95% confidence intervals are shown in the shaded area. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Given the above similar correlations among all consistency indices, we examine the relationships among these indices within each setting. Table 2 reports correlation coefficients for pairwise comparisons and shows that the average of correlation coefficients is 0.809. Whereas there have been theoretical discussions on how best to capture the extent of GARP violations (Apesteguia and Ballester, 2015; Echenique, 2021; Polisson and Quah, 2024), we show empirically that these indices are highly correlated. In subsequent analyses, for ease of presentation, we focus on results using CCEI and report results for other indices as robustness checks.

Table 2: Correlation of Consistency Indices in Experiment 1 and Scanner Dataset 1

	Exp: Risk				Scanner: Consumption		
	CCEI	HMI	1-MPI		CCEI	HMI	1-MPI
HMI	0.756***			HMI	0.698***		
1-MPI	0.944***	0.644***		1-MPI	0.890***	0.582***	
1-MCI	0.988***	0.784***	0.936***	1-MCI	0.905***	0.790***	0.791***

Notes: $N = 1,055$. Spearman's rank correlations are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.2 Experiment 2 & Scanner Dataset 2

In this section, we examine the correlations of choice consistency when making budgetary choices in Experiment 2 and consumption choices in Scanner Dataset 2. This analysis helps us see if the lack of correlation in consistency scores from risky decisions in Experiment 1 and consumption decisions in Scanner Dataset 1 is due to the different types of choices being made or different settings.

First, we calculate CCEI scores for each task in Experiment 2 (Figure A1 and Table A3) and their patterns are similar to those in Experiment 1. Next, we test pairwise correlations of CCEI scores across tasks (Figure 5). The results reveal that CCEI scores for risky decisions are moderately correlated with those both for social decisions (0.412, $p < 0.01$) and food decisions (0.305, $p < 0.01$) in Experiment 2, suggesting that consumers with higher consistency scores in the risk task tend to exhibit a higher level of consistency in the social and food tasks in the experiment. Meanwhile, the CCEI scores for social and food decisions are weakly correlated (0.254, $p < 0.01$). These lead to our next observation.

Result 2. *The correlations of consistency scores are moderate between risky and social (food) decisions, and low between social and food decisions in the experiment.*

We similarly summarize the consistency of choices in Scanner Dataset 2 (Figure A2 and Table A4), and examine correlations of consistency across non-overlapping pairs of consumption categories and time periods.[9] Figure 6 presents Spearman's rank correlations of CCEI scores. These results indicate that the correlations are positive across most food categories, though the magnitude is low (0.173, $p < 0.01$; -0.046. $p > 0.1$; and 0.096, $p < 0.01$). Similarly, there are also

[9]We do not look at the correlations of consistency scores between pairs that share the same categories (time periods) because the same consumption choices in those overlapping categories (time periods) can create correlations.

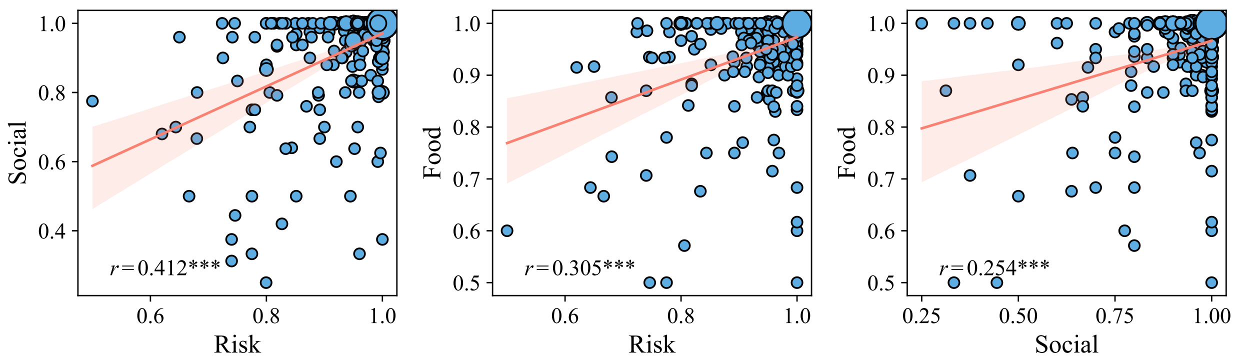


Figure 5: Correlations of CCEI Scores in Experiment 2

Notes: $N = 302$. Spearman's rank correlations (r) are reported. The size of each circle represents the number of observations at the point. The lines represent the linear fit, and 95% confidence intervals are shown in the shaded area. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

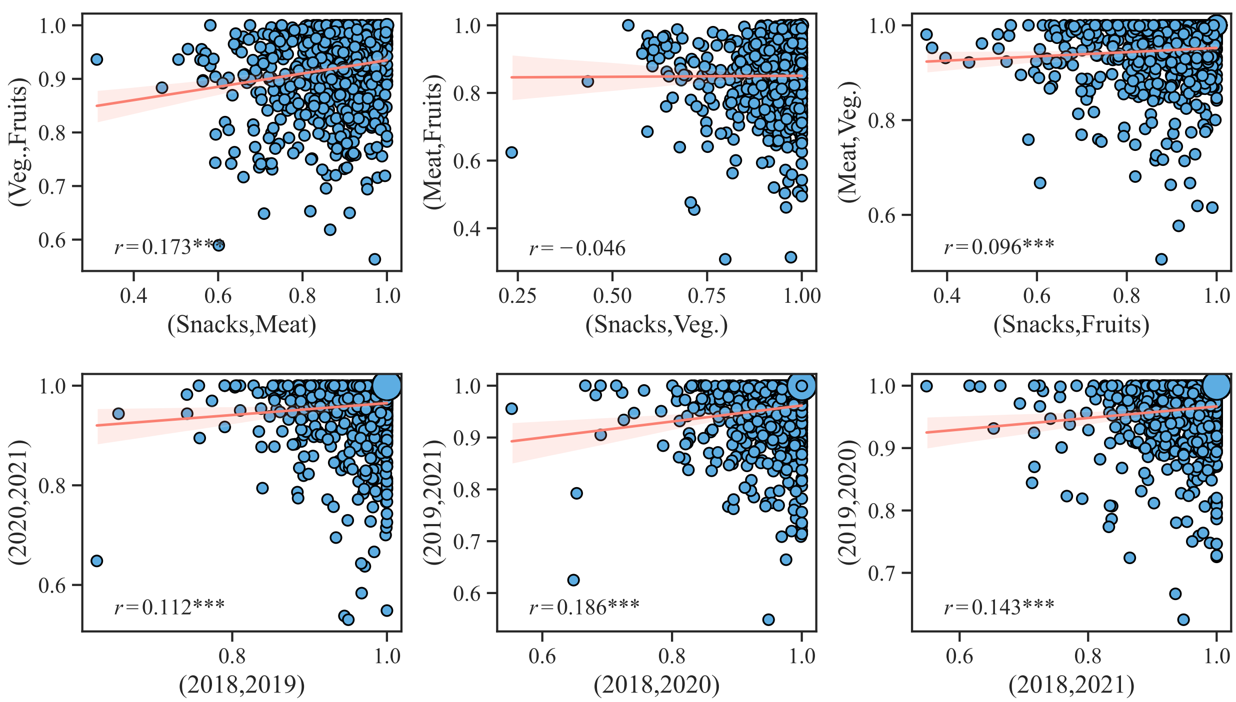


Figure 6: Correlations of CCEI Scores in Scanner Dataset 2

Notes: $N = 822$ (938) for the three correlations across categories (time periods). Spearman's rank correlations (r) are reported. The size of each circle represents the number of observations at the point. The lines represent the linear fit, and 95% confidence intervals are shown in the shaded area. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

significantly positive but low correlations across different time windows (0.112, $p < 0.01$; 0.186, $p < 0.01$; and 0.143, $p < 0.01$). We summarize these findings below.

Result 3. *The correlations of consistency scores are generally low both across consumption categories and over time periods in the supermarket.*

4.3 Robustness Checks

In this subsection, we conduct several robustness checks on our main results including alternative indices of HMI, MPI, and MCI, the power of detecting GARP violations, imposing more assumptions on preferences, and fitness in structural estimations.

Alternative Indices In addition to CCEI, alternative indices including HMI, MPI and MCI are proposed to measure the degree of GARP violations. We use these indices to examine the correlations.

Heterogeneous Power Consumers face individualized budget sets in our scanner dataset, so the power for detecting GARP varies across consumers. To address this issue, we calculate Selton Score by subtracting simulated CCEI from the observed CCEI (Dean and Martin, 2016; Selten, 1991). In addition, we use the residual from a regression of simulated CCEI on observed CCEI.

More (Realistic) Assumptions on Preferences Since GARP measures the extent of maximizing some well-behaved, i.e., continuous and strictly increasing, utility function, it is possible that our choice dataset satisfies GARP, but may fail to meet requirements on particular preferences. Thus, we consider more realistic restrictions and compute new consistency indices under these conditions. Specifically, two properties related to income are often discussed. First, we examine the adherence to the normality of demand for both goods, which suggests that the demand for commodities increases as income rises, holding their prices constant (denoted as *Normal*) (Cherchye et al., 2018).[10] Second, we consider the homothetic preference, assuming a linear relationship between income level and consumption (*Homothetic*) (Heufer and Hjertstrand, 2019; Varian, 1983). Moreover, consumers' utility may be negatively affected by the expenditure, especially in the supermarket with different expenditures across periods. Therefore, we consider the quasilinear utility $U(x) - px$, where the expenditure of goods reduces the utility in a linear form (*Quasilinear*) (Brown and Calsamiglia, 2007; Demetry et al., 2022). Meanwhile, to further evaluate choices on two contingent assets in the experiment, we also consider the compliance with first order stochas-

[10] Cherchye et al. (2018) provide the necessary and sufficient condition for the normality of both goods. To further assess how well our data align with this condition, we scale the budget in their NARP-I condition (Definition 3) by a factor e, and identify the largest e that allows the dataset to meet the condition for both goods. We then regard this maximum e as *Normal*.

tic dominance (FOSD) and expected utility (EU) maximization (Nishimura et al., 2017; Polisson et al., 2020).

Preference Estimation In addition to revealed preference analysis, we further measure the extent to which certain specific types of preferences fit the choice data. For risky decisions, we apply the disappointment aversion (DA) model, which is equivalent to rank dependent utility with binary states. For the other types of decisions, we use the constant elasticity of substitution (CES) utility functions (Fisman et al., 2007), in which parameter α captures the relative weight placed on the first good, and ρ is the curvature of indifferent curves. We report the detailed estimation procedure in Appendix B. Because the log likelihood (*LL*) measures the goodness of fit of the utility model, we may interpret it as the degree of noise when making decisions.

