Joint vs. Individual Performance in a Dynamic Choice Problem *

Logan Miller∓ **Ryan Rholes†**

University of Arkansas University of Oxford

May 2023

Abstract

This paper compares the relative ability of individuals and pairs to solve a finite, stochastic lifecycle problem that requires borrowing and saving to achieve the rational benchmark. We find that pairs significantly outperform individuals, especially when allowing subjects to account for past mistakes along conditionally-optimal consumption paths. Joint decision-makers out-earn individuals by about 23%. Though pairs and individuals both overreact to income and wealth balances, these distortions are twice as large for individuals. Analyzing chat data reveals that pairs bargain to balance idiosyncratic consumption preferences, which reduces consumption errors. We estimate consumption heuristics at the observation level and study their dynamics. We show that about half our subjects (or pairs of subjects) stick to heuristics for the majority of the experiment. These 'stable' subjects significantly outperform their 'unstable' counterparts in the dynamic optimization task. Finally, we provide suggestive evidence that subjects who have a nuanced view of debt outperform subjects who think of debt as always bad, even after controlling for cognitive ability.

JEL classifications: C91, C92, D12, D15, D16, E21

Keywords: Individual Behavior, Group Behavior, Intertemporal Household Choice; Life Cycle Models and Saving, Collaborative Consumption, Consumption; Saving; Wealth

*We thank Catherine Eckel, Don Koh, John Duffy, Luba Petersen, Sherry Li, and Pete McGee for helpful comments. We thank seminar participants at the 7th Annual TExAS conference and the University of Arkansas for valuable feedback. Finally, we also thank Tenley Getschman, Caleb Hill, Colton Atha, and Evan Graves for their outstanding research assistance.

† Rholes: Department of Economics, University of Oxford, 10 Manor Road, Oxford, Oxfordshire, England, ryan.rholes@economics.ox.ac.uk. Corresponding author.

∓ Miller: Department of Economics, University of Arkansas, Walton College of Business, Fayetteville, AR, Lmiller@walton.uark.edu.

1 Introduction

Modern macroeconomic theory typically models decision-makers as rational individuals who solve complex, dynamic choice problems. However, many important real-world macroeconomic decisions are made jointly, and distinguishing between individual and joint decision-making can matter. For example, in policy decisions (Blinder and Morgan, 2005, 2007; Lombardelli et al., 2005), in strategic decision-making within firms (Cooper and Kagel, 2005), and in selecting common consumption streams (Jackson and Yariv, 2014) while accounting for heterogeneous time-discounting preferences. In the context of intertemporal choice – typically studied in the lab using individual decision-makers (Duffy, 2016) – Bourguignon et al. (2009) shows households comprising several adult decision-makers do not behave as a representative agent and instead balance consumption preferences via joint decisions. This aligns with Cesarini et al. (2017), Fortin and Lacroix (1997), and Browning et al. (1994), who all reject the unitary model in favor of collective models of intra-household decision-making. An interesting question then is how and why individual and joint decision-making differs in the context of intertemporal resource allocation, which is hard to answer using observational data alone, especially if inputs into the decision-making process are unobservable.

We elucidate the issue using evidence from the Panel Study of Income Dynamics (PSID) data, where poor married households better smooth their consumption in response to negative income shocks than poor, single households. Why? Maybe financial constraints are more likely to bind on poor, single households (e.g. if married households have dual incomes).[1] However, it might also be that cognitive constraints are more likely to bind for poor single households following negative income shocks since married households can form decisions jointly.[2] Understanding the role of these two frictions matters from a policy perspective.

Disentangling these two channels using observational data alone is often impossible due to data limitations regarding spousal employment, cognitive limitations, and the process by which households actually make decisions. Further exacerbating the issue is that current experimental evidence suggests that joint decision-making does not necessarily improve performance in a dynamic optimization task (Carbone and Infante, 2015; Carbone et al., 2019), which is a conclusion at odds with the broader literature on joint decision-making.

In this paper, we revisit dynamic optimization in the laboratory to investigate the ability of individuals and pairs to dynamically optimize. Because pairs and individuals face identical environments, structural differences cannot account for any treatment-level differences we observe. Our interests are two-fold. First, we are interested in whether joint consumption decisions more closely align with the rational benchmark than individual decisions in a setting where joint decision-makers have identical bargaining power and consumption utility. Second, we are interested in *how* pairs form joint decisions and – if they are – *why* they are different from individuals.

To do this, we implement two experimental treatments using a between-subjects design where we compare the ability of individuals and pairs to solve a finite lifecycle problem.

[1] We provide details about this suggestive evidence in Section 7.4 of the appendix.

[2] For example, if poverty impedes cognitive function (Haushofer and Fehr, 2014; Mani et al., 2013). Evidence on this is mixed (i.e. Carvalho et al. (2016)).

Subjects in our experiment receive stochastic income and must save and borrow to smooth consumption. In our *Individuals* treatment, participants form independent decisions. In our *Pairs* treatment, we form subjects into stable pairs who then engage in unrestricted chat to form joint decisions.

We find that pairs significantly outperform individuals relative to the rational, representative benchmark. This is true when we measure performance along the unconditionally-optimal consumption path and even more pronounced along the conditionally-optimal consumption path where we allow participants to correct past mistakes. Additionally, we find that, on average, joint-decision making in our experiment leads participants to earn about 23% more than individual decision-making after accounting for fixed show-up fees paid to all subjects.

One reason for this performance difference might be that participants in *Pairs* had the chance to carefully consider and then articulate the logic of their choices when forming a joint decision. To test this, we conduct a third treatment, *Ledger*, wherein subjects make individual consumption decisions but can use the chat window used in *Pairs* to articulate the logic of their decisions in each period before making a consumption decision. Behavior in *Ledger* is statistically indistinguishable from behavior in *Individuals*.

Then why is performance in *Pairs* better than in *Individuals*? Participants' consumption in all treatments too closely tracks income and overreacts to wealth accumulation. However, joint decision-making reduces the magnitude of these biases by approximately 50%. Participants in our *Pairs* treatment engage in unrestricted communication via a chat window to form joint decisions. Analysis of this chat data reveals that participants in *Pairs* bargain to form joint decisions that balance idiosyncratic preferences. This bargaining process significantly reduces consumption errors relative to the rational benchmark and the conditionally-optimal consumption path by both reigning in overspending but also by significantly reducing under-spending. We are not the first to study chat data in a dynamic choice setting - Carbone and Infante (2015) use chat data from joint decision makers to develop a set of consumption heuristics they use in a heuristic classification exercise. However, ours is the first paper to analyze chat data directly and document bargaining as a mechanism by which joint decision-making improves performance in a dynamic optimization task.

We concluded each experimental session with a post-experiment survey of decisions that, elicited subjects' subjective outlook on debt via free-form answers. Interestingly, we find that about half of our subjects view debt as explicitly bad while the remaining half takes a more nuanced outlook on debt. Using these classifications, we provide suggestive evidence that subjects with a more nuanced outlook on debt outperform their counterparts in our optimization task. This is true even after controlling for cognitive ability, suggesting this does not necessarily result mechanically from the fact that cognitively sophisticated subjects also hold more nuanced economic perspectives.

Building on previous optimization experiments, we classify our subjects (or pairs of subjects) into a set of consumption heuristics common to the learning-to-optimize literature. However, we study the dynamics of these heuristics and show that they are unstable for about half of our observations. We then classify observations into either stable or unstable heuristic types and show that observations with stable heuristics significantly outperform observations with unstable heuristics.

The rest of this paper is organized as follows: Section 2 provides a more thorough review of the literature, Section 3 outlines the theoretical model underpinning our experiment, Section 4 describes experimental design, Section 5 details our results, and Section 6 concludes.

2 Literature Review

There is extensive literature, thoroughly discussed in Duffy (2016) and summarized here, that studies the ability of individuals to solve dynamic stochastic optimization problems. Generally, subjects deviate considerably from the optimal consumption path.

Others have studied dynamic optimization with stochastic income in the lab. Hey and Dardanoni (1988) show subjects fail to optimize in response to a stochastic income, a no-borrowing constraint, and a constant rate of return on savings. Carbone and Hey (2004) and Carbone (2006) simplify this design by eliminating discounting and by simplifying the stochastic income process and find these reductions in the complexity of the lifecycle problem do not align subject behavior with the rational consumption path. Carbone and Infante (2014) study dynamic optimization under certainty, risk, and ambiguity and find that subjects significantly under-consume when faced with ambiguity relative to risk and certainty. Carbone et al. (2021) study consumption smoothing in a Lucas Tree model where subjects trade consumption claims via a long-lived asset, with an alternative solution, where agents can trade short-lived consumption claims between periods. They find the exchange economy with short-lived assets is more efficient in encouraging consumption smoothing.

There is also an established literature studying dynamic optimization by allowing individual decision-makers to interact in various capacities. Ballinger et al. (2003) provide evidence in support of inter-generational learning in the context of dynamic choice via a 60-period life-cycle problem under income uncertainty. The authors grouped subjects into three-member "families" and randomly assigned each family member to either the first, second, or third generation. Members of the first generation had no opportunity to learn. However, members of subsequent generations could both observe and communicate with members of the previous generation for several periods before beginning to make their own decisions. This generational transmission of information improves the decisions of subsequent generations. Our study differs from theirs in that subjects in our *Pairs* treatments do not pass along knowledge but instead work together to generate knowledge, and our *Pairs* subjects form joint decisions and share the payoff of this joint decision.

Brown et al. (2009) show that allowing for social learning improves the speed of own-learning compared to rates of own-learning from subjects in private-learning treatments. In contrast, Carbone and Duffy (2014) shows that revealing the average level of past consumption causes subjects to deviate further from both the conditionally- and unconditionally-optimal consumption path. Bao et al. (2013) show that pairing subjects together and having each subject either forecast or optimize leads to quicker convergence to the rational expectations equilibrium than does having a single subject perform both tasks. Duffy and Orland (2021) test a buffer stock model in the lab and show that imposing liquidity constraints does not increase savings but higher income variation does.

Ubiquitous across these previous studies is the use of individual decision-makers. However, there are also studies comparing the behavior of groups in macroeconomic settings. For example, Blinder and Morgan (2005) and Blinder and Morgan (2007) show that groups outperform individuals setting monetary policy to stabilize an experimental economy around inflation and employment targets, that this result is robust when increasing group size to eight members, and that groups do not need a well-defined leader to produce this result. Lombardelli et al. (2005) corroborates this result and also shows that groups outperform individuals as policymakers because group decision-making moderates policy errors and because groups facilitate social learning. Similarly, Rholes and Petersen (2021) show in a learning-to-forecast experiment that aggregating over group expectations produces more stable inflation dynamics than do individual expectations.

Most closely related to our work are Carbone and Infante (2015) and Carbone et al. (2019), which both study differences between pairs and individuals in a dynamic optimization setting. The former concludes that stable pairs perform no differently than individuals in solving the lifecycle problem once experienced and that pairs with re-matching perform worse than individuals. Additionally, they find that performance differences only exist whenever they consider a rational benchmark that does not allow for corrections of past errors. We find the opposite – stable pairs in our experiment consistently outperform individuals as planners, relative to the rational benchmark and when allowing participants to account for previous errors, even after gaining experience. The latter compares group and individual performance in an optimization task while facing either risk or ambiguity and find that groups are better planners under ambiguity, but individuals are better planners under risk.

Because participants in our experiment face a stochastic income process with a known distribution and support, our experiment best matches decisions under risk in Carbone et al. (2019). However, our design differs drastically from Carbone et al. (2019) where subjects face a bimodal income distribution and, importantly, cannot borrow to smooth consumption. Thus, one can interpret our results as complimentary to their results under risk in that we show joint decision-making improves performance in a dynamic stochastic optimization problem that requires (and allows for) borrowing to achieve a rational benchmark.

Finally, we also contribute to the extensive literature that studies differences between groups and individuals, which is summarized by Charness and Sutter (2012), who conclude that pairs typically better align with game-theoretic predictions and exhibit higher cognitive sophistication. Canonical examples are Cooper and Kagel (2005), who find that teams form more strategic decisions than individuals, and Kugler et al. (2007) who show groups are less trusting than individuals but are equally trustworthy. More recent examples are Kagel and McGee (2016) who show that teams can better learn cooperative strategies in finitely repeated prisoner dilemma games and Chakraborty and Fenig (2022) who show that members of teams provide more effort than do individuals when completing the same remote work, especially when facing a cooperative rather than competitive incentive scheme.

3 Theory

Subjects in both our *Individuals* and *Pairs* treatments maximize their discounted lifetime utility, subject to an intertemporal budget constraint:

$$max\ \mathbb{E}_0 \sum_{t=1}^{t=T} \beta^t U(c_t) \tag{1}$$

$$s.t.\ \sum_{t=1}^{t=T} c_t \leq \sum_{t=1}^{t=T} w_t + a_0 \tag{2}$$

where c_t is consumption, a_0 is initial wealth, and w_t is an i.i.d. per-period stochastic income with $w \sim U\{\underline{w}, \overline{w}\}$.[3] Subjects in our experiment save freely and borrow up to $\underline{w}$ in all but the final decision period. We denote saving and borrowing throughout as s_t.

We induce the quadratic utility function[4]

$$U(c_t) = \phi c_t - \frac{1}{2}c_t^2. \tag{3}$$

This functional form is useful for several reasons. First, it allows subjects to consume zero in any period without incurring negative utility. Second, it is concave across the action space, which induces a consumption smoothing motive.[5] Finally, combining this functional form with equations (1) and (2) above yields the stochastic Euler equation from Hall (1978):

$$c_t = (1 - \kappa)\phi + \kappa \mathbb{E}_t c_{t+1} \tag{4}$$

where $\kappa \equiv \beta(1 + r)$. We set $\beta = 1,\ r = 0$ in order to reduce the complexity of our choice problem, which reduces Equation (4) to the consumption Euler equation:

$$c_t = \mathbb{E}_t c_{t+1}. \tag{5}$$

Solving by backward induction yields our unconditionally-optimal consumption path. [6]

[3] Income is drawn from a discrete uniform distribution so that per-period income is always an integer value.

[4] Carbone and Infante (2015) and Carbone et al. (2019) both induce C.A.R.A.-type utility functions. Our utility relative to theirs yields a very simple optimal consumption rule, which we describe in Equation (5). A possible trade-off is that quadratic utility can lead to increasing absolute risk aversion. To offset this, we elicit risk measures both before and after subjects complete our lifecycle problems so that we can control for risk in our regression analysis.

[5] Restrictions on ϕ are such that, across the feasible action space, the first derivative of $u(c_t)$ is strictly positive and the second derivative is strictly negative. This means that subjects in our experiment can never consume beyond the bliss point regardless of how much wealth they accumulate.

[6] Notice that if $r > 0$ then per-period consumption is lower and per-period savings are higher in most periods. This might lead to behavior similar to that found in Carbone and Infante (2015), who include a positive rate of return on savings in their experimental design.

$$c_{T-j} = \begin{cases} y_{T-j} + s_{T-j-1}, & j = 0 \\ \frac{j}{j+1}\mu + \frac{1}{j+1}(y_{T-j} + s_{T-j-1}), & j \in (1, 2, ..., T-1) \end{cases} \tag{6}$$

This solution indicates that optimal consumption is a linear function of the mean of the income distribution, μ, and period wealth. Intuitively, subjects should focus less on income and more on wealth as the game nears completion. We plot the unconditionally-optimal consumption path alongside the income processes used in all experimental sessions in Figure 1. The unconditionally-optimal path is the same for all subjects because we hold the stochastic income processes constant across all subjects.