Table 3: Robustness of Correlations of Consistency Indices

	Exp. 1 vs. Scanner Dataset 1	Exp. 2			Scanner Dataset 2					
					Across cat.			Across time		
Panel A: Consistency indices										
CCEI	○	◉	◉	◍	◍	○	○	◍	◍	◍
HMI	○	◉	◉	◉	○	○	○	○	◍	◍
MPI	○	◉	◉	◍	◍	○	○	◍	◍	◍
MCI	○	◉	◉	◍	◍	○	◍	◍	◍	◍
Panel B: Power of the test										
Selten Score	○	–	–	–	○	○	○	○	◍	◍
Power-adjusted CCEI	○	–	–	–	◍	○	○	○	◍	◍
Panel C: More (realistic) assumptions on preferences										
Normal	○	◉	◉	◉	◍	◍	◍	◉	●	◉
Homothetic	○	◉	◍	◍	◍	○	◍	◍	◉	◉
Quasilinear	○	–	–	–	◍	◉	◍	◍	◉	◉
FOSD	○	◍	◍	–	–	–	–	–	–	–
EU	○	◍	◍	–	–	–	–	–	–	–
Panel D: Preference estimation										
LL	○	●	◉	◉	◍	○	◍	◉	●	◉

Notes: ○/◍/◉/● represents that the magnitude of correlation coefficient is in [0,0.1)/[0.1,0.3)/[0.3,0.5)/[0.5,1]. Specific values are detailed from Tables A5 to A7. – denotes that the index is not applicable. Within Experiment 2, the three coefficients represent the correlations of consistency scores between risk and social tasks, risk and food tasks, and social and food tasks, respectively. In Scanner Dataset 2, the three coefficients across categories represent those between (Snacks,Meat) and (Veg.,Fruits), (Snacks,Veg.) and (Meat,Fruits), and (Snacks,Fruits) and (Meat,Veg.); those across time are correlations between (2018,2019) and (2020,2021), (2018,2020) and (2019,2021), and (2018,2021) and (2019,2020).

Table 3 summarizes the correlations using these consistency indices (see Table A5 to A7 for details). We find support for the robustness of Result 1 on no correlation between experimental and scanner data and Result 2 on moderate correlations within experimental data. Notably, after imposing more assumptions on consumption decisions in the supermarket, correlations within scanner data, as described in Result 3, increase substantially both in magnitude and significance.

5 Underlying Mechanisms

Our findings indicate that while consistency scores can be derived from choice data across various contexts, they do not always reflect a general ability to make decisions aligned with individual preferences. This resonates with existing literature highlighting the challenges in inferring true preferences from observable choices. For instance, Ariely et al. (2003) introduce the notion of coherent arbitrariness, suggesting that subjects are swayed by arbitrary anchors but respond consistently to significant changes in price, quantity, and quality. Similarly, Dean and Sautmann (2021) argue that intertemporal choices in experiments may not accurately represent true time preferences, as they can be affected by financial shocks and budget constraints. Furthermore, the frequently observed S-shaped probability-weighting function in prospect theory may arise from cognitive uncertainty, where individuals tend to default to cognitive shortcuts (Enke and Graeber, 2023), or from the complexities involved in evaluating lotteries (Oprea, 2024). Observable choices such as those in the Allais paradox and Ellsberg paradox may also reflect noise or errors (McGranaghan et al., 2024; Nielsen and Rehbeck, 2022) or cognitive challenges such as the failure of contingent thinking (Esponda and Vespa, 2024; Niederle and Vespa, 2023). Additionally, de Clippel et al. (2024) show that subjects, whose choices are consistent with utility maximization, often rely on heuristic decision rules that mimic the characteristics of utility functions rather than truly maximize utility. In a similar vein, consistency scores inferred from observable choices may be "as if" to some degree, because it is challenging to separate the influences of unobservable factors such as randomness, budget constraints, preferences, and heuristics which we elaborate below.

5.1 Randomness in Choice Behavior

We examine the randomness in the choice behaviors, which can be viewed as a measure of some unobservable factors. We apply a non-parametric one-sided test proposed by Cherchye et al. (2023) to examine whether a choice dataset is best explained by a random DM or an approximate utility maximizer. This test uses the permutation approach, which fixes the budget sets but permutes the relative consumption shares over different rounds for each DM, to operationalize the random behaviors. They show that, given a sufficient number of observed choices, the CCEI from an approximate utility maximizer has a probability close to 1 of exceeding CCEI generated by the permutation test. Accordingly, we compute the proportion of permuted CCEI scores, generated through 10,000 permutations, that surpass the observed CCEI.[11] If such proportion is smaller than the traditional significance level $\alpha = 0.05$, we can regard the DM as an approximate utility maximizer.

Table 4: Proportions of Approximate Utility Maximizers

	N	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Panel A: Risk task in Experiment 1 vs. Consumption in Scanner Dataset 1				
Risk	1,055	67.87%	77.06%	82.37%
(Meat,Veg.)	1,055	5.21%	22.65%	37.82%
Panel B: Experiment 2				
Risk	302	70.86%	86.42%	89.74%
Social	302	64.57%	79.14%	83.11%
Food	302	22.52%	44.37%	55.96%
Panel C: Scanner Dataset 2				
(Snacks,Meat)	822	4.01%	15.94%	27.25%
(Snacks,Veg.)	822	4.99%	20.44%	30.05%
(Snacks,Fruits)	822	3.65%	11.80%	21.90%
(Meat,Veg.)	822	5.47%	20.92%	35.16%
(Meat,Fruits)	822	2.80%	10.46%	18.49%
(Veg.,Fruits)	822	2.31%	9.25%	16.91%
(2018,2019)	938	8.53%	29.64%	43.92%
(2018,2020)	938	5.97%	24.84%	37.31%
(2018,2021)	938	4.58%	18.23%	31.77%
(2019,2020)	938	5.54%	22.60%	36.57%
(2019,2021)	938	7.04%	24.41%	39.13%
(2020,2021)	938	4.90%	22.49%	35.61%

We find that 77.06% participants in Experiment 1 can be classified as approximate utility maximizers, while this number is 22.65% when their scanner data are used. Furthermore, an individual classified as an approximate utility maximizer in the experiment does not have a significantly higher chance to be labeled as an approximate utility maximizer in the supermarket (22.5% vs.

[11] We follow the procedure in Cherchye et al. (2023) to abort the test when the proportion exceeds 0.2 for the first 1,000 permutations and report this value to speed up computation.

23.1%, $p > 0.1$, two-sample proportion test). We also examine the proportion in Experiment 2 and Scanner Dataset 2 and use alternative significance levels. Table 4 summarizes these proportions with the corresponding significance level, and shows that the results are robust. Our observations are comparable to those of Cherchye et al. (2023), where the proportions are 82% for social task in a two-goods setting in the experiment and 30% for real-life choices at the significance level of 0.05. Overall, this analysis shows that choices in scanner data are more random than those in the experiment, and the sources of randomness are distinct in these two settings.

5.2 Budget Constraints

One important difference across the two environments is the budget constraint. In the lab, participants are presented with well-defined budget lines using a simple interface. By contrast, the budget lines of consumers are not directly observable and are constructed by aggregating the choices in a given set of categories. In addition, consumption data often lack the power to detect GARP violations, because budget lines may not sufficiently cross (Blundell et al., 2003, 2008; Chambers and Echenique, 2016, Chapter 5). However, our scanner data do have the power to detect GARP violations. To take into account of the perceived budget constraints, we utilize two alternative approaches. The first approach is the Price Misperception Index (PMI) in de Clippel and Rozen (2023), which quantifies the minimal degree of price misperception that can rationalize choice data. The second approach is proposed by Deb et al. (2023) to examine the consistency with the generalized axiom of price preference (GAPP). This approach allows us to calculate deviations from consistent welfare predictions across different budgets, after accounting for one specific type of inattention to prices.[12] Panel A in Table 6 shows that the observed correlations are similar to those using CCEI.[13]

[12] Given $p^1 = (1, 2), x^1 = (0, 2); p^2 = (2, 1), x^2 = (3, 1)$, we can find the dataset satisfies GARP. However, it violates GAPP, because the participant can spend less to purchase bundle x^1 at period 2, thus being better off at price p^2 than at p^1. Similarly, she can also spend less to purchase bundle x^2 at period 1 and become better off at p^1 than at p^2. This leads to a contradiction in the preference over prices. GAPP is a necessary and sufficient condition for the dataset to be rationalized with $U(x, -px) : \mathbb{R}^K_+ \times \mathbb{R}_- \to \mathbb{R}$ (Theorem 1, Deb et al., 2023). According to Proposition 3 in Deb et al. (2023), since the expenditure in our experiment is fixed, their measure (ϑ^*) is equivalent to CCEI.

[13] Additionally, we also conduct a few more robustness checks. First, we restrict our analyses to the most frequently purchased subcategory, i.e. pork and leaf vegetables, within each category and calculate the consistency scores accordingly. Second, we assume the price in each category is the same for all consumers in each month (Dean and Martin, 2016; Echenique et al., 2011). Specifically, we obtain the aggregate price index P^t_I for category I in month t: $P^t_I = \sum_{i \in I} e^t_i / Q^t_I$, where e^t_i represents the total expenditure of all consumers on item i in category I in month t, and Q^t_I is the total quantity of items purchased in category I in month t. In these analyses, all coefficients are close to our baseline findings.

Table 5: Correlations of Alternative Indices for Budgets Constraints and Heuristic Rules

	Exp. 1 vs. Scanner Dataset 1	Exp. 2			Scanner Dataset 2					
					Across cat.			Across time		
PMI	○	◑	◔	◔	◔	○	○	○	◔	○
GAPP	○	–	–	–	○	○	○	◔	◔	◔

Notes: ○/◔/◑/● represents that the magnitude of correlation coefficient is in [0,0.1)/[0.1,0.3)/[0.3,0.5)/[0.5,1]. Specific values are detailed in Table A8. – denotes that the index is not applicable. Within Experiment 2, the three coefficients represent the correlations of consistency scores between risk and social tasks, risk and food tasks, and social and food tasks, respectively. In Scanner Dataset 2, the three coefficients across categories represent those between (Snacks,Meat) and (Veg.,Fruits), (Snacks,Veg.) and (Meat,Fruits), and (Snacks,Fruits) and (Meat,Veg.); those across time are correlations between (2018,2019) and (2020,2021), (2018,2020) and (2019,2021), and (2018,2021) and (2019,2020).

Table 6: Correlations of Alternative Indices for Budgets Constraints and Heuristic Rules

	Exp. 1 vs. Scanner Dataset 1	Exp. 2			Scanner Dataset 2					
					Across cat.			Across time		
Downward	○	●	◑	◑	○	○	○	○	◔	◔

Notes: ○/◔/◑/● represents that the magnitude of correlation coefficient is in [0,0.1)/[0.1,0.3)/[0.3,0.5)/[0.5,1]. Specific values are detailed in Table A8. – denotes that the index is not applicable. Within Experiment 2, the three coefficients represent the correlations of consistency scores between risk and social tasks, risk and food tasks, and social and food tasks, respectively. In Scanner Dataset 2, the three coefficients across categories represent those between (Snacks,Meat) and (Veg.,Fruits), (Snacks,Veg.) and (Meat,Fruits), and (Snacks,Fruits) and (Meat,Veg.); those across time are correlations between (2018,2019) and (2020,2021), (2018,2020) and (2019,2021), and (2018,2021) and (2019,2020).

Constructing budget lines is inherently challenging due to various factors influencing both consumer behaviors and firm pricing strategies. On the demand side, consumers often respond to price changes in a coarse manner, leading to suboptimal purchasing decisions. For instance, they may exhibit left-digit bias, where they overemphasize the significance of changes in the leftmost digit of a price (List et al., 2023), and tend to evaluate price changes based on their memory of past prices, which can distort their perceptions of current value (Bordalo et al., 2020). On the supply side, firms frequently engage in strategic pricing and adjust their prices in ways that can manipulate consumer perceptions and purchasing behaviors. As a result, budget constraints are not exogenously given, but influenced by market dynamics. Addressing these complexities requires further research to accurately measure and understand these factors.

5.3 Formation of Preferences

Formation of preferences also differs across environments. Plott (1996) proposes the notion of preference discovery, whereby choice consistency with utility maximization can be understood as a discovery process in which subjects learn their own needs through a process of reflection and practice. This perspective highlights the dynamic nature of formation of preferences, where individuals refine their understanding of their needs and objectives over time. Complementing this view, Deming (2021) distinguishes between the decision-making ability of performing a routine task and a problem-solving task, and suggests that the formation of preferences may differ depending on the task characteristics. In this regard, DM as a consumer is more likely to do a routine task and may be subjective to some preference shocks related to various contextual factors. By contrast, when participants in the lab are presented with risky decisions and social decisions, they need to learn the abstract experimental rules, since these rules are not part of their daily experience. Therefore, choice consistency in the lab may be related to the ability to "discover" one's preference in the new choice environment and "solve" the choice problem accordingly. Below we present some results testing the effect of some contextual factors.

Preference Shocks in Scanner Data We examine whether consumers' preferences on meat and vegetables change across time (years, seasons, days and hours) and other contexts (discounted vs. non-discounted purchases) in Scanner Dataset 1.

To examine the impact of each factor, we first generate the choice datasets, s_1 and s_2, for two scenarios separately.[14] Then we calculate the consistency scores for each dataset, CCEI_{s_1} and CCEI_{s_2}, and the combined dataset, $\text{CCEI}_{(s_1,s_2)}$. Thus, we can obtain $\text{CCEI}_{\text{diff}(s_1,s_2)} = \min\{\text{CCEI}_{s_1}, \text{CCEI}_{s_2}\} - \text{CCEI}_{(s_1,s_2)}$ at the consumer level. By definition, $\text{CCEI}_{\text{diff}(s_1,s_2)} \geq 0$. When $\text{CCEI}_{\text{diff}(s_1,s_2)} > 0$, there does not exist a well-behaved utility function that can rationalize the two datasets after accounting for the corresponding consistency scores, thus implying distinct preferences between the two scenarios. However, $\text{CCEI}_{\text{diff}(s_1,s_2)} > 0$ may be simply due to power of the test, namely, the combined dataset has more observations, so it is more likely to fail the test. To account for this, we build a benchmark by randomly allocating the combined dataset into two simulated choice datasets, s_1' and s_2', keeping the same number of choices between s_1 (s_2) and s_1' (s_2'). Therefore, we obtain $\text{CCEI}_{\text{diff}(s_1',s_2')}$ for each consumer. Repeating this procedure

[14]For years, s_1 (s_2) encompasses 12 choices in 2019 (2020) for each consumer n. For seasons, both s_1 and s_2 represent one season with six choices, yielding six different combinations of s_1 and s_2 across four seasons. For days, we focus on working days and non-working days. In China, the year 2019 (2020) comprises 250 (251) working days after accounting for weekends and holidays. Using purchasing scanner data for working and non-working days separately, we focus on months with positive consumption to reconstruct the monthly quantity Q^t_{In} and price P^t_{In} in each scenario. For hours, we similarly generate Q^t_{In} and P^t_{In} using purchases on meal time and non-meal time, respectively. Meal time spans from 10:00 to 14:00 and from 16:00 to 19:00, and other time is considered non-meal time. Lastly, to examine the impact of promotional discounts, we obtain s_1 for discounted purchases and s_2 for non-discounted purchases in a similar manner.

100 times, we can examine whether the observed $\text{CCEI}_{\text{diff}(s1,s2)}$ equals to the average simulated $\text{CCEI}_{\text{diff}(s_1',s_2')}$, denoted as $\overline{\text{CCEI}}_{\text{diff}(s_1',s_2')}$.

Table 7 reports the results. By conducting pairwise comparisons among the four seasons, we observe that consumers' food preferences in spring are notably different from other seasons, especially fall and winter.[15] Additionally, our analysis reveals distinct preferences across multiple dimensions: between meal and non-meal time, and between discounted and non-discounted purchases. Meanwhile, we cannot detect significant differences between the year 2020 and 2021 and between working and non-working days. Overall, these results suggest that there are contextual factors influencing choices in the supermarket setting.

Table 7: The Impact of Setting-specific Factors on CCEI in Scanner Dataset 1

Category	s_1	CCEI_{s_1}	s_2	CCEI_{s_2}	$\text{CCEI}_{\text{diff}(s_1,s_2)}$	$\overline{\text{CCEI}}_{\text{diff}(s_1',s_2')}$	p-value
Season	Spring	0.995	Summer	0.996	0.009	0.008	0.207
Season	Spring	0.995	Fall	0.996	0.012	0.009	0.005
Season	Spring	0.995	Winter	0.995	0.012	0.010	0.004
Season	Summer	0.996	Fall	0.996	0.008	0.007	0.128
Season	Summer	0.996	Winter	0.995	0.010	0.009	0.102
Season	Fall	0.996	Winter	0.995	0.009	0.008	0.971
Year	2019	0.978	2020	0.986	0.021	0.019	0.105
Working day	Yes	0.935	No	0.929	0.033	0.031	0.207
Meal time	Yes	0.937	No	0.935	0.035	0.030	0.004
Discount	Yes	0.932	No	0.932	0.033	0.030	0.028

Notes: p-values of paired t-tests are reported.

Learning in Experiments We analyze whether there is a learning effect in the experiment. We split 22 rounds of choices in Experiment 1 into the first half and the second half, and find that CCEI scores are lower for the first half of the risky decisions (0.967 vs. 0.981, $p < 0.01$, paired t-test). Using the same method as mentioned above, the underlying preferences for the two halves are also different, contributing to the deviation from utility maximization. Moreover, given that we

[15] We also apply the method by Echenique et al. (2011). Specifically, we calculate the proportion of GARP violations across every two choices within specific pairwise combination, then take the average across consumers. We find the results are similar (Table A9 and A10). By estimating the CES functions across seasons, we further observe $\hat{\alpha}$ in spring is significantly lower than in other three seasons ($p < 0.01$ for all comparisons, paired t-tests), indicating that consumers shift their preferences towards vegetables rather than meat in spring (Table A11). Two possible explanations exist. First, the majority of vegetables reach their optimal freshness in spring. Second, following the Chinese New Year—a significant temporal landmark marking the end of winter in China, individuals tend to set new year resolution, such as improving their diet or adopting healthier eating habits.

randomize the order of all three tasks in Experiment 2, we find that the Spearman's rank correlation between the first and second tasks, and that between the first and third are both 0.284 ($p < 0.01$), whereas that between the last two tasks is 0.350 ($p < 0.01$). These findings suggest that the generalizability increases as participants gain more experience throughout the experiment.

Individual Characteristics We further examine how individual characteristics affect consistency scores in these two settings. The independent variables include Raven's IQ test score, shopping frequency, Big Five personality traits as well as demographic variables including gender, family income, education, age and family size. The dependent variables are CCEI scores (Column 1) and consistency with FOSD (Column 2) in Experiment 1, as well as CCEI scores in Scanner Dataset 1 (Column 3).

Table 8 presents OLS regression results. Our results show that the estimated coefficient of IQ is significantly positive for consistency with FOSD, which is similar to the observation from Cappelen et al. (2023). Moreover, consumers with more shopping experience have significantly higher scores of CCEI in the supermarket. This is in line with List (2003), showing that individuals with more trading experiences exhibit less endowment effect and behave closer to neoclassical predictions, and suggesting that consumers can learn to overcome the endowment effect. Given the maximum score of these measures is 1 and their distributions are skewed, we also apply a Tobit model and use relative rankings as dependent variables separately and show that the results remain robust (Table A12 and A13).

5.4 Heuristic Rules

In addition to preferences, decision makers may use different heuristic rules in response to different environments. In the environment of budgetary choices, Choi et al. (2006) show that subjects' portfolio choices can be explained by some simple rules, such as a diversification heuristic—allocating the same number of points to the two accounts. Halevy and Mayraz (2024) design simple investment rules for selecting portfolios and show that most of the subjects prefer to make allocations through the rule-based interface. It is possible that in the abstract environment of experiments, participants find the decisions to be difficult and adopt a set of heuristic rules to simplify their choices. In food purchasing decisions, consumers may use different heuristic rules such as habits formed in their daily life (Havranek et al., 2017), or reference dependence in perceiving price differences (Thaler and Shefrin, 1981).

To explore such heuristic rules, in Experiment 2, we focus on "middle choosers" who choose

Table 8: OLS Regressions for Consistency Indices in Experiment 1 and Scanner Dataset 1

	Exp: CCEI	Exp: FOSD	Scanner: CCEI
	(1)	(2)	(3)
IQ	0.003	0.013***	-0.000
	(0.002)	(0.004)	(0.001)
Frequency	0.059	0.057	0.241***
	(0.074)	(0.155)	(0.042)
Constant	0.910***	0.673***	0.587***
	(0.025)	(0.054)	(0.047)
Controls	Y	Y	Y
N	1,055	1,055	1,055
R^2	0.023	0.034	0.084

Notes: *IQ* is the number of correct answers in a seven-question version of Raven's Progressive Matrices. *Frequency* is the average monthly shopping days (divided by 100) for meat and vegetables from 2019 to 2020. Control variables include gender dummy, dummies for medium income (family income between 5,001 and 10,000 RMB per month) and high income (family income more than 10,000 RMB per month), dummies for medium education (high school) and high education (above high school), dummies for medium age (born between 1970 and 1989) and elder age (born before 1969), family size, and five variables for the five dimensions in Big Five personality traits. In the scanner dataset, simulated CCEI is added as an additional control to address the heterogeneous power issue. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

near-middle options when the prices for two goods are similar.[16] The results reveal that middle choosers constitute 26.49% of participants in both risk and social tasks. Notably, conditional on being middle choosers in the risk task, the likelihood of being middle choosers in the social task is 45.00%, while this likelihood is only 19.82% among non-middle choosers in the risk task ($p < 0.01$, two-sample proportion test).