We also consider subjects' decisions relative to a conditionally-optimal level of consumption, $\hat{c_t^*}$, which accounts for past consumption errors by recalculating optimal consumption for each remaining period conditional on past mistakes.[7]

$$\hat{c_t^*} = c_t^* + \frac{(y_t - c_t^*) + s_{t-1}}{T-(t-1)}, \ \forall\ t\ \in\ \{2, ..., T-1\} \tag{7}$$

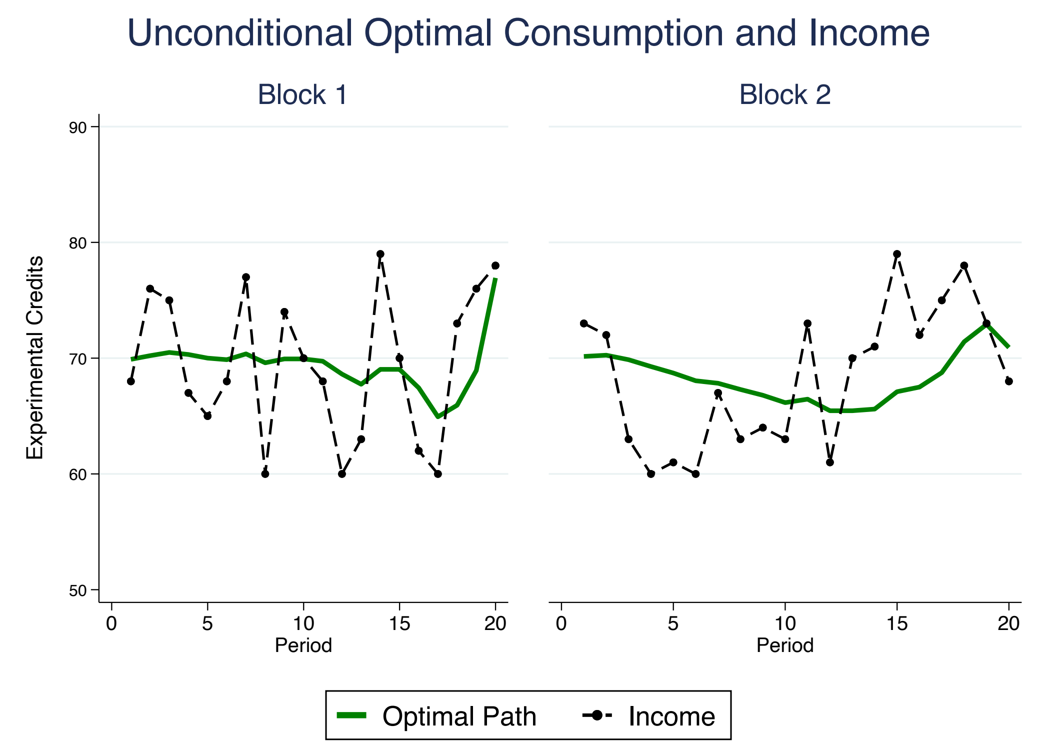


Figure 1: This figure shows the unconditionally-optimal consumption path for decision lifecycles 1 and 2 of all experimental sessions. The graph also includes the pre-drawn stochastic income processes used for lifecycles 1 and 2 in all experimental sessions.

[7] We do not plot the conditionally-optimal path here since it depends on individual deviations from the unconditionally-optimal consumption path.

4 Experimental Design

We use a between-subjects design to implement three experimental treatments built around a standard learning-to-optimize (LTO) framework where we compare the ability of individuals and pairs to solve two different twenty-period versions of the finite lifecycle problem outlined in Section 3. We set $\phi = 1,600$, $\overline{w} = 80$, $\underline{w} = 60$, $\beta = 1$, $r = 0$ for all sessions. We set the discount rate $\beta = 1$ and the nominal interest rate $i = 0$ to reduce the complexity of the choice problem. We choose ϕ so that subjects could never consume beyond the bliss point for any possible income draw. We noted above that participants could borrow up to the $\underline{w} = 60$ in all but the final period of each lifecycle. Importantly, our pre-drawn stochastic income processes are such that this per-period borrowing constraint never forces an optimizing participant off her stochastic Euler equation (Equation (4)). We illustrate the optimal level of borrowing that corresponds to the unconditionally-optimal consumption path in Figure 6, which is located in Section 7.1.

Participants in our *Individuals* treatment made decisions alone when solving both lifecycle problems. For the *Pairs* treatment, we randomly formed stable pairs[8] and allowed each pair to engage in unrestricted chat to solve the lifecycle problems. Importantly, we required participants in the *Pairs* treatment to reach a consensus decision in each period.

In addition to our two primary treatments, we include a third treatment titled *Ledger*. There are at least two possible explanations for why participants in Pairs might outperform participants in *Individuals*. First, subjects in a pair are able to discuss strategies and exchange ideas, and sometimes balance preferences in order to form a joint decision. Second, subjects in the *Pairs* treatments may carefully consider the optimization problem in order to communicate with an assigned partner. Thus, one could question if pairs do better because they are making a joint decision or instead because they are forced to more carefully consider and explain their spending, saving, and borrowing decisions.

Ledger allows us to distinguish between these two potential mechanisms. *Ledger* is identical to *Individuals*, except that subjects in *Ledger* have access to the same chat window as do subjects in Pairs. We encourage participants in *Ledger* to use this chat window as a sort of journal to articulate the logic of their individual decisions. The intuition is that if the process of carefully considering and articulating decisions drives treatment differences between *Pairs* and *Individuals*, then we ought to observe similar treatment effects when comparing *Ledger* and *Individuals* and no treatment differences between *Ledger* and *Pairs*. In order to be consistent with the *Pairs*, we neither require subjects in the *Ledger* to use the *Ledger* nor do we allow them access to *Ledger* entries from previous periods.

The consumption smoothing motive in our setting comes from the concavity of the induced quadratic utility function. Subjects spent, saved, and borrowed per-period income, allotted as experimental credits (ECs), that followed two different pre-drawn stochastic income processes. Subjects began each lifecycle with no initial wealth. Importantly, subjects received consumption points in each period equivalent to the consumption utility resulting from their consumption decision in that period. Using pre-drawn income processes allowed us to hold the income process constant across treatments for each decision lifecycle.

[8]Pairs in the *Pairs* treatment were stable within and across the two lifecycles.

Sessions began with a 6-question, individual-level Cognitive Reflection Test (CRT) introduced by Frederick (2005), also adopting questions from the Cognitive Reflection Test-Long (CRT-L) developed by Primi et al. (2016).[9] Subjects had 90 seconds to answer each CRT question and earned $.25 for each correct answer. We followed this with an individual-level Eckel-Grossman test of risk preferences (Eckel and Grossman, 2002).[10] Subjects were not time-constrained when solving the lifecycle problem in either treatment, but were given a soft reminder at 75-seconds to make a decision. We did this to help move the experiment along at a reasonable pace. All subjects were students recruited at the University of Arkansas.[11] We ended each session with a demographic survey that also included a survey of attitudes toward debt and saving.

Instructions provided subjects with detailed information about the utility function, income processes, lifecycle duration, and borrowing and saving. Importantly, subjects had sufficient information to fully solve the lifecycle problem and achieve the unconditionally-optimal consumption path. Further, we provided subjects with information about their per-period income, and their current bank account balance to help them keep track of their borrowing/savings. We also provided subjects with a consumption smoothing tool to reduce the cognitive complexity of the problem. To use the tool, subjects could propose a hypothetical level of consumption and learn the corresponding levels of utility (we called these consumption points in the game), savings or debt, and the marginal utility of consumption (we called this the 'marginal increase' in the game). Subjects could use this tool as many or as few times as desired. We provide an example of the decision screen for an individual in Figure 12 and for pairs in Figure 13 in Section 7.1 of the Appendix.

	Individuals & Ledger	**Pairs**
Instructions & Comprehension Quiz	Individual	Individual
Cognitive Reflection Test	Individual	Individual
Eckel-Grossman Risk Assessment	Individual	Individual
Two rounds of decison-making	Individual	Joint
Eckel-Grossman Risk Assessment	Individual	Joint
Demographics & Survey of Decisions	Individual	Individual

Table 1: This table describes the order of events when conducting a session and indicates whether the task was completed individually (Individual) or in a pair (Joint).

For *Individuals*, we converted consumption points to U.S. dollars at 50 points per $1. For *Pairs*, we converted consumption points at 25 points per $1.[12] This conversion scheme holds subject-level incentives constant across treatments. Subjects also received a $10 show-up fee. We conducted all sessions at the University of Arkansas's Behavioral Business Research Laboratory. We conducted the first wave of experimental session sessions between October 2019 and March 2020 and the second in September and October of 2022. On average, sessions

[9]We include the full set of questions in Section 7.7.

[10]We conduct this assessment twice in each of our treatments - once before and once after gameplay. We do this to understand if risk evolves differently for pairs and individuals. It does not. Further, our results are robust to using either measure as a control for risk attitudes.

[11]IRB protocol #: 1908210566

[12]We rounded payoffs to the next highest point. For example, we treated a score of 51.4 points as a score of 52 points, which would earn an individual $1.04 rather than $1.028.

lasted approximately 1.5 hours. We have 38 observations in *Individuals*, 38 observations in *Pairs*, and 30 in *Ledger* for a total of 106 unique observations comprising 144 participants.[13] We implemented our experiment using zTree (Fischbacher, 2007).

5 Results

We show treatment-level mean absolute unconditional and conditional consumption errors by period and treatment for *Pairs* and *Individuals* treatments in panels (a) and (c) of Figure 2, respectively.[14] Additionally, we present the corresponding difference in treatment-level mean absolute unconditional and conditional consumption errors in panels (b) and (d) of the same figure. For panels (b) and (d), observations above the x-axis denote an instance where *Pairs* outperformed *Individuals*. Visually, it appears that participants in *Pairs* outperform those in *Individuals* in solving the finite life-cycle problem along both consumption paths (we also show this using medians rather than averages in Figure 9 and also show comparisons with the *Ledger* treatment in Figure 10 and Figure 11, which are all located in Section 7.1).

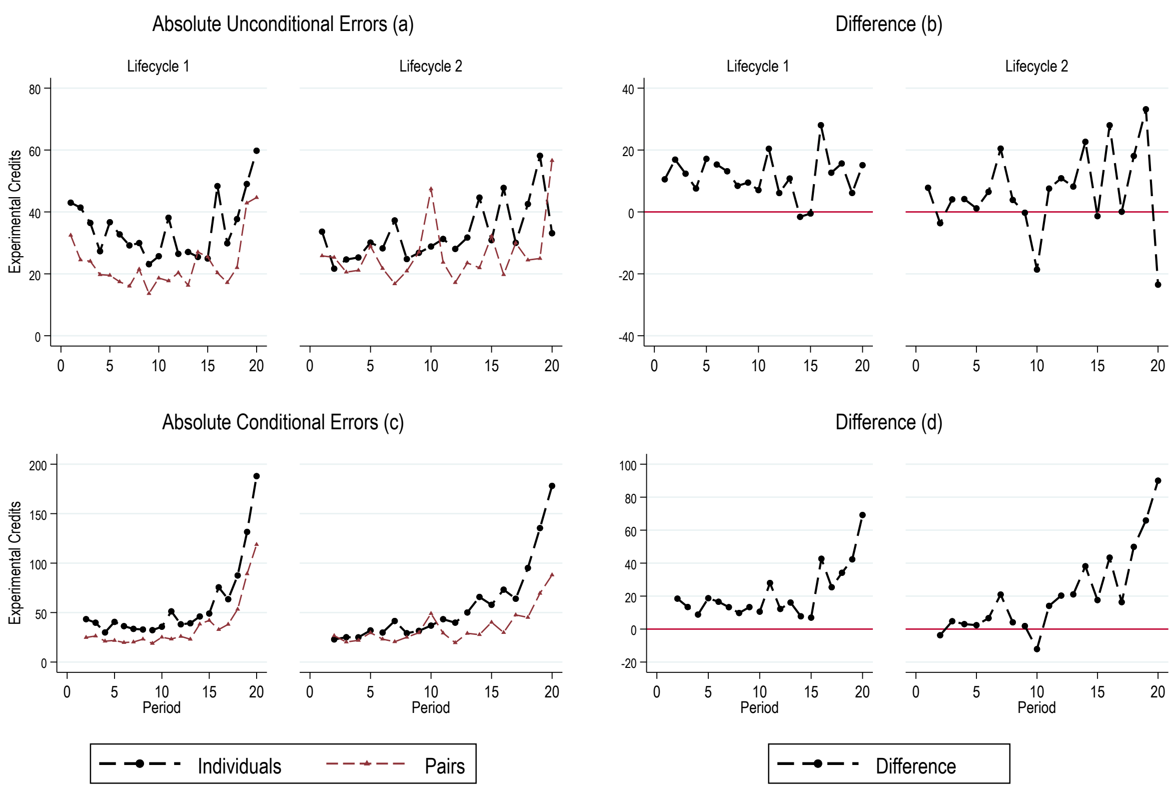


Figure 2: This figure depicts treatment-level average absolute consumption errors (panels (a) and (c)) and differences between the absolute consumption errors by treatment (panels (b) and (d)). For panels (b) and (d), values above zero indicate that *Pairs* outperformed individuals in that period.

[13] We note that our sample sizes, particularly in *Ledger* where $N = 30$, are a bit smaller than the general rule-of-thumb of $N = 40$.

[14] We show average consumption and per-period consumption heterogeneity in Figure 14 located in Section 7.2 of the Appendix. We also show granular consumption data in Section 7.1.

Also worth noting in panel (c) of Figure 2 is the gradual buildup of absolute conditional errors toward the end of each lifecycle. This is likely due to the adoption of simple consumption heuristics primarily focused on income (we discuss this in more detail in Section 5.2). Because the conditionally-optimal path assumes subjects will account for previous mistakes in remaining decisions, these heuristics are increasingly penalized when moving along the conditionally optimal consumption path.

To assess whether the apparent treatment-level performance difference in Figure 2 is statistically significant, we first conduct a series of ordinary least squares (OLS) regressions where we project observation-level measures of lifecycle performance onto a set of indicator variables denoting treatment. We estimate heteroskedasticity-robust standard errors to account for the possibility of unequal error variance across treatments. We do this by restricting data to only the first lifecycle, to only the second lifecycle, and again combining data from both lifecycles.[15]

	Observation-Level Regression Results								
	Individuals as Baseline					***Pairs* as Baseline**			
	RMSD-U *(1)*	**RMSD-C** *(2)*	**MAE-U** *(3)*	**MAE-C** *(4)*		**RMSD-U** *(5)*	**RMSD-C** *(6)*	**MAE-U** *(7)*	**MAE-C** *(8)*
	Lifecycle 1					*Lifecycle 1*			
Pairs	−19.030**	−29.430**	−11.540**	−21.450**	*Ind.*	19.030**	29.430**	11.540**	21.450**
	(9.716)	(15.34)	(4.931)	(10.02)		(9.716)	(15.34)	(4.931)	(10.02)
Ledger	−8.514	−12.020	−2.853	−8.133	*Ledger*	10.510**	17.410*	8.687***	13.320**
	(9.692)	(18.01)	(5.194)	(11.67)		(5.039)	(11.97)	(3.500)	(7.660)
Constant	52.550***	81.590***	34.620***	57.600***		33.520***	52.160***	23.080***	36.140***
	(9.027)	(14.43)	(4.418)	(9.431)		(3.596)	(5.212)	(2.189)	(3.380)
	Lifecycle 2					*Lifecycle 2*			
Pairs	−9.117	−30.830**	−6.462	−21.290**	*Ind.*	9.117	30.830**	6.462	21.290**
	(9.896)	(14.63)	(6.082)	(9.856)		(9.896)	(14.63)	(6.082)	(9.856)
Ledger	3.174	−11.870	0.694	−8.176	*Ledger*	12.290*	18.960*	7.156*	13.120*
	(11.53)	(17.75)	(6.594)	(12.03)		(9.614)	(12.57)	(5.700)	(8.715)
Constant	47.500***	80.090***	32.970***	56.670***		38.390***	49.270***	26.510***	35.380***
	(8.319)	(13.62)	(4.898)	(9.105)		(5.361)	(5.347)	(3.606)	(3.775)
	Pooled					*Pooled*			
Pairs	−14.630*	−29.960**	−9.001**	−21.370**	*Ind.*	14.630*	29.960**	9.001**	21.370**
	(9.425)	(14.48)	(5.245)	(9.330)		(9.425)	(14.48)	(5.245)	(9.330)
Ledger	−2.574	−11.260	−1.080	−8.150	*Ledger*	12.060**	18.700**	7.921**	13.220**
	(10.12)	(17.25)	(5.553)	(11.12)		(7.102)	(11.47)	(4.246)	(7.393)
Constant	51.880***	82.880***	33.790***	57.130***		37.250***	52.920***	24.790***	35.760***
	(8.391)	(13.71)	(4.490)	(8.830)		(4.293)	(4.664)	(2.712)	(3.012)
N	106	106	106	106		106	106	106	106

Table 2: This table presents results from a series of OLS regression wherein we project observation-level measures of consumption errors onto a set of indicator variables denoting treatment. We refer to the root-mean-squared deviation of consumption as RMSD, the mean absolute consumption error as MAE, the unconditional error path as U, and the conditional error path as C. Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We collapse our data in two ways. First, we calculate the mean absolute consumption

[15] Collapsing data to the observation level provides a strict approach to dealing with the fact that consumption decisions are likely serially correlated.

error (MAE) for each observation i as $MAE_i^X = \frac{\sum_{t=1}^{t=T} |c_{i,t} - c^*_{X,t}|}{T}$. Second, we calculate the root-mean-squared deviation (RMSD) of consumption errors for each observation i as $RMSD_i^X = \sqrt{\frac{\sum_{t=t}^{t=T} (c_{i,t} - c^*_{X,t})^2}{T}}$. Here, $X \in \{U,\ C\}$ corresponds to the unconditionally-optimal and conditionally-optimal consumption paths, respectively. Note that for $X = U$, we use c_t^*, as given by Equation (6). For $X = C$, we instead use $\hat{c^*_{i,t}}$, as given by Equation (7). We include both measures because of the difference in how they penalize large period-level errors. While both measures capture the magnitude of consumption errors for a given observation, the RMSD's quadratic formulation (opposed to the MAE, which is linear) more harshly punishes larger period-level errors than the MAE. Including both measures allows for some agnosticism about how best to treat large absolute consumption errors (that result from save-and-binge strategies, for example), which appear more frequently in our *Individual* observations.