Another possible heuristic is "downward-sloping demand", considering the degree of demand change in response to price change. Specifically, we calculate the Spearman's rank correlation between $\ln(x_1/x_2)$ and $\ln(p_1/p_2)$

Spearman $(\ln(x_1/x_2), \ln(p_1/p_2))$

for each individual to measure the compliance with "downward-sloping demand" property (Echenique et al., 2023). The results show that pairwise correlations range from 0.324 to 0.539 in Experiment 2 ($p < 0.01$ for all combinations) (See Panel B in Table 6).[17] This is notable because individuals do not necessarily respond consistently to prices due to specific utility functions in

[16]Specifically, we define middle choosers as participants who select the near-middle options—allocating between 40% to 60% budget to both goods (choosing the 5th, 6th or 7th option on the interface)—across six rounds in the risk task or three rounds in the social task where price ratios (p_1/p_2) range from 0.9 to 1.1. We cannot identify middle choosers in the food task as no price ratios fall within the range.

[17]In addition, we also observe certain correlations of estimated parameters, $\hat{\alpha}$ and $\hat{\rho}$, among tasks within Experiment 2 (Table A14).

particular preference tasks. By contrast, "downward-sloping demand" measures have low or very low correlations in the scanner data.

As the selection of middle options and the adherence to downward-sloping demand can be rationalized by certain utility functions, the observed correlations may be influenced by an underlying connection between risk preference and social preference. However, we believe that these correlations are more likely to arise from similar decision-making modes or heuristics used in response to the comparable budgetary interface presented in both risk and social tasks.

6 Concluding Remarks

Integrating both experimental data and scanner data, we conduct a comprehensive study to investigate the generalizability of consistency scores. Our primary findings reveal a striking lack of correlation between individuals' consistency scores for risky decisions made in the experiment and food consumption choices observed in supermarkets. Within the supermarket context, we observe generally low correlations among food consumption decisions across categories and years; within the experimental setting, we observe on average moderate correlations among different tasks using budgetary interface. We further show that these correlational patterns remain robust when using a wide range of indices of consistency proposed in the literature. Our study adds to the understanding that preferences and consistency derived from observable choices do not necessarily generalize across different environments.

We delve into several unobservable factors that may underpin the consistency scores across different environments, including randomness in choice behaviors, budget constraints, formation of preferences, and heuristic rules. In line with this, Choi et al. (2014) highlight the complexities of measuring decision quality, and argue that assessing decision quality in real-world contexts is challenging because decision makers "might have different preferences over the same outcomes, or face different but unobserved incentives and constraints, or have different information, or hold different beliefs." To address these issues, they propose the use of choice consistency as a measure of decision-making quality, particularly in relation to departures from GARP using experimental data. Their study, along with several subsequent ones, further demonstrate that consistency measure holds real-world significance, even after accounting for unobserved constraints, preferences, and beliefs. Here we present additional results to pin down some underlying factors of consistency measures. In experimental data, we find that consistency is shaped by learning effects, heuristic rules of choosing the middle option and responding to price changes, and the cognitive ability of individuals. In the scanner data, consistency is influenced by contextual factors such as seasonality, the time of day, promotional discounts, as well as individual shopping experiences.

Overall, these findings suggest that the observed lack of correlation in consistency scores between the two settings may stem partly from measurement errors, particularly in how budget constraints are constructed, as well as the influence of contextual factors. This disconnect may be partly due to the multidimensional nature of decision-making quality as well. Specifically, participants in the experiment are often faced with novel and abstract problems, while consumers in supermarkets typically make repeated daily decisions in familiar environments. This distinction closely resembles the difference between the decision-making capacity of a routine task and a problem-solving task (Deming, 2021). It is also reflected in the contrast between fluid intelligence, which is the ability to reason and solve problems in unfamiliar situations, and crystallized intelligence, which involves accumulated knowledge, skills, and experiences (Cattell, 1943).

To sum, our study contributes to a fundamental aspect of economics—choice consistency within the framework of utility maximization under budget constraints (Afriat, 1967; Samuelson, 1938; Varian, 1982). It highlights a disconnect between experimental settings and supermarket environments, moderate correlations within experimental tasks and weak to moderate correlations in supermarket settings. Our results suggest that revealed preference measures have yet to capture the intricate and multidimensional nature in decision-making. Looking ahead, it is crucial for future research to explore alternative measures and frameworks of choice consistency that can more accurately capture the complexities of decision making in controlled laboratory experiments, field experiments, and everyday life.

References

Afriat, Sydney N. 1967. "The construction of utility functions from expenditure data." *International Economic Review*, 8(1): 67–77.

Afriat, Sydney N. 1972. "Efficiency estimation of production functions." *International Economic Review*, 13(3): 568–598.

Ahn, David, Syngjoo Choi, Douglas Gale, and Shachar Kariv. 2014. "Estimating ambiguity aversion in a portfolio choice experiment." *Quantitative Economics*, 5(2): 195–223.

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

Apesteguia, Jose, and Miguel A. Ballester. 2015. "A measure of rationality and welfare." *Journal of Political Economy*, 123(6): 1278–1310.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "'Coherent arbitrariness': Stable demand curves without stable preferences." *Quarterly Journal of Economics*, 118(1): 73–106.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working over time: Dynamic inconsistency in real effort tasks." *Quarterly Journal of Economics*, 130(3): 1067–1115.

Banks, James, Leandro S. Carvalho, and Francisco Perez-Arce. 2019. "Education, decision making, and economic rationality." *Review of Economics and Statistics*, 101(3): 428–441.

Barseghyan, Levon, Jeffrey Prince, and Joshua C. Teitelbaum. 2011. "Are risk preferences stable across contexts? Evidence from insurance data." *American Economic Review*, 101(2): 591–631.

Blundell, Richard, Martin Browning, and Ian Crawford. 2003. "Nonparametric Engel curves and revealed preference." *Econometrica*, 71(1): 205–240.

Blundell, Richard, Martin Browning, and Ian Crawford. 2008. "Best nonparametric bounds on demand responses." *Econometrica*, 76(6): 1227–1262.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. "Memory, attention, and choice." *Quarterly Journal of Economics*, 135(3): 1399–1442.

Bronars, Stephen G. 1987. "The power of nonparametric tests of preference maximization." *Econometrica*, 55(3): 693–698.

Brown, Donald J., and Caterina Calsamiglia. 2007. "The nonparametric approach to applied welfare analysis." *Economic Theory*, 31(1): 183–188.

Bruhin, Adrian, Ernst Fehr, and Daniel Schunk. 2019. "The many faces of human sociality: Uncovering the distribution and stability of social preferences." *Journal of the European Economic Association*, 17(4): 1025–1069.

Burks, Stephen, Jeffrey Carpenter, Lorenz Götte, and Aldo Rustichini. 2012. "Which measures of time preference best predict outcomes: Evidence from a large-scale field experiment." *Journal of Economic Behavior & Organization*, 84(1): 308–320.

Camerer, Colin F. 2015. "The promise and success of lab–field generalizability in experimental economics: A critical reply to Levitt and List." In *Handbook of Experimental Economic Methodology*. Oxford University Press.

Cappelen, Alexander W., Shachar Kariv, Erik Ø Sørensen, and Bertil Tungodden. 2023. "The development gap in economic rationality of future elites." *Games and Economic Behavior*, 142: 866–878.

Carvalho, Leandro, Arna Olafsson, and Dan Silverman. 2024. "Misfortune and mistake: The financial conditions and decision-making ability of high-cost loan borrowers." *Journal of Political Economy*, 132(9): 3173–3213.

Carvalho, Leandro S., Stephan Meier, and Stephanie W. Wang. 2016. "Poverty and economic decision-making: Evidence from changes in financial resources at payday." *American Economic Review*, 106(2): 260–284.

Cattell, Raymond B. 1943. "The measurement of adult intelligence." *Psychological Bulletin*, 40(3): 153–193.

Chambers, Christopher P., and Federico Echenique. 2016. *Revealed preference theory.* Cambridge:Cambridge University Press.

Chapman, Jonathan, Erik Snowberg, Stephanie W. Wang, and Colin Camerer. 2024. "Looming large or seeming small? Attitudes towards losses in a representative sample." *Review of Economic Studies*, rdae093.

Chapman, Jonathan, Mark Dean, Pietro Ortoleva, Erik Snowberg, and Colin Camerer. 2023. "Econographics." *Journal of Political Economy Microeconomics*, 1(1): 115–161.

Chen, Yiting, Tracy Xiao Liu, You Shan, and Songfa Zhong. 2023. "The emergence of economic rationality of GPT." *Proceedings of the National Academy of Sciences*, 120(51): e2316205120.

Cherchye, Laurens, Thomas Demuynck, and Bram De Rock. 2018. "Normality of demand in a two-goods setting." *Journal of Economic Theory*, 173: 361–382.

Cherchye, Laurens, Thomas Demuynck, Bram De Rock, and Joshua Lanier. 2023. "Are consumers (approximately) rational? Shifting the burden of proof." *Review of Economics and Statistics*, forthcoming.

Cherchye, Laurens, Thomas Demuynck, Bram De Rock, and Khushboo Surana. 2020. "Revealed preference analysis with normal goods: Application to cost-of-living indices." *American Economic Journal: Microeconomics*, 12(3): 165–188.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv. 2006. "Substantive and procedural rationality in decisions under uncertainty." *Working Paper*.

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

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. “Who is (more) rational?” *American Economic Review*, 104(6): 1518–1550.

de Clippel, Geoffroy, and Kareen Rozen. 2023. “Relaxed optimization: How close is a consumer to satisfying first-order conditions?” *Review of Economics and Statistics*, 105(4): 883–898.

de Clippel, Geoffroy, Ryan Oprea, and Kareen Rozen. 2024. “As if.” *Working Paper*.

Cohen, Jacob. 1988. *Statistical power analysis for the behavioral sciences (2nd Edition).* New York:Routledge.

Crawford, Ian. 2010. “Habits revealed.” *Review of Economic Studies*, 77(4): 1382–1402.

Crawford, Ian, and Bram De Rock. 2014. “Empirical revealed preference.” *Annual Review of Economics*, 6(1): 503–524.

Dean, Mark, and Anja Sautmann. 2021. “Credit constraints and the measurement of time preferences.” *Review of Economics and Statistics*, 103(1): 119–135.

Dean, Mark, and Daniel Martin. 2016. “Measuring rationality with the minimum cost of revealed preference violations.” *Review of Economics and Statistics*, 98(3): 524–534.

Dean, Mark, and Pietro Ortoleva. 2019. “The empirical relationship between nonstandard economic behaviors.” *Proceedings of the National Academy of Sciences*, 116(33): 16262–16267.

Deb, Rahul, Yuichi Kitamura, John K.-H. Quah, and Jörg Stoye. 2023. “Revealed price preference: Theory and empirical analysis.” *Review of Economic Studies*, 90(2): 707–743.

Dembo, Aluma, Shachar Kariv, Matthew Polisson, and John K.-H. Quah. 2024. “Ever since allais.” *Working Paper*.

Demetry, Marcos, Per Hjertstrand, and Matthew Polisson. 2022. “Testing axioms of revealed preference in Stata.” *Stata Journal*, 22(2): 319–343.

Deming, David J. 2021. “The growing importance of decision-making on the job.” *Working Paper*.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner. 2011. “Individual risk attitudes: Measurement, determinants, and behavioral consequences.” *Journal of the European Economic Association*, 9(3): 522–550.

Dziewulski, Paweł. 2020. “Just-noticeable difference as a behavioural foundation of the critical cost-efficiency index.” *Journal of Economic Theory*, 188: 105071.

Echenique, Federico. 2021. “On the meaning of the critical cost efficiency index.” *Working Paper*.

Echenique, Federico, Sangmok Lee, and Matthew Shum. 2011. “The money pump as a measure of revealed preference violations.” *Journal of Political Economy*, 119(6): 1201–1223.

Echenique, Federico, Taisuke Imai, and Kota Saito. 2023. "Approximate expected utility rationalization." *Journal of the European Economic Association*, 21(5): 1821–1864.

Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R. Cullen. 2012. "How general are risk preferences? Choices under uncertainty in different domains." *American Economic Review*, 102(6): 2606–2638.

Ellis, Andrew, and David J. Freeman. 2024. "Revealing choice bracketing." *American Economic Review*, 114(9): 2668–2700.

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive uncertainty." *Quarterly Journal of Economics*, 138(4): 2021–2067.

Esponda, Ignacio, and Emanuel Vespa. 2024. "Contingent thinking and the sure-thing principle: Revisiting classic anomalies in the laboratory." *Review of Economic Studies*, 91(5): 2806–2831.

Fehr, Ernst, and Andreas Leibbrandt. 2011. "A field study on cooperativeness and impatience in the tragedy of the commons." *Journal of Public Economics*, 95(9-10): 1144–1155.

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

Galizzi, Matteo M., and Daniel Navarro-Martinez. 2019. "On the external validity of social preference games: A systematic lab-field study." *Management Science*, 65(3): 976–1002.

Georganas, Sotiris, Paul J. Healy, and Roberto A. Weber. 2015. "On the persistence of strategic sophistication." *Journal of Economic Theory*, 159: 369–400.

Halevy, Yoram, and Guy Mayraz. 2024. "Identifying rule-based rationality." *Review of Economics and Statistics*, 106(5): 1369–1380.

Halevy, Yoram, Dotan Persitz, and Lanny Zrill. 2018. "Parametric recoverability of preferences." *Journal of Political Economy*, 126(4): 1558–1593.

Harbaugh, William T., Kate Krause, and Timothy R. Berry. 2001. "GARP for kids: On the development of rational choice behavior." *American Economic Review*, 91(5): 1539–1545.

Havranek, Tomas, Marek Rusnak, and Anna Sokolova. 2017. "Habit formation in consumption: A meta-analysis." *European Economic Review*, 95: 142–167.

Heufer, Jan, and Per Hjertstrand. 2019. "Homothetic efficiency: Theory and applications." *Journal of Business & Economic Statistics*, 37(2): 235–247.

Horn, John L., and Raymond B. Cattell. 1966. "Refinement and test of the theory of fluid and crystallized general intelligences." *Journal of Educational Psychology*, 57(5): 253–270.

Houtman, Martijn, and Julian Maks. 1985. "Determining all maximal data subsets consistent with revealed preference." *Kwantitatieve Methoden*, 19(1): 89–104.

Kim, Hyuncheol Bryant, Syngjoo Choi, Booyuel Kim, and Cristian Pop-Eleches. 2018. "The role of education interventions in improving economic rationality." *Science*, 362(6410): 83–86.

Levitt, Steven D., and John A. List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic Perspectives*, 21(2): 153–174.

Levitt, Steven D., John A. List, and Sally E. Sadoff. 2011. "Checkmate: Exploring backward induction among chess players." *American Economic Review*, 101(2): 975–990.

List, John A. 2003. "Does market experience eliminate market anomalies?" *Quarterly Journal of Economics*, 118(1): 41–71.

List, John A., Ian Muir, Devin Pope, and Gregory Sun. 2023. "Left-digit bias at Lyft." *Review of Economic Studies*, 90(6): 3186–3237.

Li, Wenchao, Zhiming Leng, Junjian Yi, and Songfa Zhong. 2023. "A multifaceted poverty reduction program has economic and behavioral consequences." *Proceedings of the National Academy of Sciences*, 120(10): e2219078120.

McGranaghan, Christina, Kirby Nielsen, Ted O'Donoghue, Jason Somerville, and Charles D. Sprenger. 2024. "Distinguishing common ratio preferences from common ratio effects using paired valuation tasks." *American Economic Review*, 114(2): 307–347.

McGrew, Kevin S. 2009. "CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research." *Intelligence*, 37(1): 1–10.

Nielsen, Kirby, and John Rehbeck. 2022. "When choices are mistakes." *American Economic Review*, 112(7): 2237–2268.

Niederle, Muriel, and Emanuel Vespa. 2023. "Cognitive limitations: Failures of contingent thinking." *Annual Review of Economics*, 15(1): 307–328.

Nishimura, Hiroki, Efe A. Ok, and John K.-H. Quah. 2017. "A comprehensive approach to revealed preference theory." *American Economic Review*, 107(4): 1239–1263.

Oprea, Ryan. 2024. "Decisions under risk are decisions under complexity." *American Economic Review*, 114(12): 3789–3811.

Plott, Charles R. 1996. "Rational individual behavior in markets and social choice processes: The discovered preference hypothesis." In *The Rational Foundations of Economic Behavior*. Macmillan.

Polisson, Matthew, and John K.-H. Quah. 2013. "Revealed preference in a discrete consumption space." *American Economic Journal: Microeconomics*, 5(1): 28–34.

Polisson, Matthew, and John K.-H. Quah. 2024. "Rationalizability, cost-rationalizability, and Afriat's efficiency index." *Working Paper*.

Polisson, Matthew, John K.-H. Quah, and Ludovic Renou. 2020. "Revealed preferences over risk and uncertainty." *American Economic Review*, 110(6): 1782–1820.

Rubinstein, Ariel. 2016. "A typology of players: Between instinctive and contemplative." *Quarterly Journal of Economics*, 131(2): 859–890.

Samuelson, Paul A. 1938. "A note on the pure theory of consumer's behaviour." *Economica*, 5(17): 61–71.

Selten, Reinhard. 1991. "Properties of a measure of predictive success." *Mathematical Social Sciences*, 21(2): 153–167.

Snowberg, Erik, and Leeat Yariv. 2021. "Testing the waters: Behavior across participant pools." *American Economic Review*, 111(2): 687–719.

Spearman, C. 1904. "General intelligence, objectively determined and measured." *American Journal of Psychology*, 15(2): 201–293.

Stango, Victor, and Jonathan Zinman. 2023. "We are all behavioral, more or less: Measuring and using consumer-level behavioral sufficient statistics." *Review of Economic Studies*, 90(3): 1470–1498.

Thaler, Richard H., and H. M. Shefrin. 1981. "An economic theory of self-control." *Journal of Political Economy*, 89(2): 392–406.

Varian, Hal R. 1982. "The nonparametric approach to demand analysis." *Econometrica*, 50(4): 945–973.

Varian, Hal R. 1983. "Non-parametric tests of consumer behaviour." *Review of Economic Studies*, 50(1): 99–110.

Weber, Elke U., Ann-Renee Blais, and Nancy E. Betz. 2002. "A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors." *Journal of Behavioral Decision Making*, 15(4): 263–290.