We report results from these regressions in Table 2 where we test the null hypothesis that performance in *Individuals* (and also in *Ledger*) is at least as good as in *Pairs*. The left half (right half) of Table 2 treats *Individuals* (*Pairs*) as our baseline treatment. Results using data from only the first lifecycle are in the top panel, from the second lifecycle in the middle panel, and combining data from both lifecycles in the bottom panel. We refer to this notion of combined data as *Pooled* in the rest of our analysis.

First, we note that regardless of how we collapse our data or which subset of decisions we consider, we observe no significant treatment-level differences between *Individuals* and *Ledger*. By contrast, we observe significant differences in the performance of *Pairs* and *Individuals* along both consumption paths and using both measures in the first lifecycle, and along the conditional consumption path using both measures in the second lifecycle. Though not statistically significant, we also see that our coefficient of interest is qualitatively consistent in lifecycle 2 along the unconditional path for both performance measures. When combining data from both lifecycles, performance in *Pairs* is significantly better than in *Individuals* along both consumption paths for both performance measures.

When treating *Pairs* as our baseline treatment, we find that for both measures and along both consumption paths, participants in both *Individuals* (denoted *Ind.*) and *Ledger* perform significantly worse than their counterparts in *Pairs* in the first lifecycle and when combining data from both life cycles. The same result holds along the conditionally-optimal path in lifecycle 2 but not along the unconditionally-optimal path.

Overall, our results show that participants in *Pairs* outperform participants in *Individuals*. Further, the fact that our coefficient estimates for *Ledger* are insignificant demonstrates that participants in *Pairs* outperform subjects in *Individuals* because they are participating in joint decision-making and not just because they are more carefully considering their decisions or articulating the logic of their decisions. That said, it is worth noting that our results indicate, at least qualitatively, that allowing subjects to articulate their thoughts in *Ledger* improves decision-making relative to *Individuals*.

We also conduct a series of Mann-Whitney U (MWU) and Kolmogorov-Smirnov (KS) tests on both the RMSD and the MAE. The two tests are similar in that they are non-parametric tests that make no assumptions about underlying distributional parameters. However, the

	Summary of Non-Parametric Tests					
	Unconditional Absolute Error			Conditional Absolute Error		
	Lifecycle 1 *(1)*	**Lifecycle 2** *(2)*	**Pooled** *(3)*	**Lifecycle 1** *(4)*	**Lifecycle 2** *(5)*	**Pooled** *(6)*
	Excluding Ledger Observations ($N_I = 38, N_P = 38$)					
Mean MAE - I	34.62	32.97	33.79	57.60	56.67	57.13
Mean MAE - P	23.08	26.51	24.80	36.14	35.38	35.76
MWU	0.05	0.48	0.21	0.11	0.11	0.16
KS	0.07	0.39	0.39	0.12	0.12	0.19
Mean RMSD - I	52.55	47.50	51.88	83.71	82.18	85.03
Mean RMSD - P	33.52	38.39	37.25	53.52	50.55	54.29
MWU	0.03	0.55	0.26	0.19	0.16	0.25
KS	0.04	0.39	0.19	0.12	0.12	0.07
	Including Ledger Observations ($N_I = 68, N_P = 38$)					
Mean MAE - I	33.36	33.28	33.31	54.01	53.06	53.54
MWU	0.01	0.18	0.05	0.03	0.08	0.07
KS	0.01	0.12	0.10	0.02	0.10	0.05
Mean RMSD - I	48.79	48.90	50.74	78.27	76.80	79.93
MWU	0.01	0.21	0.06	0.06	0.10	0.10
KS	0.01	0.12	0.07	0.02	0.06	0.01

Table 3: This table reports p-values from a series of Mann-Whitney U (MWU) tests and Kolmogorov–Smirnov (KS) tests of equality across our *Individuals* (I) and *Pairs* (P) treatments of the root-mean-squared deviation (RMSD) of consumption and mean absolute consumption errors (MAE). Columns labeled Lifecycle 1 report results using data from only the first lifecycle, columns labeled Livecycle 2 reports results using data from only the second lifecycle, and columns labeled Pooled report results using data from both lifecycles.

MWU test primarily assesses whether our treatment-wise distributions of observation-level performance measures differ in their central tendencies whereas the KS test assesses whether the distributions underlying our two samples are themselves different.

We again do this using data from each lifecycle independently and then combining data from both lifecycles. Additionally, based on results in Table 2 that demonstrate that there are no statistical differences between the *Individuals* and *Ledger* treatments, we also report results from our non-parametric tests where we treat *Ledger* observations as *Individuals* observations. We report these results in Table 3. Note that this leads to unbalanced sample sizes ($N_{Ind.} = 68, N_{Pairs} = 38$). This is a non-issue for these statistical tests (or for t-tests) since they make no assumptions about relative sample sizes.[16]

We first note (rows 1,2 and 5,6) that the mean differences in the MAE and RMSD across *Pairs* and *Individuals* align with Figure 2. For both observational-level measures, along both consumption paths, and for each method of partitioning, we see that pairs, on average, seem to outperform individuals. Overall, results are qualitatively consistent with the average

[16] Unbalanced group sizes can yield a higher type II error rate than using balanced groups of an equivalent total sample size since both compute statistics that depend on the product of the two group sizes.

treatment effects we estimate using a parametric approach in Table 2. However, results are not always significant, which is perhaps due to a lack of statistical power.

Focusing on lifecycle 1, when we compare just the *Individual* data to the *Pairs* data (top half of Table 3) we see that the MWU and KS tests are significant for both the MAE and RMSD for absolute unconditional errors. When looking at absolute conditional errors, the MWU test is significant for only the MAE, while the KS tests are significant for both the MAE and RMSD. When evaluating performance difference in lifecycle 2, we note that neither test is significant when testing the difference in the absolute unconditional errors. When testing the difference in absolute conditional errors, the KS test is only marginally insignificant for both the MWU and the MAE tests ($p \approx 0.11$ and $p \approx 0.12$, respectively). Finally, with pooled data, we note that the only significant result is the KS test on the difference in absolute conditional errors.

We repeat our non-parametric tests including observations from *Ledger*. First, note in the bottom half of Table 3 that doing this leads to very little change in the average values of the MAE or RMSD along either consumption path or for either way we partition our data. If anything, incorporating *Ledger* observations slightly attenuates errors, which should make it harder for us to detect treatment differences. Thus, changes in significance that result from incorporating *Ledger* observations result from an increase in statistical power and not because performance in *Ledger* is worse than in *Individuals* (a conclusion corroborated by Table 2). Doing this, we now see that performance in *Pairs* is significantly better than in *Individuals* using both performance measures and along both consumption paths in all instances in all but lifecycle 2 along the unconditionally optimal path.

Overall, the balance of evidence from the observation-level analysis indicates that joint decision-making leads to significantly better decisions in the finite lifecycle problem faced by our participants. This is especially true when we consider the conditionally-optimal consumption path where we allow for the correction of past mistakes. Further, our results show that the superior performance observed in *Pairs* cannot be explained by the fact that pairs have the opportunity to articulate the logic of their consumption decisions. Instead, the mechanism underlying our treatment differences seems to hinge critically on the interaction of subjects within a pair while reaching a joint decision.

We next exploit decision-level data in order to better understand what leads to consumption errors by estimating a series of multi-level mixed-effects linear regressions. This approach allows us to control for time-varying structural features of our economic environment (i.e. per-period income) and features that result directly from the decisions of our participants (i.e. wealth accumulation). We estimate

$$Y_{i,t} = \alpha + \sum_j \gamma_j Group_j + \beta_{i,t} X_{i,t} + \zeta_i Z_i + \epsilon_{i,t} \tag{8}$$

where our outcomes of interest, $Y_{i,t}$ are either absolute unconditional consumption errors or absolute conditional consumption errors. Here, $X_{i,t}$ is a vector holding time-varying characteristics and Z_i a vector holding time-invariant characteristics of participants. Finally, $Group_j,\ j \in \{Individual, Pairs, Ledger\}$, denotes a set of indicator variables denoting treatment. We estimate Equation (8) while restricting our data by treatment and also for our full data sample.[17] We report the results of these estimation exercises in Table 4.[18]

Columns 2-6 of this table report results using unconditional absolute consumption errors while columns 7-11 report results using conditional absolute consumption errors. Columns labeled with treatment names use data from only the corresponding treatment while columns labeled 'Pooled' use data from all treatments combined.

Row labels correspond to right-hand-side variables. Our coefficients of interest are *Pairs* and *Ledger*, which correspond to indicator variables denoting treatment. If the results here are consistent with our observation-level results, we should expect significant negative coefficients for *Pairs* and insignificant coefficients for *Ledger*. Wealth refers to subjects' accumulated savings (or debt) in ECs while Income represents current-period income in ECs. We include these controls because both can be important determinants of consumption.This is shown experimentally by Carbone and Hey (2004) and empirically by Flavin (1981), Hayashi (1982), and Zeldes (1989). Lifecycle2 is an indicator variable for the second lifecycle, which we include to capture the increased complexity of this lifecycle that results from a need to borrow extensively to achieve optimality. GPA and CRT correspond to a subject's grade point average CRT score, which we take as proxies for intelligence or cognitive ability. Note that for *Pairs* observations, each measure represents the within-pair average of the two individuals. Though these two measures are correlated, the correlation coefficient is only 0.24. This suggests that these two measures are capturing different components of intellectual capacity. Additionally, we control for gender at the observation level via Male. Here 1.Male refers to a male participant in the *Individuals* treatment and a mixed-gender pair in the *Pairs* treatment. The variable 2.Male represents an all-male pair in the *Pairs* treatment. Note we do control for gender in columns 6 or 11 because categorical definitions are different across treatments and so would yield nonsensical interpretations. Finally, we include interaction terms, Lifecycle2**Treatment*, which indicate differences in performance across treatments induced by lifecycle 2. We include this variable to capture whether pairs and individuals are differently impacted by lifecycle 2, where optimal behavior requires extensive borrowing.

We first note that results here are consistent with our observation-level results in that performance in *Pairs* is significantly better than in *Individuals* and that performance *ledger* is statistically indistinguishable from *Individuals*. Additionally, we again see that this effect is more pronounced along the conditional path where we allow participants to learn and account for past errors.

Two things stand out as potential explanations for these patterns. First, we see that consumption errors react differently to wealth balances along each consumption path. Accumulating wealth attenuates errors in *Individuals* along the unconditional consumption path but exacerbates them along the conditional consumption path. The effect is similar in *Pairs*, albeit insignificant along the unconditional path and about 50% smaller along the conditional path than in *Individuals*. Second, we corroborate a common finding in the LTO literature by documenting that consumption overreacts to income along both consumption paths and in

[17] Though random effects models are common in this literature (examples are Carbone and Duffy (2014), Ballinger et al. (2011)), a Hausman test indicates the need to control for potential fixed effects, which perhaps result from static session effects (Fréchette, 2012).

[18] We include an alternative specification in Section 7.1 where we instead group data by lifecycles (Table 15).

	Regression Results - Mixed Effects Estimations									
	Unconditional Absolute Error					Conditional Absolute Error				
	Individuals (1)	***Pairs (2)***	***Ledger (3)***	**Pooled *(4)***	**Pooled *(5)***	***Individuals (6)***	***Pairs (7)***	***Ledger (8)***	**Pooled *(9)***	**Pooled *(10)***
Pairs				-9.001^{*}	-13.03^{**}				-21.37^{**}	-15.39^{***}
				(5.195)	(6.563)				(9.241)	(5.442)
Ledger				-1.080	-1.365				-8.150	2.326
				(5.500)	(6.319)				(11.01)	(5.974)
Wealth	-0.0486^{**}	-0.0188	-0.0413		-0.0361^{**}	0.251^{***}	0.0923^{***}	0.135^{***}		0.196^{***}
	(0.0230)	(0.0214)	(0.0287)		(0.0151)	(0.0621)	(0.0353)	(0.0519)		(0.0512)
Income	0.673^{**}	0.345^{**}	0.444		0.465^{***}	2.684^{***}	1.457^{***}	1.769^{***}		1.966^{***}
	(0.326)	(0.148)	(0.364)		(0.148)	(0.590)	(0.165)	(0.424)		(0.239)
Lifecycle2	-4.607	3.133	0.748		-4.431	8.369^{**}	4.960^{**}	2.806		5.415
	(4.241)	(2.789)	(5.867)		(3.869)	(3.841)	(2.232)	(6.038)		(4.029)
CRT	-0.944	-6.295^{***}	-5.457		-3.047^{*}	-0.467	-5.463^{***}	-5.034		-1.466
	(3.474)	(2.171)	(3.424)		(1.566)	(1.850)	(2.059)	(3.128)		(1.366)
GPA	-25.89^{*}	-10.02	-12.98		-13.89^{*}	-10.02	-7.132	-6.003		-7.145
	(15.18)	(6.737)	(10.16)		(7.278)	(12.22)	(4.730)	(10.12)		(5.997)
1.Male	39.30^{***}	-10.84^{*}	-1.404			19.89	-14.78^{***}	1.104		
	(10.25)	(6.150)	(10.45)			(13.18)	(3.833)	(11.13)		
2.Male		-2.358					-5.213			
		(6.280)					(5.227)			
Lifecycle2**Pairs*					7.059					4.312
					(4.718)					(4.294)
Lifecycle2**Ledger*					5.219					-2.208
					(6.722)					(5.974)
α	88.34	61.85^{**}	58.94	33.79^{***}	71.71^{**}	-128.4^{*}	-20.94	-47.39	57.13^{***}	-64.70^{*}
	(55.00)	(28.02)	(38.85)	(4.447)	(29.55)	(71.57)	(22.10)	(39.79)	(8.746)	(34.11)
N	1080	1520	800	4240	3400	1026	1444	760	4028	3230
Clusters	27	38	20	106	85	27	38	20	106	85

Table 4: This table contains the results of a series of mixed effects regressions. Column 1 lists variable names, columns 2 through 6 report results using absolute unconditional consumption errors as the dependent variable, and columns 7 through 11 report results using absolute conditional consumption errors as the dependent variable. Columns labeled using treatment names use data from their corresponding treatment only. Columns labeled 'Pooled' use all data. Robust standard errors clustered at the observation level in parentheses. Note differences in clusters and observations across columns result from some participants choosing to not provide demographic data.

both treatments. However, we again see that participants in *Individuals* are reacting about twice as strongly to income as are participants in *Pairs*.

We also note that both measures of cognitive ability are positively correlated with performance in our lifecycle problem. While CRT is not indicative of a subject's ability to use backward induction, a necessary skill for this setting, it is indicative of their ability to resist knee-jerk responses and more carefully consider problems. Interestingly, we see this coefficient is always significant in *Pairs*, which aligns with the general finding in Charness and Sutter (2012) that group decisions typically reflect more cognitive sophistication. GPA is only a significant predictor for *Individuals* along the unconditionally-optimal path, but we note that the sign is consistent for all estimates and suggests a higher GPA leads to smaller consumption errors. Overall, these results are consistent with Ballinger et al. (2011), which shows that cognitive ability can be a strong predictor of performance in experiments focused on savings behavior.

Recall that 1.Male signifies a male participant in the *Individuals* treatment and a mixed-gender pair in the *Pairs* treatment. We see that males perform significantly worse in *Individuals* along the unconditionally-optimal path. This is qualitatively true along the conditionally-optimal path, albeit not statistically significant. Interestingly, we see that mix-gendered pairs outperform all-male pairs along both consumption paths.

We repeat these mixed effect regressions reclassifying *Ledger* observations as *Individual* observations, as we did with our non-parametric tests. We report these results in Table 5. Results are qualitatively identical and quantitatively very similar.