Online Appendices

A Additional Figures and Tables

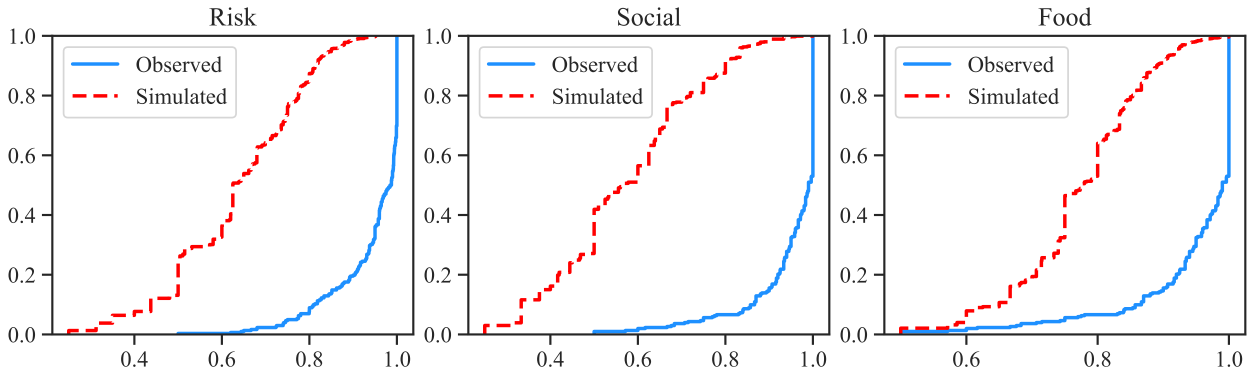


Figure A1: CCEI Scores Across Tasks in Experiment 2

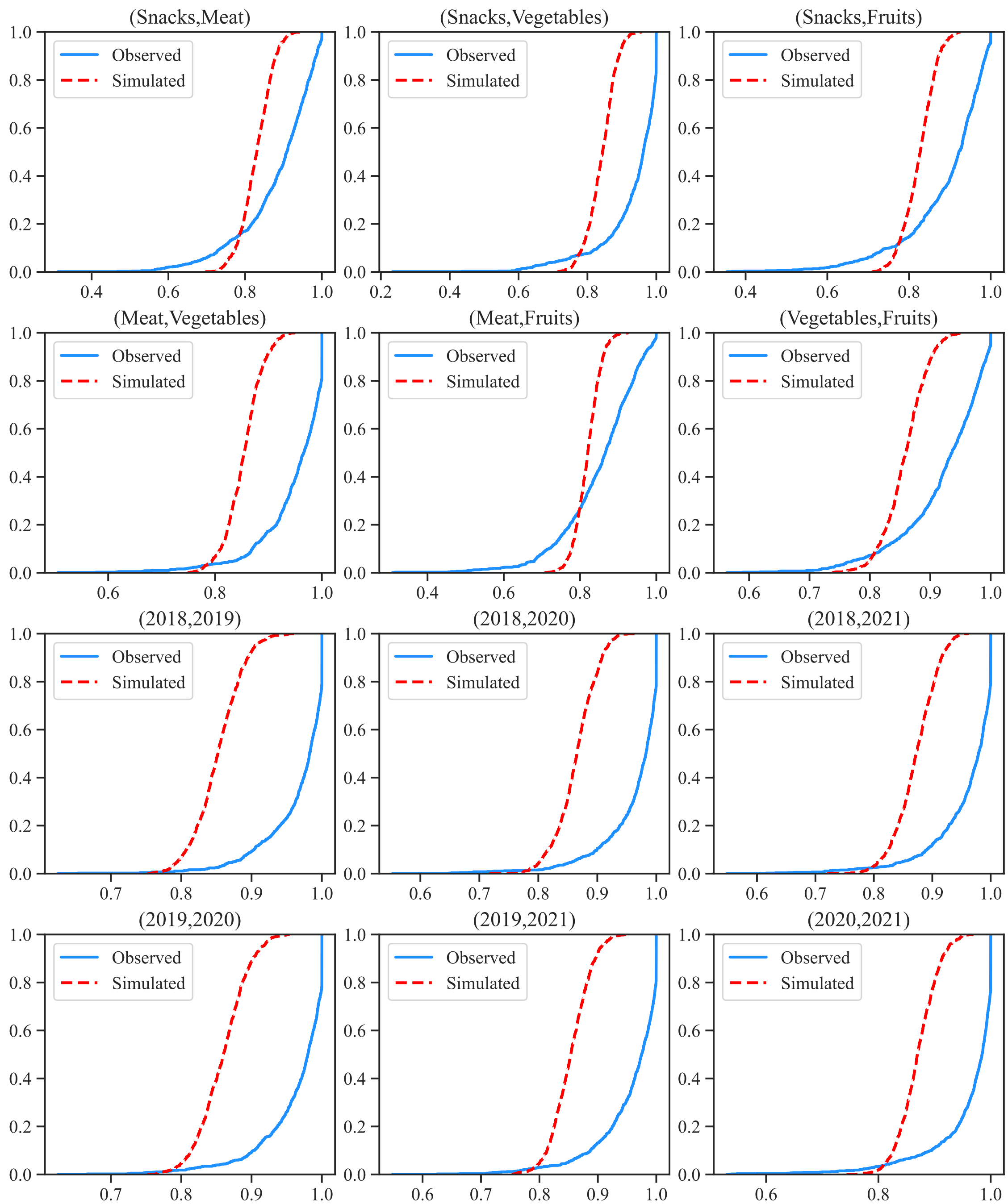


Figure A2: CCEI Scores in Scanner Dataset 2

Table A1: Summary Statistics of Participants in Two Experiments

Panel A: Demographics			
Variable	Category	Percent	
		Experiment 1	Experiment 2
Gender	Female	79.9%	64.6%
	Male	20.1%	35.4%
Age	Before 1969	19.5%	18.5%
	in 1970-1989	75.0%	50.7%
	After 1990	5.5%	30.8%
Family Monthly Income	$<$ RMB 5,000	24.4%	19.2%
	RMB 5,000-10,000	40.0%	29.8%
	$>$ RMB 10,000	35.6%	51.0%
Education	$<$ High school	44.9%	36.8%
	High school	29.8%	19.5%
	$>$ High school	25.3%	43.7%
Family Size	One	1.7%	2.3%
	Two	7.1%	10.6%
	Three	36.9%	35.4%
	Four	31.8%	32.5%
	>Four	22.5%	19.2%
N		1,055	302

Panel B: Consumption behavior (Meat,Veg.) in 2019 and 2020		
Variable	Experiment 1	Experiment 2
# Months with purchase records	24.0	8.9
Monthly expenditure (RMB)	482.2	245.3
Monthly shopping days	11.8	5.1
N	1,055	117

Notes: For participants in Experiment 2, we can match 117 of them who have consumption records of both meat and vegetables in the supermarket in at least one month.

Table A2: Consistency Indices in Experiment 1 and Scanner Dataset 1

	Observed						Simulated					
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max
Panel A: Risk task in Experiment 1												
CCEI	1,055	0.941	0.094	0.350	0.988	1	1,000	0.635	0.148	0.250	0.625	0.952
HMI	1,055	0.883	0.094	0.591	0.909	1	1,000	0.697	0.074	0.455	0.682	0.909
MPI	1,055	0.045	0.060	0	0.020	0.415	1,000	0.191	0.051	0.048	0.192	0.355
MCI	1,055	0.010	0.028	0	0.001	0.435	1,000	0.125	0.077	0.005	0.112	0.623
Panel B: Consumption (Meat,Veg.) in Scanner Dataset 1												
CCEI	1,055	0.946	0.063	0.597	0.965	1	1,055	0.847	0.034	0.750	0.847	0.953
HMI	1,055	0.925	0.053	0.708	0.917	1	1,055	0.860	0.026	0.771	0.860	0.932
MPI	1,055	0.056	0.050	0	0.045	0.294	1,055	0.125	0.022	0.050	0.125	0.203
MCI	1,055	0.004	0.006	0	0.002	0.049	1,055	0.018	0.008	0.002	0.017	0.065

Table A3: CCEI Scores in Experiment 2

	Observed						Simulated					
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max
Risk	302	0.943	0.084	0.500	0.986	1	1,000	0.635	0.148	0.250	0.625	0.952
Social	302	0.927	0.138	0.250	1	1	1,000	0.572	0.162	0.250	0.562	1
Food	302	0.950	0.089	0.500	0.990	1	1,000	0.773	0.096	0.500	0.775	1

Table A4: CCEI Scores in Scanner Dataset 2

	Observed						Simulated					
	N	Mean	SD	Min	Median	Max	N	Mean	SD	Min	Median	Max
Panel A: CCEI across categories												
(Snacks,Meat)	822	0.883	0.099	0.312	0.909	1	822	0.829	0.041	0.698	0.831	0.945
(Snacks,Veg.)	822	0.933	0.090	0.234	0.962	1	822	0.841	0.041	0.715	0.844	0.956
(Snacks,Fruits)	822	0.892	0.102	0.354	0.922	1	822	0.825	0.039	0.711	0.827	0.924
(Meat,Veg.)	822	0.947	0.063	0.506	0.966	1	822	0.853	0.034	0.750	0.854	0.948
(Meat,Fruits)	822	0.850	0.101	0.308	0.867	1	822	0.820	0.033	0.709	0.820	0.925
(Veg.,Fruits)	822	0.920	0.071	0.563	0.936	1	822	0.858	0.035	0.739	0.860	0.953
Panel B: CCEI across time periods												
(2018,2019)	938	0.966	0.046	0.625	0.982	1	938	0.852	0.034	0.753	0.852	0.961
(2018,2020)	938	0.963	0.053	0.553	0.982	1	938	0.864	0.035	0.718	0.865	0.970
(2018,2021)	938	0.959	0.057	0.549	0.979	1	938	0.871	0.036	0.721	0.872	0.961
(2019,2020)	938	0.963	0.049	0.625	0.980	1	938	0.859	0.034	0.753	0.860	0.953
(2019,2021)	938	0.956	0.056	0.549	0.975	1	938	0.853	0.033	0.754	0.854	0.950
(2020,2021)	938	0.961	0.062	0.530	0.985	1	938	0.870	0.034	0.745	0.869	0.968

Table A5: Correlations of Consistency Indices Between Experiment 1 and Scanner Dataset 1

Exp: Risk	Scanner: Food	Correlation
Panel A: Baseline indices		
CCEI	CCEI	-0.011
HMI	HMI	0.050
MPI	MPI	-0.006
MCI	MCI	-0.007
Panel B: Heterogeneous power for budget constraints		
CCEI	Selten Score	-0.003
CCEI	Power-adjusted CCEI	-0.009
Panel C: More (realistic) assumptions on preferences		
Normal	Normal	-0.016
Homothetic	Homothetic	-0.005
CCEI	Quasilinear	-0.060*
FOSD	CCEI	-0.042
EU	CCEI	-0.045
Panel D: Preference estimation		
LL	LL	0.042

Notes: $N = 1,055$. Spearman's rank correlations are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6: Correlations of Consistency Indices in Experiment 2

	Risk vs. Social	Risk vs. Food	Social vs. Food
Panel A: Baseline indices			
CCEI	0.412***	0.305***	0.254***
HMI	0.449***	0.325***	0.306***
MPI	0.408***	0.309***	0.241***
MCI	0.408***	0.339***	0.249***
Panel B: More (realistic) assumptions on preferences			
Normal	0.426***	0.349***	0.324***
Homothetic	0.495***	0.283***	0.204***
FOSD	0.260***	0.187***	–
EU	0.263***	0.191***	–
Panel C: Preference estimation			
LL	0.522***	0.379***	0.383***

Notes: $N = 302$. Spearman's rank correlations are reported. Each coefficient is calculated using the index as a measure for both consistency scores. Two exceptions are FOSD and EU, which are applied exclusively to the risk task; for these, we calculate the correlations between them and CCEI scores for other tasks. – denotes that the measure is not applicable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A7: Correlations of Consistency Indices in Scanner Dataset 2

	Across categories ($N = 822$)			Across time ($N = 938$)		
	(Snacks,Meat) vs. (Veg.,Fruits)	(Snacks,Veg.) vs. (Meat,Fruits)	(Snacks,Fruits) vs. (Meat,Veg.)	(2018,2019) vs. (2020,2021)	(2018,2020) vs. (2019,2021)	(2018,2021) vs. (2019,2020)
Panel A: Baseline indices						
CCEI	0.173***	-0.046	0.096***	0.112***	0.186***	0.143***
HMI	0.061*	-0.016	0.100***	0.096***	0.114***	0.132***
MPI	0.194***	-0.035	0.099***	0.142***	0.220***	0.160***
MCI	0.154***	-0.048	0.102***	0.131***	0.192***	0.156***
Panel B: Heterogeneous power for budget constraints						
Selten Score	0.098***	-0.069**	0.041	0.045	0.127***	0.134***
Power-adjusted CCEI	0.121***	-0.070**	0.016	0.047	0.127***	0.117***
Panel C: More (realistic) assumptions on preferences						
Normal	0.191***	0.214***	0.114***	0.449***	0.502***	0.482***
Homothetic	0.226***	0.033	0.137***	0.218***	0.341***	0.342***
Quasilinear	0.265***	0.329***	0.165***	0.260***	0.379***	0.430***
Panel D: Preference estimation						
LL	0.290***	-0.003	0.172***	0.421***	0.529***	0.468***

Notes: Spearman's rank correlations are reported. Each coefficient is calculated using the index as a measure for both consistency scores. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A8: Correlations of Alternative Indices for Budgets Constraints and Heuristic Rules

	PMI	GAPP	Downward
Panel A: Experiment 1 vs. Scanner Dataset 1			
Risk vs (Veg.,Meat)	0.023	-0.025	-0.029
Panel B: Experiment 2			
Risk vs Social	0.378***	–	0.539***
Risk vs Food	0.220***	–	0.324***
Social vs Food	0.208***	–	0.345***
Panel C: Scanner Dataset 2			
(Snacks,Meat) vs. (Veg.,Fruits)	0.151***	0.092***	0.084**
(Snacks,Veg.) vs. (Meat,Fruits)	0.014	-0.009	0.042
(Snacks,Fruits) vs. (Meat,Veg.)	0.081**	0.083**	0.098***
(2018,2019) vs. (2020,2021)	0.068**	0.198***	0.091***
(2018,2020) vs. (2019,2021)	0.122***	0.235***	0.224***
(2018,2021) vs. (2019,2020)	0.097***	0.279***	0.199***

Notes: Spearman's rank correlations are reported. $-$ denotes that the measure is not applicable. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A9: Average Proportions of GARP Violation Across Combinations of Seasons in Scanner Dataset 1

	Spring	Summer	Fall	Winter
Spring	1.04%	1.02%	1.08%	1.03%
Summer		0.87%	0.93%	0.91%
Fall			0.88%	0.84%
Winter				0.85%

Notes: For each combination of seasons, the number in each cell represents the proportion of month pairs that exhibit violations of GARP within all month pairs, averaged across all consumers.

Table A10: Pairwise Comparisons Across Difference Combination of Seasons in Scanner Dataset 1

	(Spring,Fall)	(Spring,Winter)	(Summer,Fall)	(Summer,Winter)	(Fall,Winter)
(Spring,Summer)	0.200	0.803	0.119	0.043	
(Spring,Fall)		0.269	0.006		0.000
(Spring,Winter)				0.017	0.000
(Summer,Fall)				0.610	0.057
(Summer,Winter)					0.140

Notes: The table reports the pairwise comparisons of results in Table A9 between different season combinations. p-values of paired t-tests are reported.

Table A11: Comparisons of Estimated Parameters for Food Preference Across Seasons in Scanner Dataset 1

		$\hat{\alpha}$			$\hat{\rho}$		
Season: s_1	Season: s_2	Mean (s_1)	Mean (s_2)	p-value	Mean (s_1)	Mean (s_2)	p-value
Spring	Summer	0.371	0.422	0.000	-6.356	-5.816	0.108
Spring	Fall	0.371	0.445	0.000	-6.356	-6.303	0.886
Spring	Winter	0.371	0.425	0.000	-6.356	-6.786	0.224
Summer	Fall	0.422	0.445	0.124	-5.816	-6.303	0.186
Summer	Winter	0.422	0.425	0.861	-5.816	-6.786	0.009
Fall	Winter	0.445	0.425	0.174	-6.303	-6.786	0.197

Notes: p-values of paired t-tests are reported.

Table A12: Tobit Regressions for Consistency Indices in Experiment 1 and Scanner Dataset 1

	Lab: CCEI (1)	Lab: FOSD (2)	Scanner: CCEI (3)
IQ	0.003	0.014***	0.000
	(0.002)	(0.005)	(0.001)
Frequency	0.065	0.036	0.272***
	(0.097)	(0.168)	(0.049)
Conscientiousness	0.001	0.002	-0.002*
	(0.002)	(0.004)	(0.001)
Extraversion	0.001	-0.001	0.001
	(0.002)	(0.004)	(0.001)
Agreeableness	-0.004*	-0.006	0.001
	(0.002)	(0.004)	(0.001)
Openness	-0.004*	-0.006	0.002
	(0.002)	(0.004)	(0.001)
Neuroticism	-0.007***	-0.009**	-0.000
	(0.002)	(0.004)	(0.001)
Female	-0.003	0.027	-0.011*
	(0.010)	(0.019)	(0.006)
Medium Income	-0.006	0.011	0.001
	(0.010)	(0.019)	(0.005)
High income	0.017	0.046**	-0.001
	(0.012)	(0.020)	(0.006)
Medium education	-0.003	-0.005	-0.003
	(0.010)	(0.017)	(0.005)
High education	0.011	0.030	-0.009
	(0.011)	(0.019)	(0.006)
Medium age	0.022	0.063*	-0.005
	(0.020)	(0.036)	(0.010)
Elder age	0.018	0.068*	-0.006
	(0.022)	(0.038)	(0.010)
Family size	-0.001	-0.004	0.001
	(0.004)	(0.007)	(0.002)
Simulated CCEI			0.503***
			(0.062)
Constant	0.924***	0.680***	0.507***
	(0.032)	(0.058)	(0.055)
N	1,055	1,055	1,055

Notes: The upper limit for Tobit regressions is 1. *IQ* is the number of correct answers in a seven-question version of Raven's Progressive Matrices. *Frequency* is the average monthly shopping days (divided by 100) for meat and vegetables from 2019 to 2020. Dummy for medium (high) income indicates family income between RMB 5,001 and 10,000 (more than RMB 10,000) per month. Dummy for medium (high) education indicates education degree of high school (above high school). Dummy for medium (elder) age refers to those born between 1970 and 1989 (before 1969). In the scanner dataset, simulated CCEI is added as an additional control to address the heterogeneous power issue. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A13: OLS Regressions for Relative Ranking of Consistency Indices in Experiment 1 and Scanner Dataset 1

	Lab: CCEI (1)	Lab: FOSD (2)	Scanner: CCEI (3)
IQ	0.005	0.015***	0.005
	(0.006)	(0.006)	(0.005)
Frequency	0.121	0.040	1.158***
	(0.254)	(0.221)	(0.220)
Conscientiousness	-0.004	-0.002	-0.010*
	(0.006)	(0.005)	(0.005)
Extraversion	0.003	0.001	0.000
	(0.006)	(0.005)	(0.005)
Agreeableness	-0.008	-0.006	0.005
	(0.006)	(0.005)	(0.005)
Openness	-0.008	-0.006	0.004
	(0.006)	(0.005)	(0.005)
Neuroticism	-0.017***	-0.014***	-0.003
	(0.006)	(0.005)	(0.005)
Female	-0.012	0.013	-0.033
	(0.028)	(0.025)	(0.024)
Medium Income	-0.023	0.005	0.016
	(0.027)	(0.024)	(0.023)
High income	0.047	0.053**	0.004
	(0.031)	(0.027)	(0.026)
Medium education	0.023	-0.001	-0.013
	(0.025)	(0.022)	(0.022)
High education	0.032	0.042	-0.044*
	(0.031)	(0.027)	(0.026)
Medium age	0.045	0.068	-0.031
	(0.048)	(0.043)	(0.042)
Elder age	0.034	0.068	-0.029
	(0.053)	(0.047)	(0.045)
Family size	-0.001	-0.011	-0.000
	(0.011)	(0.010)	(0.010)
Simulated CCEI			2.407***
			(0.256)
Constant	0.474***	0.386***	-1.625***
	(0.081)	(0.074)	(0.225)
N	1,055	1,055	1,055
R^2	0.027	0.033	0.108

Notes: Dependent variables are $Relative\ ranking_n = (N - rank(consistency_n) + 1)/N$, where n denotes subject, N is the number of observations, and $rank(consistency_n)$ is the rank of subject n's consistency index in descending order. *IQ* is the number of correct answers in a seven-question version of Raven's Progressive Matrices. *Frequency* is the average monthly shopping days (divided by 100) for meat and vegetables from 2019 to 2020. Dummy for medium (high) income indicates family income between RMB 5,001 and 10,000 (more than RMB 10,000) per month. Dummy for medium (high) education indicates education degree of high school (above high school). Dummy for medium (elder) age refers to those born between 1970 and 1989 (before 1969). In the scanner dataset, simulated CCEI is added as an additional control to address the heterogeneous power issue. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A14: Correlations of Estimated Parameters in Experiment 2

	Risk vs Social	Risk vs Food	Social vs Food
$\hat{\alpha}$	0.150***	0.109*	0.067
$\hat{\rho}$	0.366***	0.081	0.160***

Notes: Spearman's rank correlations are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Economic Specifications and Estimations of Preferences

Given any period, a DM's choices are denoted as (x_1, x_2). For risk preference, x_1 (x_2) corresponds to the better (worse) outcome of contingent assets, respectively. Regarding social preference, x_1 (x_2) represents her own (opponent's) payment. For food preference, they are the consumption quantities on two types of food. For risk preference, we apply the commonly used disappointment aversion (DA) model, which is equivalent to rank dependent utility when there are two equiprobable states, specified in Equation 3; while for social and food preferences, we use the constant elasticity of substitution (CES) utility function (Fisman et al., 2007) specified in Equation 4.

$$U(x_1, x_2) = \alpha u(x_1) + (1-\alpha)u(x_2), \alpha \in [0,1], u(z) = \begin{cases} \frac{1}{\rho}z^\rho, \rho \leq 1(\rho \neq 0) \\ \ln(z), \rho = 0 \end{cases} \tag{3}$$

$$U(x_1, x_2) = [\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{\frac{1}{\rho}}, \alpha \in [0,1], \rho \leq 1 \tag{4}$$

Specifically, α captures the relative weight on better outcome in risk task, the degree of selfishness for social choices, and the weight on the first commodity in food purchase. ρ indicates the degree of risk-seeking in risk preference. For social and food preferences, ρ captures the curvature of the indifference curves. As ρ approaches 1, the two accounts (goods) become perfect substitutes. Conversely, as ρ approaches negative infinity, the two accounts (goods) become perfect complements.

Referring to the method by Fisman et al. (2007), the demand function for the two utility models is given by

$$x_1 = \left[\frac{g}{(p_1/p_2)^m + g}\right]\frac{E}{p_1},$$

where E is the expenditure, $m = \rho/(1-\rho)$, and $g = [\alpha/(1-\alpha)]^{1/(1-\rho)}$. This generates the following econometric specification for estimation:

$$\frac{p_1^t x_1^t}{E^t} = \frac{g}{(p_1^t/p_2^t)^m + g} + \varepsilon^t,$$

where ε^t is assumed to be normally distributed with mean zero and variance σ^t. We generate the estimates $\hat{g}$ and $\hat{m}$, using the nonlinear Tobit maximum likelihood method, then use these to calculate parameters $\hat{\alpha}$ and $\hat{\rho}$, and regard log likelihood (*LL*) as a measure of goodness of fit.