	Regression Table: Mixed Effects Estimations Including *Ledger* as *Individuals*							
	Unconditional Absolute Error				**Conditional Absolute Error**			
	Individuals ***(1)***	***Pairs*** ***(2)***	**Pooled** ***(3)***	**Pooled** ***(4)***	***Individuals*** ***(5)***	***Pairs*** ***(6)***	**Pooled** ***(7)***	**Pooled** ***(8)***
Pairs			-8.525^{**}	-12.42^{**}			-17.78^{***}	-16.37^{***}
			(3.929)	(4.846)			(6.460)	(3.966)
Wealth	-0.0426^{**}	-0.0188		-0.0358^{**}	0.240^{***}	0.0923^{***}		0.195^{***}
	(0.0184)	(0.0214)		(0.0149)	(0.0446)	(0.0353)		(0.0509)
Income	0.575^{**}	0.345^{**}		0.465^{***}	2.293^{***}	1.457^{***}		1.966^{***}
	(0.241)	(0.148)		(0.149)	(0.384)	(0.165)		(0.239)
Lifecycle2	-2.216	3.133		-2.202	5.872^{**}	4.960^{**}		4.468
	(3.406)	(2.789)		(3.256)	(2.667)	(2.232)		(3.082)
CRT	-3.535	-6.295^{***}		-3.034^{*}	-1.152	-5.463^{***}		-1.464
	(2.361)	(2.171)		(1.567)	(1.804)	(2.059)		(1.375)
GPA	-14.46	-10.02		-13.66^{*}	-3.633	-7.132		-6.974
	(11.48)	(6.737)		(7.756)	(8.114)	(4.730)		(6.056)
1.Male	13.99^{*}	-10.84^{*}			7.450	-14.78^{***}		
	(7.780)	(6.150)			(6.255)	(3.833)		
2.Male		-2.358				-5.213		
		(6.280)				(5.227)		
Lifecycle2**Pairs*				4.844				5.244
				(4.255)				(3.771)
α	61.08	61.85^{**}	33.32^{***}	70.26^{**}	-106.2^{**}	-20.94	53.54^{***}	-64.22^{*}
	(41.38)	(28.02)	(2.867)	(29.62)	(44.05)	(22.10)	(5.730)	(33.70)
N	1880	1520	4240	3400	1786	1444	4028	3230
Clusters	47	38	106	85	47	38	106	85

Table 5: This table shows of mixed effects regressions. Column 1 lists variable names, where maxCRT (min) refers to the highest (lowest) CRT score in the pair. For *Individuals*, maxCRT refer to the individual subject's CRT score. Columns 2 thru 5 report results using the absolute unconditional consumption error as the dependent variable and columns 6 thru 9 report results using the absolute conditional consumption error as the dependent variable. Columns labeled as 'Individuals' or 'Pairs' use only the data from their corresponding treatment. Columns labeled 'Pooled' use all data. We report robust standard errors in parentheses. Note that the difference in N arises because there is no conditional error in the first period of either decision lifecycle.

Since subjects in our experiment are concerned with earnings maximization, it perhaps makes the most sense to consider average earnings differences between subjects in our *Pairs* and *Individuals* treatments. Subjects in the *Individuals* treatment earned \$21.66 on average, while subjects in the *Pairs* treatment earned an average of \$24.33. Because we are concerned with earnings differences that result from differences in decisions, we subtract from these averages the fixed show-up fee of 10. We see that subjects in the *Pairs* treatment earned approximately $\frac{\$14.33-\$11.66}{\$11.66} = 22.90\%$ more, on average, than subjects in the *Individuals* treatment. Without making this adjustment, earnings differences are still significantly different: *Pairs* earn approximately 12.33% more than *Individuals.* The result of an MWU test indicates this treatment-level earnings difference is statistically significant at the 1% level.

We also quantify differences between the *Pairs* and *Individuals* treatments by comparing the performance of *Pairs* to synthetic pairs formed using subjects in our *Individuals* treatment. Our interest is in how much we must improve the performance of these synthetic pairs before their decisions are no longer statistically distinguishable from real *Pairs* at a 10% level of significance where the average error of synthetic pairs remains larger than that of real pairs. To do this, we randomly match *Individuals* into synthetic pairs and assume each pair consumed in a given period the average of what the two individuals consumed in that period. We repeat this matching process for all possible pairings and average results for all observations.

We find that we can reduce the conditional consumption error of synthetic pairs by approximately 33%, on average, before the performance of real and synthetic pairs becomes indistinguishable. In level terms, this reduces the average conditional consumption error of synthetic pairs from 63.91 to approximately 42.82 experimental credits.

5.1 Textual Analysis

Because subjects in the *Pairs* treatment of our experiment engaged in unrestricted chat to make joint decisions, we are able to use textual analysis to gain deeper insight into how subjects frame the dynamic optimization problem and develop heuristics.

Following Cooper and Kagel (2005), we establish a set of categories we use to classify the language used by subjects in our *Pairs* treatment, which we describe in Table 6. These categories are neither exhaustive nor mutually exclusive. Rather, the categories are complementary, which allows for some nuance in classification despite the binary coding system. We trained two research assistants (RAs) who then worked independently to classify language into our pre-selected categories. As an example, if a pair discussed how to allocate resources in terms of spending but never in terms of savings, the research assistants would likely code 'Discuss Savings' as a zero and 'Discuss Spending' as a one.

We use these codings from our RAs to construct a measure that captures, on average, how often chat aligns with a given category. We construct this measure by first summing over all periods, sessions, and pairs for both research assistants and then dividing this sum by two times the total number of periods times the total number of pairs. Thus, we report a number bounded between zero and one where a value of one means all pairs used language compatible with that category in all periods. Anything less than one means that there is at least one pair who does not use that language in at least one period. To measure classification agreement, we divide the number of times the RA's disagree by the number of opportunities to code a discussion category, subtract this from one, and then convert to percentage terms. We report both measures in Table 6.

Notice in Table 6 the relatively high frequency of the "Discuss Spending" category (87% of interactions in *Pairs*), which indicates that *Pairs* mostly frame discussions around spending. On the one hand, this is not surprising since subjects in our experiment earn money via consumption. On the other, the stochastic income process, coupled with the consumption

Category	Description	Mean	Agreement(%)
Discuss Saving	frames discussion in terms of saving	.090	98.09
Discuss Spending	frames discussion in terms of spending	.863	98.16
Save More	Someone proposes saving more relative to previous suggestion/period	.034	99.41
Save Less	Someone suggests saving less relative to previous suggestion/period	.008	99.34
Spend More	Someone proposes spending more relative to previous suggestion/period	.189	96.58
Spend Less	Someone proposes spending less relative to previous suggestion/period	.149	99.67
Nominal Target	Pair discusses a nominal target (i.e. consumption points)	.084	94.01
Real Target	Pair discusses a real target (i.e. total dollar earnings)	.014	99.61
Marginal Target	Pair targets a 'marginal increase' target	.060	99.61
Savings Target	Pair tries to maintain a certain amount of savings	.009	99.21
Period Earnings Target	Pair discusses a per-period earnings target	.011	98.10
Total Earnings Target	Pair discusses a lifetime earnings target	.012	99.14
Proportional Spender	Pair discusses spending a proportion of income or total wealth	.050	98.88
Borrow	Pair discusses borrowing against future income	.046	99.63
Constant Spending	Pair discusses spending a constant amount	.036	99.87
Save & Binge	Pair discusses saving heavily to spend a large lump sum later	.044	99.87

Table 6: This table provides information regarding our textual analysis. The first two columns define the categories used by two research assistants (RAs) who worked independently to classify the language used by pairs when forming joint decisions. The third column provides a measure of how frequently subjects in *Pairs* used language consistent with each category. The fourth column provides a measure of the level of classification agreement between our two RAs. We construct values in column three by summing over all periods, sessions, and pairs for both RAs, and dividing this sum by two times the total number of periods times the total number of pairs. We construct our agreement measure by dividing the number of times the RAs disagree about a given classification by the number of opportunities to code a discussion category, subtracting this from one, and then converting to percentage terms.

smoothing motive, makes saving and borrowing important components of earnings maximization in our experiment (e.g., Figure 6 demonstrates that optimal consumption requires frequent borrowing and saving in both lifecycles). We also see that subjects, explicitly or implicitly, often discuss spending strategies that fix consumption proportionally, which aligns with our results in Section 5.2, where we show that participants in both *Pairs* and *Individuals* use spending strategies with constant propensities to consume (see Figure 8, for example).

These sorts of simple heuristics greatly reduce the cognitive complexity of the optimization task but might fail subjects whenever saving or borrowing is necessary for optimization. For example, a pair that spends a fixed proportion of the per-period endowment would not borrow whenever necessary to spend at the unconditionally- or conditionally-optimal level. This aligns with Carbone and Hey (2001) and Hey and Knoll (2011), which both find that people solving dynamic optimization problems are likely to develop simple decision criteria that reduce the cognitive complexity of the choice task. These proportional consumption heuristics might explain the upward trend in conditionally-optimal consumption errors that we do not see in the unconditionally-optimal consumption errors. This is because a heuristic that leads to an absolute error in one period will, on average, lead to a similar absolute error in a later period. The invariant nature of the heuristic could prevent subjects from avoiding current-period errors and adjusting for past errors.

The tendency of participants in *Pairs* to develop simple heuristics leads to considerable under-borrowing in our experiment. However, we do not see in our chat data that participants in *Pairs* often express disdain for borrowing. Thus, under borrowing may result from subjects developing simple heuristics (i.e. proportional spending rules) that overlook borrowing and not from the fact that they are actively averse to debt. To better understand this, we evaluate responses to a question included in our post-experiment survey of decisions that asked subjects to provide their subjective outlook on debt using an open-ended question that asked subjects "Do you believe it is good or bad to have debt?" .[19] Of the 134 responses we collected, 69 subjects reported viewing debt as explicitly bad (*Bad*) while the remaining 65 subjects argued that the goodness or badness of debt depends on the benefit, purpose, and magnitude of debt (*Nuanced*).[20]

To understand how a participant's subjective outlook on debt impacts performance in our dynamic optimization task relative to other individual characteristics, we estimate a set of mixed-effects models wherein we project consumption errors along both paths onto an indicator variable denoting whether a subject had a nuanced outlook on debt along with controls all other available demographics. Results of these estimations are in Table 7. Note that we exclude participants from *Pairs* from these estimations because including them would require that we take a stance on how to categorize each of the six possible within-pair combinations of debt opinions. Additionally, we are missing data on subjective debt outlook for 9 of our 38 participants from *Individuals* due to a network failure and some subjects in both treatments chose not to provide demographic information.

	Regression Results - Subjective Debt Outlook					
	Unconditional Absolute Error			**Conditional Absolute Error**		
	Lifecycle 1* *(1)	***Lifecycle 2* *(2)***	**Pooled *(3)***	***Lifecycle 1* *(4)***	***Lifecycle 2* *(5)***	**Pooled *(6)***
Nuanced	-9.684^{+}	-5.822	-7.753	-25.55^{**}	-15.54	-20.54^{*}
	(6.138)	(7.736)	(6.584)	(12.27)	(11.96)	(11.34)
GPA	-13.15	-7.848	-10.50	-35.22^{**}	-24.50	-29.86^{*}
	(9.947)	(9.871)	(9.573)	(17.67)	(15.57)	(15.54)
CRT	-1.825	-2.643	-2.234	-5.029	-5.464	-5.246
	(1.902)	(2.341)	(2.031)	(3.323)	(3.503)	(3.208)
Male	2.884	10.75	6.816	20.53	31.26^{**}	25.90^{*}
	(6.229)	(7.276)	(6.215)	(14.46)	(14.98)	(13.64)
Constant	100.4^{**}	67.44^{*}	83.91^{**}	223.9^{***}	149.9^{**}	186.9^{***}
	(39.87)	(36.11)	(36.33)	(73.52)	(59.78)	(60.39)
N	920	920	1840	874	874	1748
Clusters	46	46	46	46	46	46

Table 7: This table provides suggestive evidence of the impact of subjective debt outlook on performance in our optimization task. Regressions include data from all participants in our *Individuals* and *Ledger* treatment for whom we have demographic data. Coefficient estimates proceed using mixed-effects regressions. Robust standard errors clustered at the observation level in parentheses. $^{+}$ <.12 * $p < .10$, ** $p < .05$, *** $p < .01$

We find suggestive evidence that having a *Nuanced* debt outlook improves performance in our dynamic optimization task along the conditionally-optimal consumption path. Interestingly, this is true even though we are controlling for cognitive ability, which suggests that this result

[19] We include the full survey in Section 7.7.

[20] We consider any response that indicated debt is always good or could be good under any circumstance as *Nuanced*.

is not driven simply by a positive correlation between cognitive ability and debt outlook. These results suggest that having a more nuanced outlook on debt is approximately as important a determinant of performance in our dynamic optimization task as gender or cognitive ability.

These findings align with Meissner (2016) and Ahrens et al. (2022), which both demonstrate that individuals perform worse when solving dynamic optimization problems that require borrowing relative to saving. This also aligns with Martinez-Marquina and Shi (2022), which shows that holding randomly-assigned debt distorts financial decision-making in an investment game where experimental subjects must choose between investing income and reducing debt balances. This leaves for future research a question that is beyond the scope of this paper – who forms a nuanced outlook on debt and what, if anything, can shift people toward a more positive outlook on debt?

We also observe in the chat data that subjects more often frame discussions in nominal rather than real terms. This is not surprising, given that our *Pairs* tend toward simple heuristics that reduce the complexity of dynamic choice. Since nominal and real earnings are isomorphic, it might be the case that subjects prefer nominal framing because it avoids the added complexity of real framing. This aligns with Petersen and Winn (2014), who find that nominal inertia arising in a choice task results from cognitive complexity and that money illusion exerts second-order effects in the same task.

Finally, we see that our *Pairs* discuss saving and binging as a strategy with surprising frequency. It is easy to assume that such behavior, first documented by Noussair and Matheny (2000), is reactionary since it demonstrates a misunderstanding of the consumption smoothing motive. However, we see here that this behavior can be thoughtful, planned, and forward-looking.

5.1.1 Bargaining Over Consumption

We also see in our chat data that pairs often engage in bargaining to balance consumption preferences. An interesting question is whether bargaining in order to balance preferences and reach a joint decision improves performance in our dynamic optimization task. We evaluate this in Table 8, which presents estimation results from a series of mixed-effects regressions where we project either the absolute unconditional or absolute conditional errors onto a measure of how much bargaining changed a pair's consumption choice and a vector of controls. Note that we restrict our sample to instances where a pair actually bargained to reach a joint decision.

Our coefficient of interest, *Bargain*, measures how bargaining within a period affected the consumption decision for that period. We construct this variable by taking the difference at the period-pair level between the initial consumption spending proposal and the agreed-upon consumption spending amount. Thus, this variable is negative if bargaining increases spending relative to the initial proposal and positive if bargaining decreased consumption spending relative to the initial proposal. Finally, note that we re-scale this variable for ease of interpretation so that it represents the effect of changing a spending decision by 100 ECs via bargaining. CRT and GPA are the same as in Table 4. MixedGender is an indicator variable denoting a mixed-gender pair and BothMale denotes a pair where both participants

	Regression Results - Bargaining					
	Unconditional Absolute Error			Conditional Absolute Error		
	Lifecycle 1	***Lifecycle 2***	**Pooled**	***Lifecycle 1***	***Lifecycle 2***	**Pooled**
	(1)	***(2)***	***(3)***	***(4)***	***(5)***	***(6)***
Bargain	-26.56^{*}	-16.48	-22.62^{**}	-34.00^{***}	-16.00	-26.07^{**}
	(14.13)	(13.25)	(9.199)	(11.71)	(13.74)	(12.38)
GPA	-5.779	-33.49	-18.82	-5.484	-41.78^{*}	-10.35
	(6.222)	(20.77)	(12.06)	(5.535)	(22.70)	(8.754)
CRT	-5.474^{***}	-6.147^{**}	-5.573^{***}	-1.915	-5.882^{*}	-4.833^{**}
	(1.768)	(3.089)	(2.031)	(2.381)	(3.332)	(2.003)
Income	0.454^{**}	0.268	0.354^{**}	1.738^{***}	1.246^{***}	1.436^{***}
	(0.187)	(0.290)	(0.167)	(0.269)	(0.297)	(0.187)
Wealth	-0.0367	-0.114^{**}	-0.0437^{**}	0.173^{***}	-0.134^{***}	0.0931^{*}
	(0.0238)	(0.0498)	(0.0190)	(0.0593)	(0.0497)	(0.0548)
MixedGender	-3.428	-45.26^{**}	-24.99^{*}	-14.66^{**}	-53.45^{**}	-23.70^{**}
	(6.894)	(22.78)	(13.83)	(6.671)	(25.35)	(9.337)
BothMale	1.458	-32.16	-16.96	-1.830	-36.40	-12.39
	(6.269)	(21.92)	(13.38)	(7.221)	(24.48)	(9.926)
Lifecycle2			0.703			3.884^{*}
			(2.271)			(2.098)
Constant	34.10	189.4^{*}	108.9^{*}	-65.04^{**}	167.0	-1.990
	(28.64)	(107.0)	(61.41)	(27.10)	(115.2)	(47.99)
N	637	611	1248	603	578	1181
Clusters	36	36	36	36	36	36

Table 8: This table reports estimates from mixed-effects regressions that estimate how bargaining within a pair impacts performance in our dynamic optimization problem. We restrict our sample to only those observations where a pair in our *Pairs* treatment engaged in bargaining to form a joint decision. Robust standard errors clustered at the observation level in parentheses. * $p < .10$, ** $p < .05$, *** $p < .01$

in a pair are male. We cluster at the observation level and use heteroskedasticity-robust standard errors.

We see that bargaining leads to statistically significantly lower consumption errors along both the unconditionally- and conditionally-optimal consumption paths in two ways. First, bargaining is welfare improving whenever it reigns in overspending. Second, we see in our data that bargaining is also an effective tool for reducing under-spending, which leads to significantly lower errors whenever the initial proposer can convince their partner that drastically reducing spending is unwise. This happens quite often in our data and leads to considerably less under-spending in *Pairs* along both consumption paths relative to both treatments where subjects form individual decisions. We show this graphically by plotting cumulative distribution functions of signed consumption errors along each consumption path in Figure 3.[21]. These distributions are significantly different ($p < .001$ using a KS test).

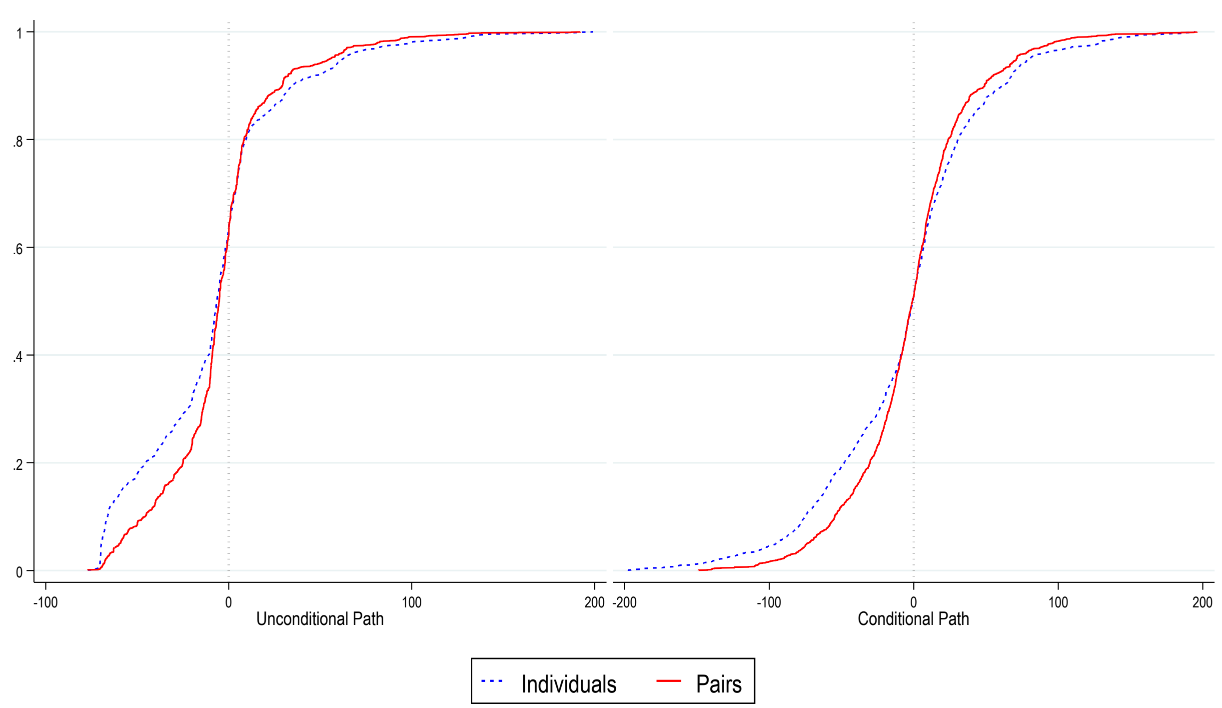


Figure 3: This shows the cumulative distribution functions of signed errors along both the unconditionally-optimal and conditionally-optimal consumption paths.

This finding aligns with Bourguignon et al. (2009), who show that it is more appropriate to treat households with multiple decision-makers as a pair that is trying to balance preferences, rather than as a single rational agent. future research might consider how intrapair bargaining dynamics might change (and how that would impact consumption errors) if subjects in *Pairs* had to earn their income, if bargaining power was asymmetric but exogenous, or if bargaining power was endogenous. However, these questions are beyond the scope of this paper.

5.2 Consumption Heuristics

We also consider which of five different consumption heuristics common in the learning-to-optimize literature best describe the consumption decisions of our participants (See Carbone (2005) and Tasneem and Engle-Warnick (2018) for examples of consumption heuristics).

H1 assumes that a subject consumes all of her income in each period. This is equivalent to having a fixed marginal propensity to consume (MPC) of 1 in each period. A real-world equivalent is an individual or family that lives paycheck-to-paycheck. H2 assumes that subjects optimize perfectly along the unconditionally-optimal path. This heuristic captures the behavior of a fully rational agent in the context of our finite lifecycle problem. H3 assumes that subjects optimize along the conditionally-optimal path, which is akin to learning the

[21] Note that this figure combines data from *Individuals* and *Ledgers*. We exclude extreme outliers in order to provide a better perspective.

unconditionally-optimal solution and then adjusting for past consumption errors. H4 supposes that a subject spends a constant value in each period regardless of income. H5 assumes that an observation i spends a fixed proportion α_i of income in each period.

Table 9: Forecasting heuristics

Model	Heuristic Name	Abbreviation	Model
H1	Hand-to-mouth	H-to-M	$C_t = Y_t$
H2	Unconditional Optimizer	U. Opt.	$C_t = C_t^*$
H3	Conditional Optimizer	C. Opt.	$C_t = C_t^* + \frac{(Y_t - C_t^*) + S_{t-1}}{T-(t-1)}$
H4	Constant Spending	ConSpend	$C_t = C_{t-1} = ... = C_1$
H5	Constant M.P.C.	ConMPC	$C_t = Y_t \frac{\gamma}{30}$, $\gamma = \{1, 2, 3, ..., 29\}$

For each period, we calculate what an observation i would consume according to each consumption heuristic, $C_{i,t}^H$ and the corresponding consumption error $C_{i,t} - C_{i,t}^H$. We then calculate the RMSD for each heuristic for i as $RMSD_i^H = \sqrt{\frac{\sum_{t=1}^{t=j}(C_{i,t} - C_{i,t}^H)^2}{j}}$. We then classify i to whichever heuristic produces the smallest RMSD.

We use this classification method to sort observations into a heuristic every five periods using all available decisions from the current lifecycle. For example, we classify subjects in period ten of lifecycle 1 using consumption decisions from periods one through ten of lifecycle 1. We classify subjects in period 15 of lifecycle 2 using data from periods one through fifteen of lifecycle 2, ignoring all decisions from lifecycle 1. This approach allows us to understand how an observation's heuristic evolves within a lifecycle. Classifications in the final period of each lifecycle correspond to how we would classify each observation if we classified subjects only once using all decisions within a lifecycle. We show results from this exercise at the aggregate level in Figure 4 and at the observation level in Figure 8. Several interesting patterns emerge.

First, we see that the majority of observations in each treatment initially adhere to ConMPC (roughly 60% in lifecycle 1 and 40% in lifecycle 2 in both treatments) but that the proportion of subjects using this heuristic decreases almost monotonically in both treatments and lifecycles as play progresses (to approximately 20% in both treatments and lifecycles). In both *Pairs* and *Individuals* the proportion of ConMPC starts around 60% for lifecycle 1 and around 40% in lifecycle 2 and decreases to about 20% - 30% in both lifecycles of both treatments.

Conversely, we see that the proportion of subjects adhering to H-to-M in both treatments and lifecycles increases with time, eventually emerging as the predominant heuristic in both treatments and in both lifecycles. This is particularly interesting because it indicates that hand-to-mouth consumption observable in observational data could, to some extent, be the outcome of bounded rationality rather than budgetary constraints or financial frictions.

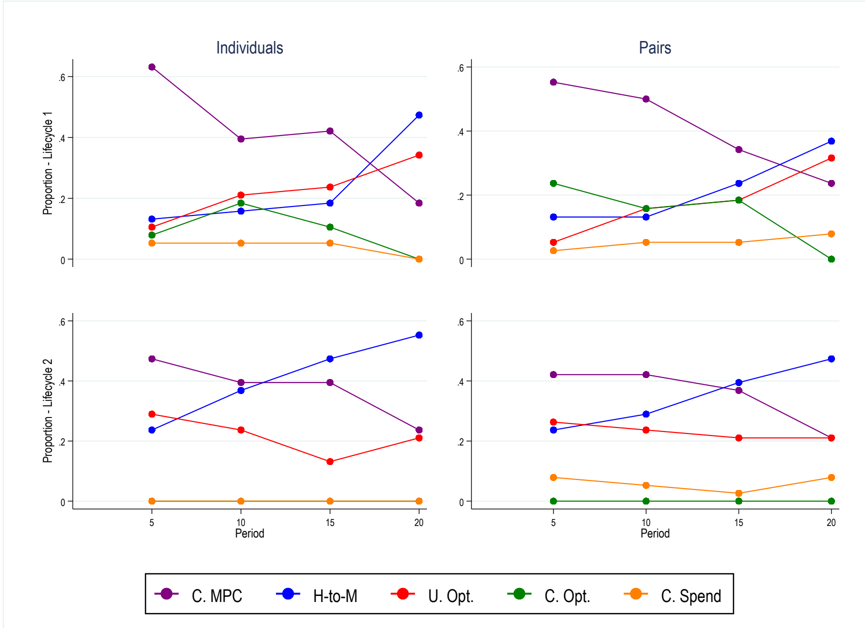


Figure 4: Consumption heuristics by treatment over time.

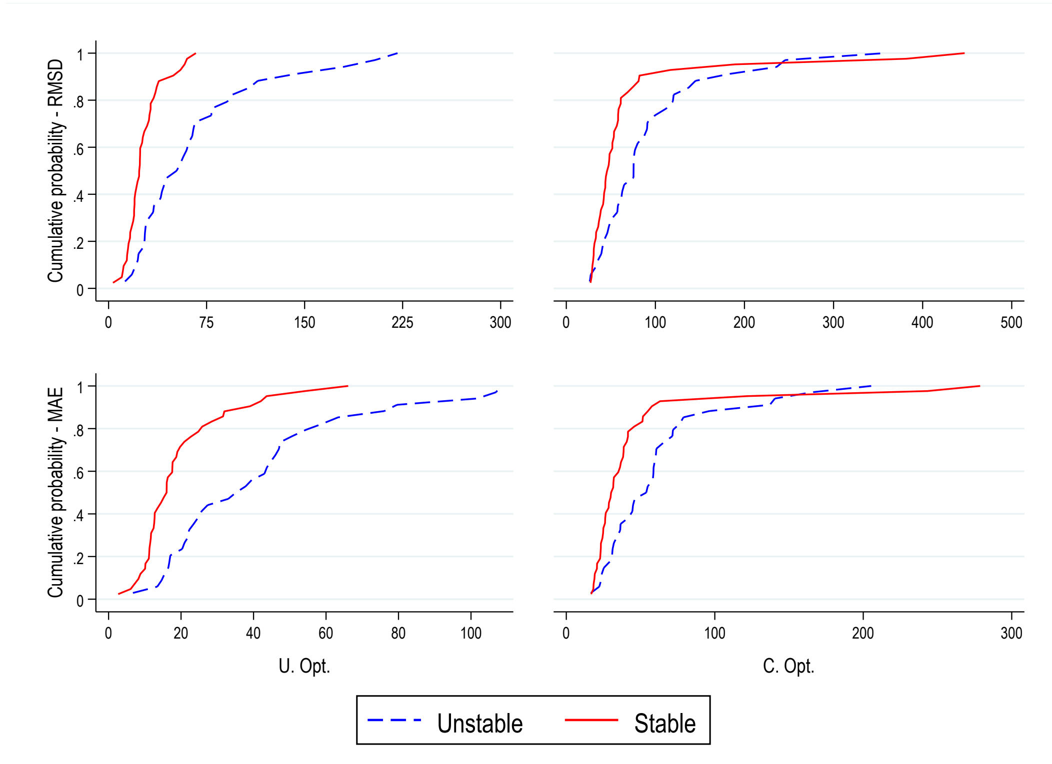
One possibility for why ConMPC decreases is because some ConMPC subjects shift their MPC upward such they become H-to-M subjects. Though both heuristics decrease the complexity of the choice problem relative to calculating the optimal consumption path, H-to-M consumption removes a layer of complexity from ConMPC since subjects do not need to calculate consumption as a proportion of income. This aligns with results in Figure 8 where we see that many observations originally classified as ConMPC transition to H-to-M. Additionally, we see that in both treatments the majority of subjects are using some version of proportional spending, where many have an MPC that is either close to or equal to one. This aligns with results from our textual analysis section where we find that subjects in our *Pairs* treatment typically frame decisions in terms of spending and develop consumption heuristics based on proportional spending, which aligns with results from Carbone (2005).

Second, we note that the proportion of U. Opt. is quite similar in both treatments in both levels and trend. Interestingly, U. Opt. best describes approximately 40% of observations in both treatments by the end of lifecycle 1 and about 20% of observations by the end of lifecycle 2. As depicted by Figure 6 in Section 7.1, unconditionally-optimal behavior requires considerably more borrowing in lifecycle 2 than lifecycle 1. This need for extensive borrowing to achieve the rational benchmark may explain why we see so many fewer unconditional optimizers in lifecyle 2 than in lifecycle 1, given that we know debt aversion typically distorts consumption behavior (Ahrens et al., 2022; Meissner, 2016).

In Figure 8 we show the heuristics subjects in our *Individuals* (top half of the figure) and *Pairs* (bottom half of the figure) adhere to, using a rolling average. Thus, there are 38 rows that correspond to each individual or pair. We classify them based on their consumption decisions every 5 periods, so there are four columns per 20-period lifecycle.

Interestingly, we see in Figure 8 that many of our subjects settle into a heuristic by our second iteration of classification in the experiment and stick with that heuristic while many others

Figure 5: Performance over Stability Classificaiton



switch heuristics throughout a lifecycle. To explore how the relative stability of heuristics impacts behavior in our primary treatments of interest, we first classify observations as either 'stable' or 'unstable'. We consider an observation stable across the experiment if they maintained the same heuristic for at least the last 75% of their decisions and classify them as unstable otherwise. We then compare observation-level measures of performance based on these classifications using both the MWU and KS tests.[22] We show the results of these non-parametric tests in Table 10. We also plot the CDFs of observation-level performance measures over stability classifications in Figure 5, which is located in Section 7.1.[23]

We find that unstable observations significantly underperform along both consumption paths relative to stable observations. This is clear when evaluating the CDFs in Figure 5. Further, this is true using both observation-level measures of performance. The starkness of this result is surprising. On the one hand, we might expect observations that employ stable heuristics to outperform observations employing unstable heuristics if stability is a byproduct of having found a suitable solution to the lifecycle consumption problem. On the other, we might expect them to under-perform relative to unstable observations if stability is a byproduct of an unwillingness to explore the solution space.

We find that unstable observations significantly under-perform along both consumption paths relative to stable observations. This is clear when evaluating the CDFs in Figure 5. Further, this is true using both observation-level measures of performance. The starkness

[22] A two-sample proportions test confirms that the number of unstable observations (18 in *Individuals* and 16 in *Pairs*) and stable pairs (20 in *Individuals* and 22 in *Pairs*) are balanced across treatments ($p = .645$).

[23] We also provide a summary of stability by heuristic in Table 12 in Section 7.1.

Table 10: Performance Differences: Stable vs. Unstable

	RMSD - U *(1)*	RMSD - C *(2)*	MAE - U *(3)*	MAE - C *(4)*
Unstable	66.644	84.879	41.317	56.268
Stable	26.686	57.35	19.560	38.496
MWU	$p < .001$	.001	$p < .001$	.001
KS	$p < .001$	.001	$p < .001$	.002

The top two rows of this table depict mean values for the corresponding observation-level performance measure split by whether observations use stable or unstable heuristics.

of this result is surprising. On the one hand, we might expect observations that employ stable heuristics to outperform observations employing unstable heuristics if stability is a byproduct of having found a suitable solution to the lifecycle consumption problem. On the other, we might expect observations with stable heuristics to under-perform relative to unstable observations if stability is a byproduct of an unwillingness to expend additional effort solving the dynamic optimization problem.[24]

6 Conclusion

This paper revisits the learning-to-optimize literature to study the relative ability of pairs and individuals to solve a finite-period, dynamic optimization problem. We find that joint decision-making leads to significantly better performance along both the unconditionally- and conditionally-optimal consumption paths. This performance gap, on average, leads to subjects in our *Pairs* treatment earning about 23% more than subjects in our *Individuals* treatment.

Our results demonstrate that simple household differences – whether or not there are multiple decision-makers in the household – can lead to systematic differences in budgetary decisions and, as a consequence, systematic differences in welfare. Though we abstract considerably from the complexity of the real world, our experimental design sheds some light on why we observe in observational data that married households in America's bottom income quartile better smooth negative income shocks than do single households and why researchers have so consistently empirically rejected unitary models of intra-household decision making. One implication of our results is that increasing access to financial and budgetary planning services might be a reasonably cheap and affordable way to increase welfare for lower-income, single households.

We show that joint decision-makers often engage in iterative bargaining that improves their decisions. In our experiment, both members of a pair faced the same utility function, shared

[24]Note that these results also hold for *Ledger*, where we see 15 stable and 15 unstable observations. We show a summary of heuristics by stability including *Ledger* observations in Table 14 and stability results including *Leger* observations in Table 13, both located in Section 7.1

the benefits of consumption equally, were equal contributors to household income (since income was exogenous), and held symmetric bargaining power. These design choices yield a setting that is likely a drastic simplification of real-world intra-household dynamics. Future research could relax these design features to understand how they impact joint decision-making.

Our results differ from the few other studies that compare the performance of pairs and individuals in a dynamic optimization task. Importantly, our paper is the first to offer evidence that joint decision-making can lead to better budgetary decisions than individual decision-making. This result better aligns the experimental literature on joint decisions in dynamic optimization with the broader literature on joint decision-making by suggesting that pairs can better smooth consumption than individuals in at least some contexts.

There are several differences between our design and the designs used in Carbone and Infante (2015) and Carbone et al. (2019) that can rationalize these differences. Because of this, we view our results as complementary to theirs. These two previous studies both feature positive nominal interest rates and do not provide subjects with a consumption calculator. We set interest rates to zero and provide a consumption calculator. Both design choices may work to reduce the complexity of the choice problem for subjects. It is reasonable to think that a choice problem can be either sufficiently easy that there is no room for performance differences or sufficiently complex that forming joint decisions is unlikely to matter. If so, it is possible our design lies somewhere between these two extremes. Further, our experimental design requires and allows for borrowing to achieve the rational benchmark, whereas Carbone and Infante (2015) and Carbone et al. (2019) do not. It is also possible that joint decision-makers are better able to handle this feature. Future work could more carefully consider how group decisions respond to debt, especially in a setting that focuses on durable goods consumption where debt is more likely to matter.

We use textual analysis from *Pairs* chat data and from a post-experiment survey-of-decisions to try and understand why we observe these performance differences and also how people approach solving dynamic optimization problems. Chat data suggests that pairs often negotiate joint consumption decisions by updating toward one another. This is corroborated by our analysis of chat data that shows bargaining within pairs leads to statistically significant improvements in consumption errors. It is further supported by responses to our survey-of-decisions question "What was your strategy for overcoming disagreements?" where the overwhelming majority of pairs indicated that they used mutual compromise to reach a joint decision. This suggests that at least one benefit of forming joint decisions is a sort of 'wisdom of the crowd' effect. This moderation of more extreme decisions reduces boom-bust consumption cycles leading to less extreme errors. We note that income in our setting is exogenous, members of a pair split consumption utility equally, and bargaining power is both exogenous and symmetric. Future research might study how relaxing these assumptions impacts intra-pair bargaining since each likely constitutes a meaningful simplification of real-world household dynamics.

Additionally, pairs almost exclusively frame discussions in terms of spending even though the stochastic per-period income process, coupled with the consumption smoothing motive, makes saving and borrowing important components of earnings maximization. Further, we see that Pairs develop simple heuristics that can lead to persistent errors that compound over time, which helps explain why average absolute conditional errors are larger than absolute unconditional errors. Finally, we see that saving-and-binging can be the outcome of forward-looking behavior rather than the result of extreme myopia or lack of a strategy.

We also provide suggestive evidence that having a more nuanced outlook on debt leads to better performance in a consumption smoothing problem where borrowing is a necessary component of optimal behavior. A potential implication of this is that financial education focused on the potential benefits and safe use of debt could improve budgetary decisions insofar as it eases strictly negative outlooks on debt.

Classifying subjects into heuristics reveals that a substantive proportion of subjects are best categorized as unconditional optimizers. This is perhaps due to the inclusion of a consumption tool that reduces the complexity of our optimization task. If so, this suggests that providing increased access to budgetary tools and/or advice may lead real-world households to behave in a more theory-consistent way.

Finally, we show that consumption heuristics are not necessarily stable over time. This might be because heuristics evolve with experience or perhaps because income dynamics influence heuristics. We show that observations with relatively stable heuristics significantly outperform observations with unstable heuristics.

Bibliography

Ahrens, S., C. Bosch-Rosa, and T. Meissner (2022). Intertemporal consumption and debt aversion: A replication and extension. *arXiv preprint arXiv:2201.06006*.

Ballinger, T., E. Hudson, L. Karkoviata, and N. Wilcox (2011). Saving behavior and cognitive abilities. *Experimental Economics 14*(3), 349–374.

Ballinger, T. P., M. G. Palumbo, and N. T. Wilcox (2003). Precautionary saving and social learning across generations: an experiment. *The Economic Journal 113*(490), 920–947.

Bao, T., J. Duffy, and C. Hommes (2013). Learning, forecasting and optimizing: An experimental study. *European Economic Review 61*, 186–204.

Blinder, A. S. and J. Morgan (2005). Are two heads better than one? monetary policy by committee. *Journal of Money, Credit and Banking*, 789–811.

Blinder, A. S. and J. Morgan (2007). Leadership in groups: A monetary policy experiment.

Bourguignon, F., M. Browning, and P. Chiappori (2009). Efficient intra-household allocations and distribution factors: implications and identification. *The Review of Economic Studies 76.2*, 503–528.

Brown, A. L., Z. E. Chua, and C. F. Camerer (2009). Learning and visceral temptation in dynamic saving experiments. *The Quarterly Journal of Economics 124*(1), 197–231.

Browning, M., F. Bourguignon, P. Chiappori, and V. Lechene (1994). Income and outcomes: A structural model of intrahousehold allocation. *Journal of Political Economy 102*(6), 1067–1096.

Carbone, E. (2005). Demographics and behaviour. *Experimental Economics 8*(3), 217–232.

Carbone, E. (2006). Understanding intertemporal choices. *Applied Economics 38*(8), 889–898.

Carbone, E. and J. Duffy (2014). Lifecycle consumption plans, social learning and external habits: Experimental evidence. *Journal of Economic Behavior & Organization 106*, 413–427.

Carbone, E., K. Georgalos, and G. Infante (2019). Individual vs. group decision-making: an experiment on dynamic choice under risk and ambiguity. *Theory and Decision 87*(1), 87–122.

Carbone, E., J. Hey, and T. Neugebauer (2021). An experimental comparison of two exchange economies: Long-lived asset vs. short-lived asset. *Management Science 67*(11), 6946–6962.

Carbone, E. and J. D. Hey (2001). A test of the principle of optimality. *Theory and Decision 50*(3), 263–281.

Carbone, E. and J. D. Hey (2004). The effect of unemployment on consumption: an experimental analysis. *The Economic Journal 114*(497), 660–683.

Carbone, E. and G. Infante (2014). Comparing behavior under risk and under ambiguity in a lifecycle experiment. *Theory and Decision 77*(3), 313–322.

Carbone, E. and G. Infante (2015). Are groups better planners than individuals? an experimental analysis. *Journal of Behavioral and Experimental Economics 57*, 112–119.

Carvalho, L. S., S. Meier, and S. W. Wang (2016). Poverty and economic decision-making: Evidence from changes in financial resources at payday. *American Economic Review 106*(2), 260–84.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Østling (2017). The effect of wealth on individual and household labor supply: evidence from swedish lotteries. *American Economic Review 107*(12), 3917–46.

Chakraborty, A. and G. Fenig (2022). Together everyone achieves more (team): Incentives for productivity. Technical report.

Charness, G. and M. Sutter (2012). Groups make better self-interested decisions. *Journal of Economic Perspectives 26*(3), 157–76.

Cooper, D. J. and J. H. Kagel (2005). Are two heads better than one? team versus individual play in signaling games. *American Economic Review 95*(3), 477–509.

Duffy, J. (2016). Macroeconomics: a survey of laboratory research. *Handbook of Experimental Economics 2*, 1–90.

Duffy, J. and A. Orland (2021). Liquidity constraints and buffer stock savings: Theory and experimental evidence. *Available at SSRN 3672126*.

Eckel, C. C. and P. J. Grossman (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior 23*(4), 281–295.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Flavin, M. A. (1981). The adjustment of consumption to changing expectations about future income. *Journal of Political Economy 89.5*, 974–1009.

Fortin, B. and G. Lacroix (1997). A test of the unitary and collective models of household labour supply. *The Economic Journal 107*(443), 933–955.

Fréchette, G. (2012). Session-effects in the laboratory. *Experimental Economics 15.3*, 485–498.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives 19*(4), 25–42.

Hall, R. E. (1978). Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence. *Journal of Political Economy 86*(6), 971–987.

Haushofer, J. and E. Fehr (2014). On the psychology of poverty. *Science 344*(6186), 862–867.

Hayashi, F. (1982). The permanent income hypothesis: estimation and testing by instrumental variables. *Journal of Political Economy 90.5*, 895–916.

Hey, J. D. and V. Dardanoni (1988). Optimal consumption under uncertainty: An experimental investigation. *The Economic Journal 98*(390), 105–116.

Hey, J. D. and J. A. Knoll (2011). Strategies in dynamic decision making–an experimental investigation of the rationality of decision behaviour. *Journal of Economic Psychology 32*(3), 399–409.

Jackson, M. O. and L. Yariv (2014). Present bias and collective dynamic choice in the lab. *American Economic Review 104*(12), 4184–4204.

Kagel, J. H. and P. McGee (2016). Team versus individual play in finitely repeated prisoner dilemma games. *American Economic Journal: Microeconomics 8*(2), 253–76.

Kugler, T., G. Bornstein, M. G. Kocher, and M. Sutter (2007). Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic Psychology 28*(6), 646–657.

Lombardelli, C., J. Proudman, and J. Talbot (2005). Committees versus individuals: An experimental analysis of monetary policy decision-making. *International Journal of Central Banking.*

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao (2013). Poverty impedes cognitive function. *Science 341*(6149), 976–980.

Martinez-Marquina, A. and M. Shi (2022). The opportunity cost of debt aversion. *Working paper.*.

Meissner, T. (2016). Intertemporal consumption and debt aversion: an experimental study. *Experimental Economics 19*(2), 281–298.

Noussair, C. and K. Matheny (2000). An experimental study of decisions in dynamic optimization problems. *Economic Theory 15*(2), 389–419.

Petersen, L. and A. Winn (2014). Does money illusion matter? comment. *American Economic Review 104*(3), 1047–62.

Primi, C., K. Morsanyi, F. Chiesi, M. A. Donati, and J. Hamilton (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (irt). *Journal of Behavioral Decision Making 29*(5), 453–469.

Rholes, R. and L. Petersen (2021). Should central banks communicate uncertainty in their projections? *Journal of Economic Behavior & Organization 183*, 320–341.

Tasneem, D. and J. Engle-Warnick (2018). Decision rules for precautionary and retirement savings. Technical report, CIRANO.

Zeldes, S. P. (1989). Consumption and liquidity constraints: an empirical investigation. *Journal of Political Economy 97.2*, 305–346.

7 Appendix

7.1 Tables and Figures

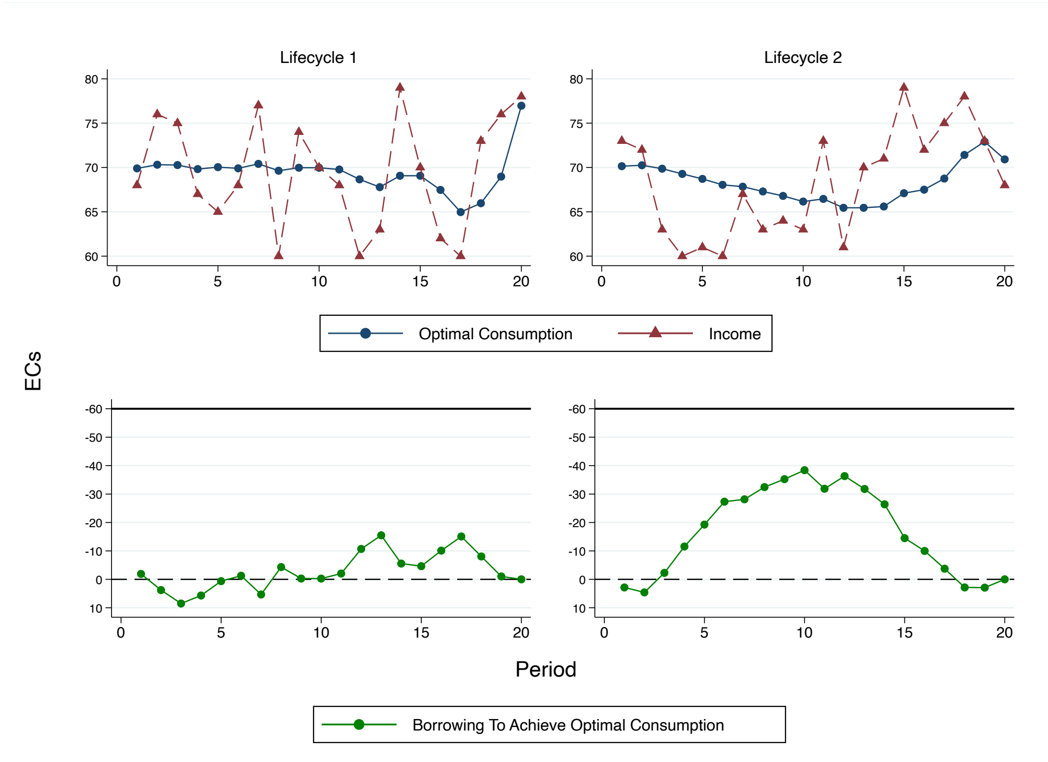


Figure 6: The top panel of this figure illustrates the unconditionally-optimal consumption paths alongside the pre-drawn stochastic income paths for both lifecycles. The bottom panel of this figure denotes the level of borrowing necessary to move along the corresponding unconditionally-optimal consumption paths. Note the solid black lines in the bottom figures denote the per-period borrowing constraint of 60 ECs, which applies to periods 1 through 19 of each lifecycle. Finally, note that negative values in the bottom panel denote saving while positive values denote borrowing. Note that the borrowing constraint is never binding for subjects in our experiment.

We provide an example, corresponding to Figure 12, that explains how an individual might use the consumption tool and the available information to play this game.

Notice under ‘Income this period’ that our hypothetical subject has received an endowment of 68 experimental credits (ECs) in period 1. This is reflected in the “Bank account balance,” which updates each period to account for per-period and previous saving/borrowing. The subject may then explore the outcome of all possible consumption decisions using the ‘Potential consumption spending’ slider or by entering hypothetical levels of consumption in the gray box labeled ‘Potential consumption spending’.

For this example, our subject could spend between 0 and 128 ECs, since subjects could borrow up to 60ECs in all but the final period of a lifecycle. Moving the slider or entering a value in the box and clicking calculate will update all other variables. In Figure 12, our hypothetical subject has selected a potential consumption value of 40. Notice that all available information has been updated to reflect this. “Consumption this period” is updated

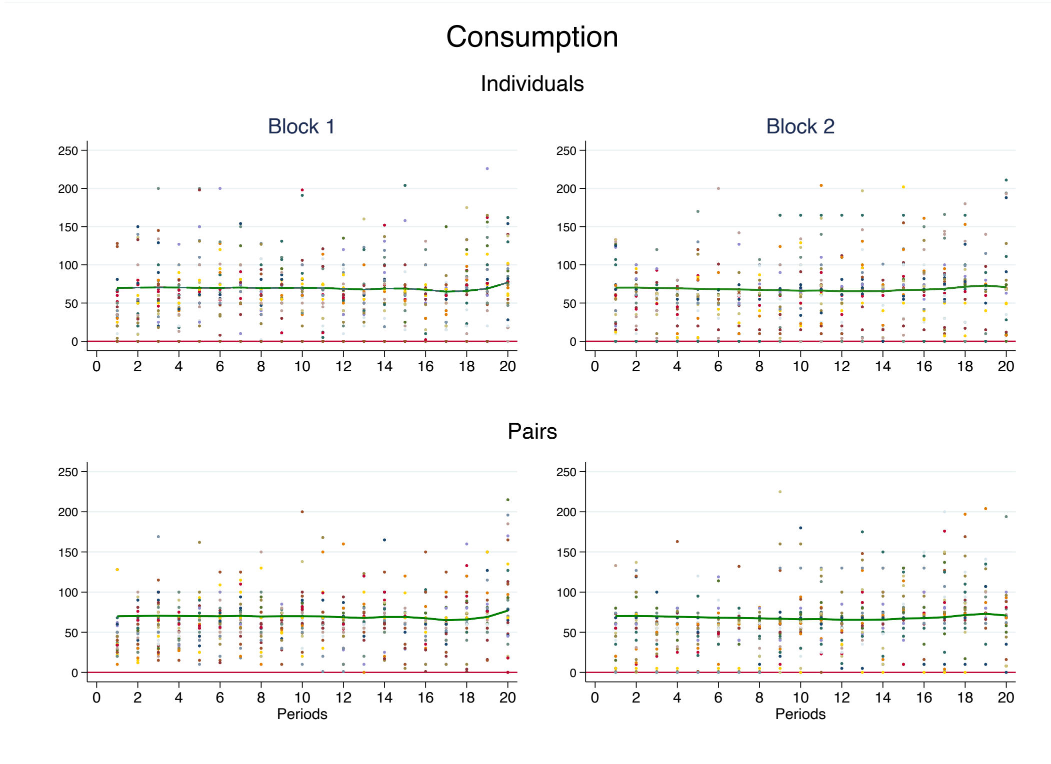


Figure 7: This figure shows all observation-level consumption decisions made in all periods and in both lifecycles for *Individuals* and *Pairs*. The solid green line in each panel denotes the optimal level of consumption.

to reflect the chosen value of 40.

The "Saving/Borrowing" field updates to 28 to reflect the 28 ECs that would remain in the subject's bank account after spending 40 of the available 68 ECs[25] This balance is also shown in the "Bank account balance" field within the consumption calculator.[26] Further, the "Consumption points" field updates to show the consumption points earned under a choice of spending 40 ECs on consumption, which is 6.

The subject is also shown the marginal utility from using one more EC on consumption in the "Marginal increase" field, which is 1.200. The subject then enters their chosen value for consumption in the 'Consumption spending' box and presses the red button labeled 'Continue' to proceed.

7.2 Average Consumption and Consumption Heterogeneity

This section of the appendix provides details on the average consumption and consumption heterogeneity by period for both *Pairs* and *Individuals*. We measure consumption heterogeneity as the cross-sectional standard deviation of consumption decisions within a period. We graph both in Figure 14.

[25]This number would be negative if the subject decided to spend more than 68 ECs.
[26]These numbers match because this is period 1. They would not necessarily match in later periods.

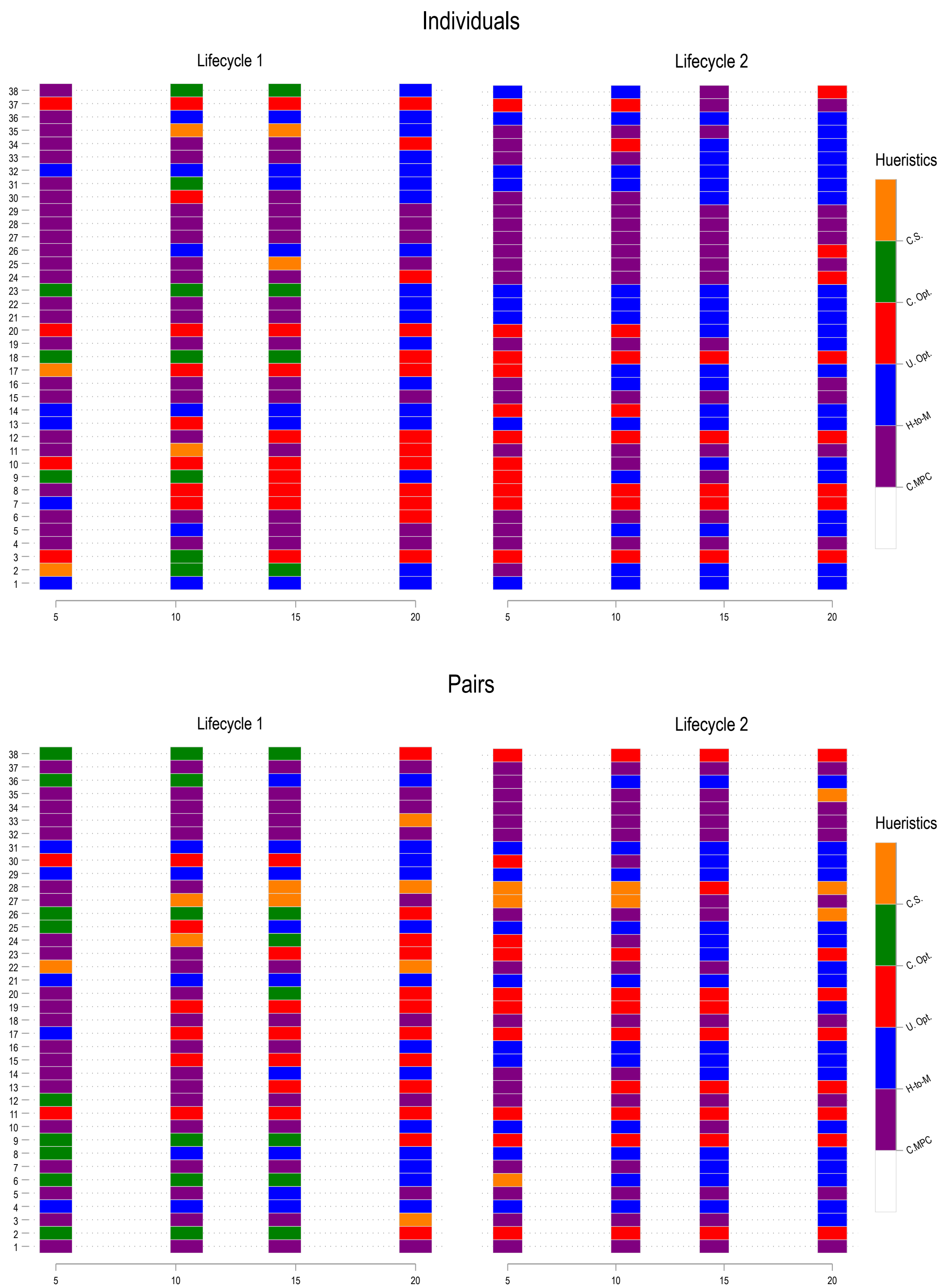


Figure 8: CDFs of observational-level performance measures by heuristic stability

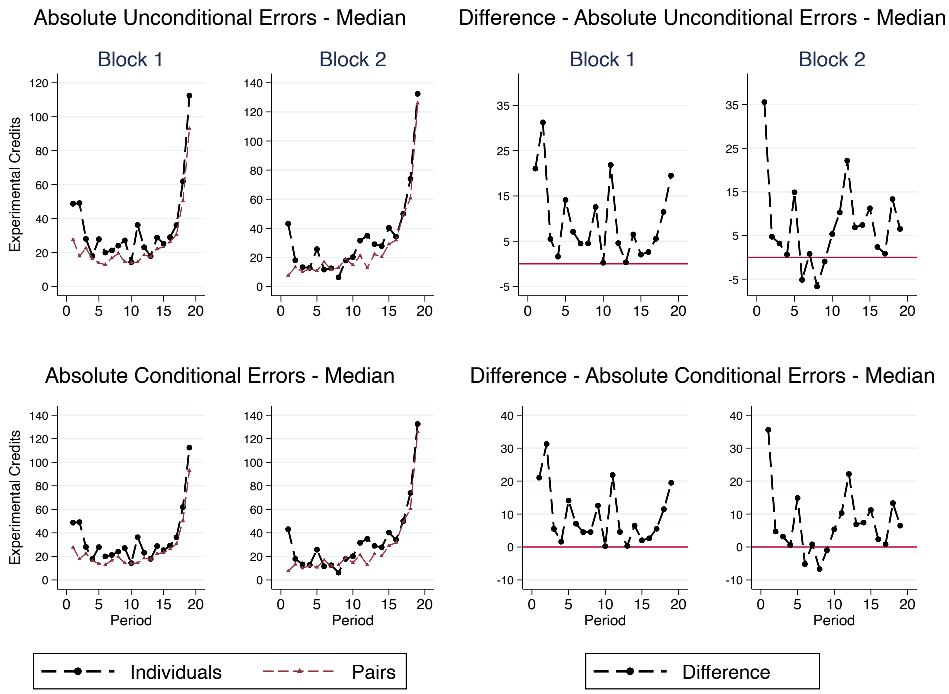


Figure 9: This figure depicts treatment-level median absolute consumption errors and their differences. Values above zero in the differences panels indicate that Pairs outperformed *Individuals* in that period.

Worth noting in Figure 14 is that average consumption increases over time for both *Individuals* and *Pairs* in both decision lifecycles. Our results replicate a common finding in this literature that subjects under-consume in early periods and over-consume in later periods. This is especially true in the first half of our second lifecycle where subjects must borrow to consume along the unconditionally-optimal path. To see this, refer to the stochastic income process depicted in Figure 1, and note that per-period income was consistently below the unconditionally-optimal level of consumption. Finally, we note that participants in our *Individuals* treatment seem to exhibit more heterogeneity in consumption than do subjects in our *Pairs* treatment.

	Balance Table		
	Individuals	**Pairs**	***Ledger***
	(1)	*(2)*	*(3)*
Avg. GPA	3.41	3.45	3.61
Gender (% Male)	72.41	59.2	36.67
Avg. Age	23.29	23.29	–
Avg. Outside Debt ($)	6975.86	12563.51	16585.61

Table 11: This table reports the balance across treatments.

Table 12: Performance Differences: Stable vs. Unstable

Heuristic	**Stable** *(1)*	**Unstable** *(2)*	**Total** *(3)*
ConMPC	15	03	18
H-to-M	16	23	29
U. Opt.	11	05	16
C. Opt.	00	00	00
ConSpend	00	03	03
Total	42	34	76

This table shows the final heuristic classification of stable and unstable observations.

Table 13: Performance Differences: Stable vs. Unstable (Including *Ledger* Observations)

	RMSD - U *(1)*	**RMSD - C** *(2)*	**MAE - U** *(3)*	**MAE - C** *(4)*
Unstable	63.68	87.00	40.14	57.14
Stable	30.02	56.22	21.44	38.25
MWU	p < .001	.001	p < .001	.001
KS	p < .001	.002	p < .001	.004

The top two rows of this table depict mean values for the corresponding observation-level performance measure split by whether observations use stable or unstable heuristics.

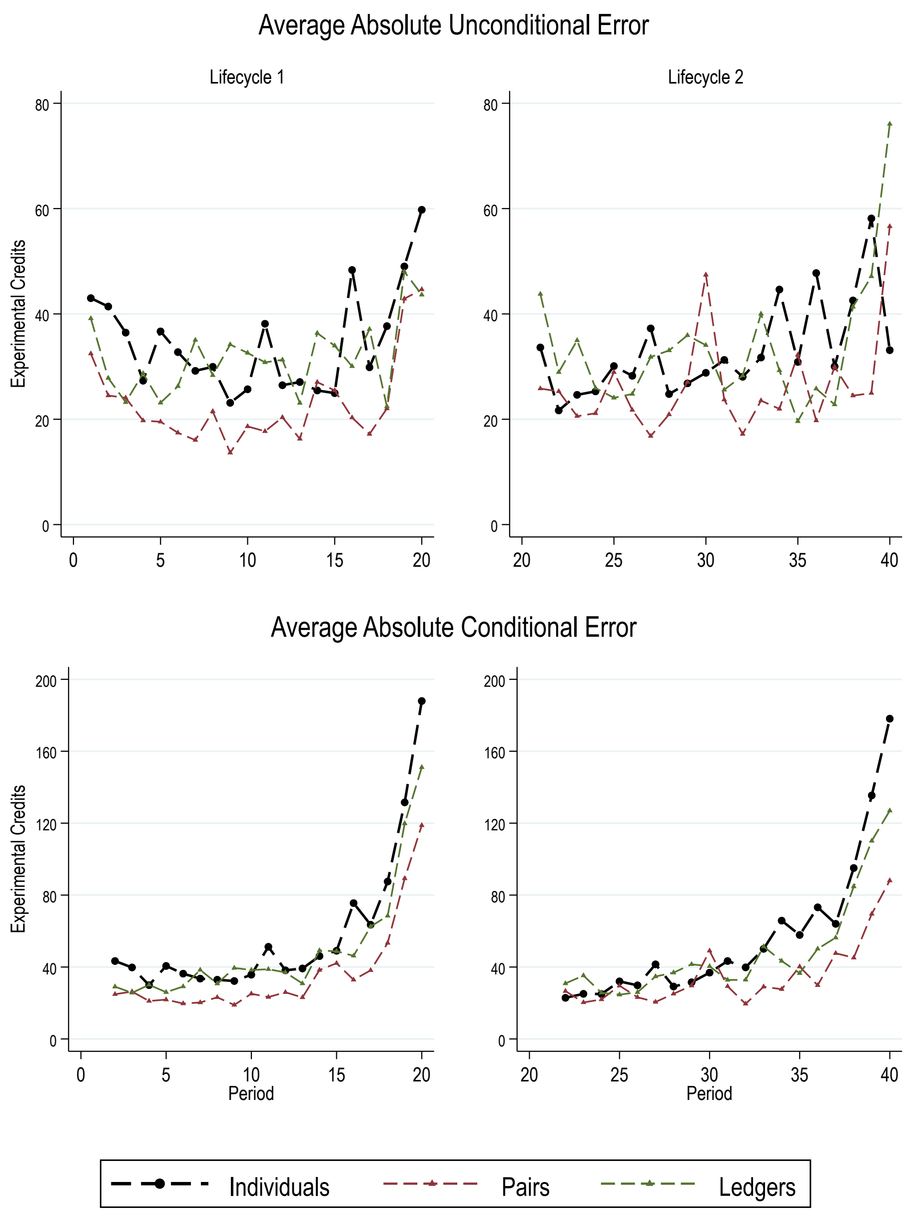


Figure 10: This figure depicts treatment-level absolute consumption errors along the unconditional (top row) and conditional (bottom row) absolute consumption paths for *Ledger* (green dashed lines), *Individual* (black dashed lines), and for *Pairs* (red dashed lines) treatments. The left column depicts values from Lifecycle 1 while the right depicts values from Lifecycle 2.

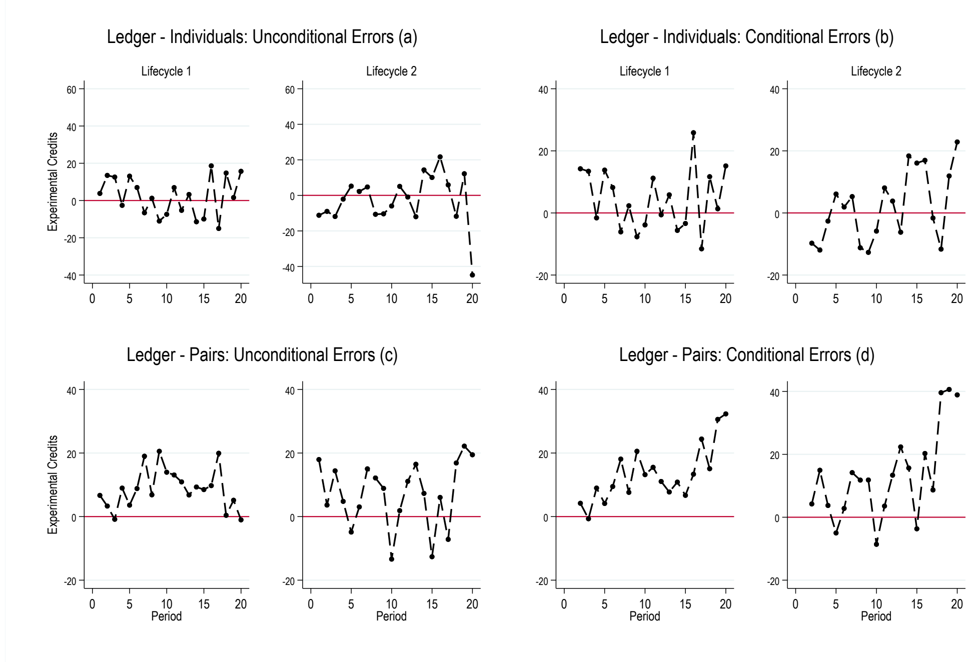


Figure 11: This figure depicts treatment-level differences in unconditional and conditional absolute consumption errors for *Ledger*and *Individual* treatments (panels a and b, respectively), and for *Ledger* and *Pair* treatments (panels c and d, respectively). Values above zero in all panels indicate an instance where the treatment-level absolute consumption error was higher in the *Ledger* treatment than in the comparison treatment.

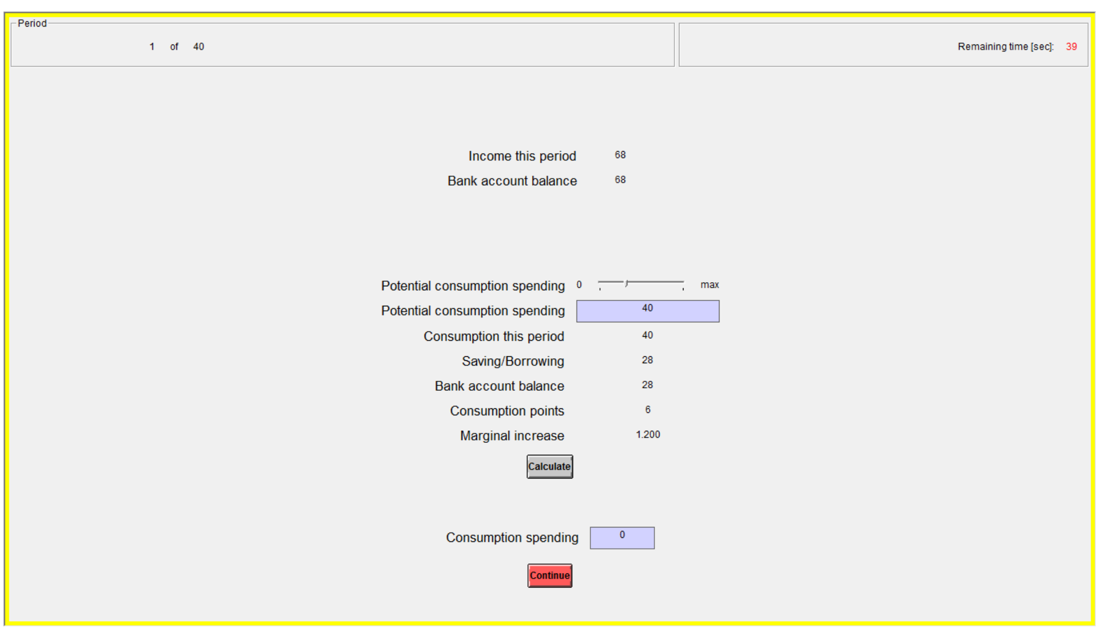


Figure 12: Decision screen for *Individuals* treatment.

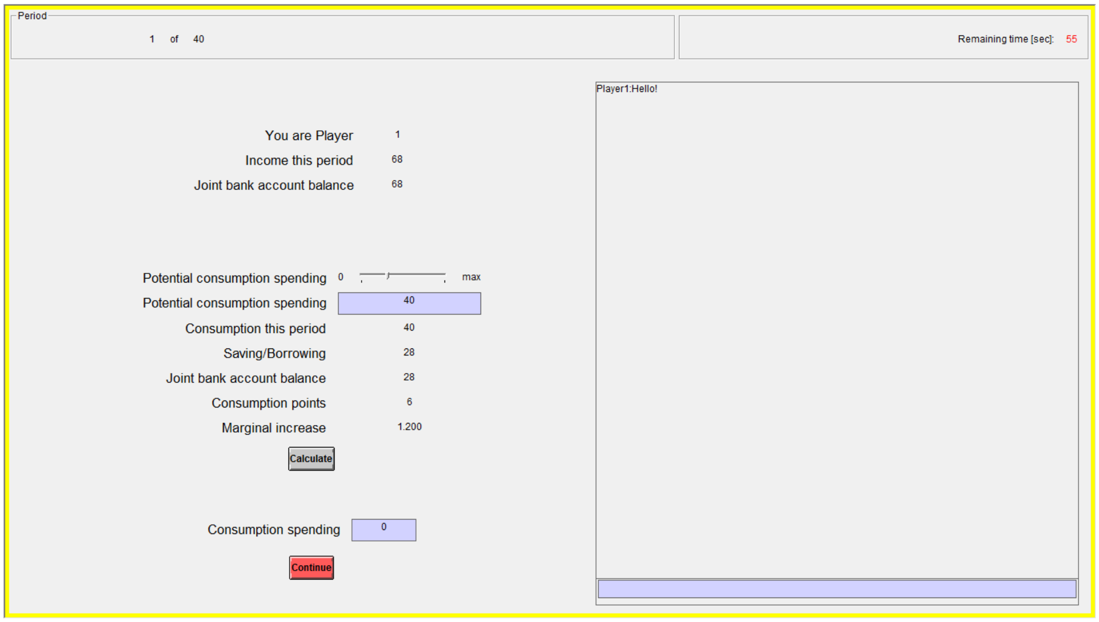
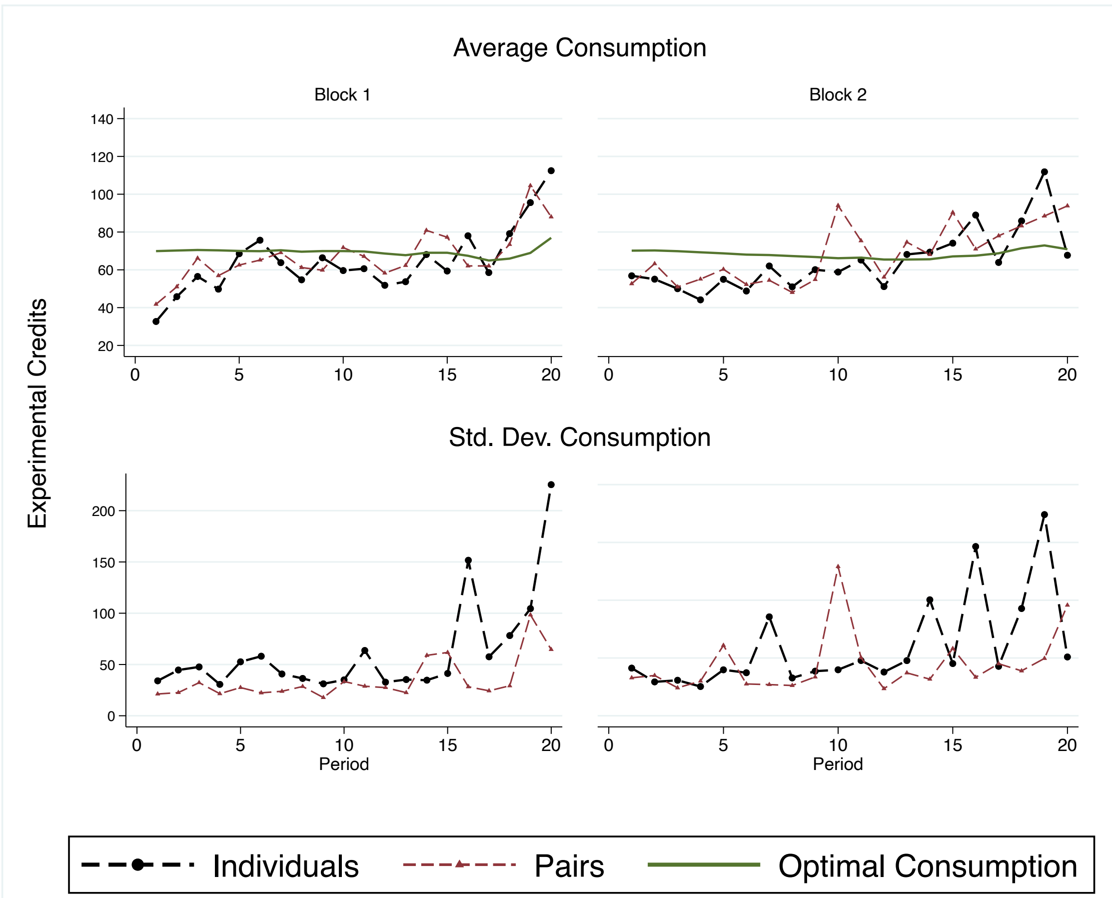


Figure 13: Decision screen for Pairs treatment.

Table 14: Performance Differences: Stable vs. Unstable

Heuristic	**Stable** *(1)*	**Unstable** *(2)*	**Total** *(3)*
ConMPC	19	06	25
H-to-M	24	30	54
U. Opt.	13	08	21
C. Opt.	01	00	01
ConSpend	00	05	05
Total	49	57	106

This table shows the final heuristic classification of stable and unstable observations.



Average Consumption
Block 1
Block 2
Experimental Credits
140
120
100
80
60
40
20
0
5
10
15
20
Std. Dev. Consumption
200
150
100
50
0
Period
Individuals
Pairs
Optimal Consumption

Figure 14

	Mixed Effects Estimations By Lifecycle											
	Unconditional Absolute Error						**Conditional Absolute Error**					
	L1	***L1***	***L2***	***L2***	***Pooled***	***Pooled***	***L1***	***L1***	***L2***	***L2***	***Pooled***	***Pooled***
	(1)	***(2)***	***(3)***	***(4)***	***(5)***	***(6)***	***(7)***	***(8)***	***(9)***	***(19)***	***(11)***	***(12)***
Individual	-11.54^{**}	-14.44^{**}	-6.462	-9.307	-9.001^{*}	-11.04^{*}	-21.45^{**}	-15.69^{***}	-21.29^{**}	-10.72^{*}	-21.37^{**}	-13.55^{***}
	(4.884)	(6.664)	(6.024)	(8.662)	(5.195)	(6.594)	(9.923)	(5.380)	(9.762)	(5.806)	(9.241)	(5.021)
Ledger	-2.853	-0.659	0.694	5.708	-1.080	3.050	-8.133	5.172	-8.167	1.914	-8.150	2.613
	(5.144)	(6.732)	(6.530)	(9.941)	(5.500)	(6.898)	(11.56)	(5.678)	(11.91)	(7.921)	(11.01)	(5.919)
CRT		-4.022**		-5.222^{**}		-4.258^{***}		-1.313		-3.192^{*}		-2.500^{*}
		(1.730)		(2.233)		(1.650)		(1.350)		(1.674)		(1.395)
GPA		-10.68		-12.39		-10.48		-4.143		-5.667		-5.360
		(7.965)		(9.427)		(7.523)		(5.532)		(7.224)		(5.914)
1.Male		5.034		9.548		6.345		0.980		6.217		3.805
		(4.999)		(8.290)		(5.209)		(4.846)		(6.592)		(4.885)
2.Male		6.253		12.66		8.821		7.056		5.594		6.100
		(5.546)		(9.848)		(6.503)		(5.037)		(7.940)		(5.665)
Wealth		-0.0440^{*}		-0.0661^{**}		-0.0361^{**}		0.225^{***}		0.180^{***}		0.192^{***}
		(0.0236)		(0.0273)		(0.0146)		(0.0392)		(0.0631)		(0.0521)
Income		0.477^{**}		0.454^{**}		0.465^{***}		2.059^{***}		1.908^{***}		1.902^{***}
		(0.219)		(0.227)		(0.149)		(0.361)		(0.263)		(0.233)
Constant	34.62^{***}	48.07	32.97^{***}	51.85	33.79^{***}	44.35	57.60^{***}	-97.90^{***}	56.67^{***}	-71.42^{**}	57.13^{***}	-75.40^{**}
	(4.376)	(34.07)	(4.851)	(36.21)	(4.447)	(29.90)	(9.341)	(35.56)	(9.017)	(30.90)	(8.746)	(30.75)
N	2120	1700	2120	1700	4240	3400	2014	1615	2014	1615	4028	3230
Clusters	106	85	106	85	106	85	106	85	106	85	106	85

Table 15: This table shows estimations from mixed effects regressions where we partition data by lifecycle. Columns(1) - (6) use unconditional absolute errors as the outcome variable and columns (7) - (12) use conditional absolute errors. Columns labeled ***L1*** use data from Lifecycle 1, ***L2*** from Lifecycle 2, and ***Pooled*** from both lifecycles. Robust standard errors clustered at the unit of obsevation are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7.3 Empirical Data

We analyze data from the Panel Study of Income Dynamics (PSID) at the family-level, or household level, to evaluate consumption smoothing. The PSID is a longitudinal household survey that has been conducted by the Institute for Social Research at the University of Michigan since 1968.[27]

Administrators of the PSID survey ask participants to report on their consumption expenditure totals across a large range of items over the course of the previous year. Examples of these items include food, utilities, transportation, education, childcare, and health care. Because respondents self-report data, people may over- or under-report their incomes and consumption expenditures for personal reasons, or due to memory lapse. Additionally, the PSID over-samples low-income families. However, we control for this in our analysis.

We restrict our sample to the years 1999 - 2017. We do this to account for two major changes in PSID data collection that came in 1999. First, surveyors began collecting data biannually instead of annually. Second, the PSID became a richer data source as surveyors began collecting additional information about household consumption and income.

For the purposes of this exercise, we made certain sample selection decisions when cleaning the PSID data. We restrict the sample to household heads aged 20 to 65. We used the OECD-modified adult equivalence scale to adjust for the increase that is proportionate per adult necessary to maintain some standard of living given a change in demographic circumstances, like the birth of a new child. We then adjusted all consumption and income measures by the personal consumption expenditure (PCE) index, to account for changes in prices, and by the OECD-modified equivalence scale. We drop all observations from the original Survey of Economic Opportunity (SEO) sample and the branches of this original sample to avoid the bias that would be introduced from the over sample of poor households, restricting our sample to just the Survey Research Center (SRC) sample. We drop observations where the household head reported working more than 5,200 hours or the household head reported working more than 520 hours at half of the minimum wage. We also drop observations where consumption expenditures are reported to be zero or negative. Thus, we restrict the sample to observations that only report positive consumption expenditures. Finally, we restrict the sample to the lowest income quartile.

7.4 Empirical Motivation

This section provides suggestive evidence that further motivates our laboratory experiment. We do not intend for this to constitute causal evidence that single households are less effective at smoothing over negative income shocks than non-single households. Instead, we use this exercise to propose that there may be differences between individual and joint decision-making that cannot be untangled using observational data alone.

Using data from the Panel Study of Income Dynamics (PSID) ranging from 1999 to 2017, we estimate the relationship between marital status and income growth for households experi-

[27]Surveyors collect data on a range of topics including education, employment, income, wealth, and expenditures, which makes it well-suited for the study of consumption smoothing.

encing a negative income shock. We define a negative income shock as having occurred when the head of household reports spending at least one month unemployed and experiencing negative income growth.

Specifically, we estimate the following equation via feasible generalized least squares (FGLS) estimation:

$$\ln(\Delta c_{i,t}) = \alpha + \beta_1 \text{Single} + \sum_{j=1}^{4} Y_j + \gamma_t + \eta_t X_{i,t} + \nu Z_i + \epsilon_{i,t}. \tag{9}$$

Here, Single is an indicator variable that takes a value of 1 if a head of household is single, $\sum_{j=1}^{4} \beta_2^j Y_j$ denotes group-level fixed effects for income quartile (with $j \in \{1, 2, 3, 4\}$), γ_t represents time fixed effects, Xi, t denotes a vector of time-varying household characteristics for education level, household size, and wealth balances, Z_i is a vector of time-invariant household characteristics like gender and race, and $\epsilon_{i,t}$ denotes our error term.

We estimate this equation using FGLS to account for potential heteroskedasticity and autocorrelation in the error term, leading to more efficient and consistent parameter estimates. We include time-level fixed effects to control for unobserved time-specific factors that may affect income growth or marital status, such as macroeconomic conditions or policy changes. We include income group fixed effects[28] to control for unobserved group-level characteristics that could influence the relationship between marital status and income growth. Note that our outcome of interest is the natural log of the first difference of consumption $ln(\Delta c_{i,t}) = ln(c_{i,t} - c_{i,(t-1)})$, where we consider food consumption, non-durable consumption, durable consumption, and total consumption. Given this log-linear specification, the interpretation of our coefficient of interest, β_1, is that $100 * \beta_1$ gives the percentage point (pp) difference in consumption growth for a given consumption class for a single household relative to a married household.

Columns one through three in Table 16 report $\hat{\beta}_1$ for food consumption, consumption of non-durable goods, and durable goods. Column four reports $\hat{\beta}_1$ when using total consumption to construct $\ln(\Delta c_{i,t})$. These estimates indicate that, on average, single households respond to a negative income shock by reducing food consumption 20.18 pp more than married households, reducing non-durable consumption by 9.46 pp more than married households, durable consumption by 10.86 pp more than married households, and total consumption by 12 pp more than married households.

One data limitation we face in this exercise is that very few heads of household report sustaining negative income shocks. We outline this in Table 17, which reports the number of income shocks we observe in our sample. From a total of 27,963 observations, we observe only 568 observations of negative income shocks. The majority of these are concentrated among the lowest-income quartile.

Though this considerably hinders our ability to provide more nuanced insight, we do have sufficient data to focus our analysis on the lowest income quartile, where we might expect

[28] Our results are robust to controlling for income instead of income groups as well. The results do not change qualitatively and experience no meaningful change quantitatively.

Variables	Food (1)	Non-durable (2)	Durable (3)	Total (4)
Single	-0.2301**	-0.1354*	-0.0992	-0.1137*
	(0.0912)	(0.0742)	(0.0844)	(0.0608)
Year FE	✓	✓	✓	✓
Y Group FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Observations	435	437	423	437
R^2	0.0755	0.0693	0.0601	0.0963

Table 16: This table shows estimations from a feasible generalized least squares regression on four consumption measures where we differentiate between single and married households. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Quartile 1	**Quartile 2**	**Quartile 3**	**Quartile 4**
Shocks	424	79	32	33

Table 17: This table shows the distribution of income shocks across the four income groups.

that both cognitive and structural frictions are most likely to bind following a negative