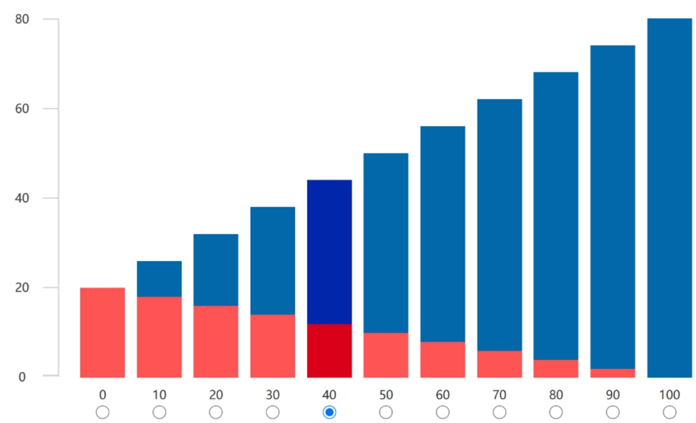
C Experimental Instructions

C.1 Risk Task

As your choices may influence the bonus you receive, please ensure that you understand the following instructions and answer carefully.

- This experiment consists of 22 rounds. In every round, you have 100 tokens, which can be allocated between the blue membership account (Blue Account) and red membership account (Red Account).
- In the following example:
  - Blue Account: 1 token = RMB 0.8, that is, each token for the Blue Account is worth RMB 0.8
  - Red Account: 1 token = RMB 0.2, that is, each token for the Red Account is worth RMB 0.2

1 token = RMB 0.80 for Blue Account and 1 token = RMB 0.20 for Red Account, which allocation would you choose?



The current choice indicates that
You will have a 50% chance of getting RMB 32.0
and another 50% chance of getting RMB 12.0

- As shown in the figure, you need to create an allocation plan. Different allocation plans influence the amount of money in your accounts, that is,

- if 0% of the tokens are allocated to the Blue Account and 100% to the Red Account, then there will be RMB 0 in the Blue Account and RMB 20 in the Red Account;
  - if 10% of the tokens are allocated to the Blue Account and 90% to the Red Account, then there will be RMB 8 in the Blue Account and RMB 18 in the Red Account;
  - ...
  - if 100% of the tokens are allocated to the Blue Account and 0% to the Red Account, then there will be RMB 80 in the Blue Account and RMB 0 in the Red Account.

- In each decision-making round, you have a 50% chance of getting the money in the Blue Account and another 50% chance of getting the money in the Red Account. As shown in the figure, when you choose 40 (40% of the tokens are allocate to the Blue Account and 60% to the Red Account), you will find the following note below the bar graph, "You will have a 50% chance of getting RMB 32.0 and another 50% chance of getting RMB 12.0". When you change the option, the amounts of money in the note will change accordingly.

- Different choices will lead to different gains and risks.
  - The gain is reflected in the total length of the two-colored bars: the longer the bar is, the greater the gain is; and the shorter the bar is, the smaller the gain is.
  - The risk is reflected in the difference between the length of the two colors within the bar: the greater the difference is, the higher the risk is; and the smaller the difference is, the lower the risk is.

- At the end of the experiment, the computer will select one decision round randomly, where each account has an equal probability of being chosen, and the participant will be paid the amount he/she has earned in that round.

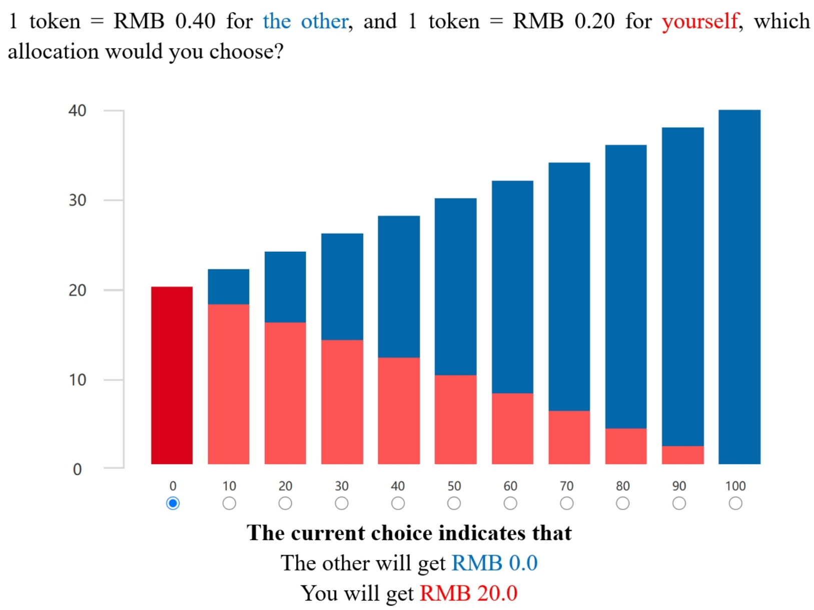
C.2 Social Task

As your choices may influence the bonus you receive, please ensure that you understand the following instructions and answer carefully.

- This experiment consists of 22 rounds. In every round, you have 100 tokens, which can be allocated between yourself and another supermarket consumer.

- In the following example:

- 1 token = RMB 0.4 for the other
- 1 token = RMB 0.2 for yourself



- As shown in the figure, you need to create an allocation plan. Different allocation plans influence the amount of money between yourself and the other, that is,
  - if 0% of the tokens are allocated to the other and 100% to yourself, then the other will have RMB 0, and you will have RMB 20;
  - if 10% of the tokens are allocated to the other and 90% to yourself, then the other will have RMB 4, and you will have RMB 18;
  - ...
  - if 100% of the tokens are allocated to the other and 0% to yourself, then the other will have RMB 40, and you will have RMB 0.
- As shown in the figure, when you choose 0 (0 % of the tokens are allocate to the other and 100% to yourself), you will find the following note below the bar graph, "The other will get RMB 0.0, and you will get RMB 20.0". When you change the option, the amounts of money in the note will change accordingly.

- At the end of the experiment, the computer will select one decision round randomly, and the participant and another randomly matched participant will be paid the amounts they have earned in that round.

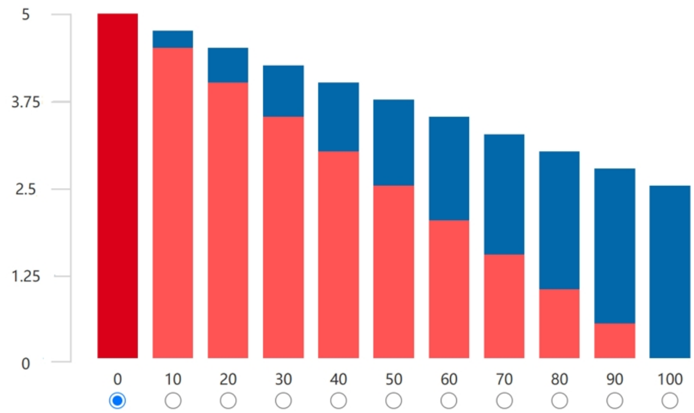
C.3 Food Task

As your choices may influence the bonus you receive, please ensure that you understand the following instructions and answer carefully.

- This experiment consists of 22 rounds. In every round, you have RMB 50, which can be allocated to buy tomatoes and hams.
- In the following example:
  - tomato: RMB 10/kg
  - picnic ham: RMB 20/kg


- As shown in the figure, you need to create an allocation plan. Different allocation plans influence the amount of picnic hams and tomatoes you buy, that is,
  - if 0% of the expenditure is allocated to hams and 100% to tomatoes, then you will buy 0 kg of picnic hams and 5 kg of tomatoes;
  - if 10% of the expenditure is allocated to hams and 90% to tomatoes, then you will buy 0.25 kg of hams and 4.5 kg of tomatoes;
  - ...
  - if 100% of expenditure is allocated to hams and 0% to tomato, then you will buy 2.5 kg of hams and 0 kg of tomatoes.

1 kg of hams = RMB 20, and 1 kg of tomatoes = RMB 10, which allocation would you choose?



The current choice indicates that
You will get 0.0 kg of hams
and get 5.0 kg of tomatoes

- As shown in the figure, when you choose 0 (0% of the expenditure is allocated to hams and 100% to tomatoes), you will find the following note below the bar graph: "You will get 0.0 kg of hams and 5.0 kg of tomatoes". When you change the option, the amounts of products in the note will change accordingly.
- At the end of the experiment, the computer will select one decision round randomly, and the vouchers of products the participant has earned in that round will be deposited into his/her membership account.

C.4 Post-Experiment Questionnaire

- Big Five Personality Traits Test
  (7 Likert Scale) I see myself as someone who
  - is reserved
  - is generally trusting
  - tends to be lazy
  - is relaxed, handles stress well
  - has few artistic interests
  - is outgoing, sociable
  - tends to find fault with others
  - does a thorough job
  - gets nervous easily
  - has an active imagination
- Raven's IQ Test
  In each of the following questions, a part of the graph is missing from its lower right side, please find the appropriate graph to fill in the gap. There is only one correct answer for each question.

1、

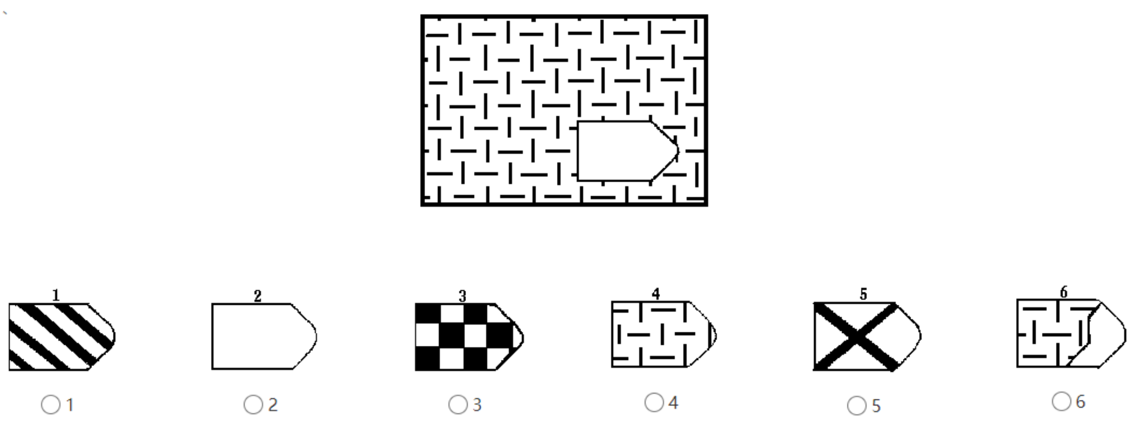


Figure C1: Sample Question of Raven's IQ Test

- Demographics

- Gender:
- Year of Birth:
- Number of household members in your household:
- Your Hukou (residency status) is: Urban/Rural
- Individual monthly income:
- Household monthly income:
- Your education level: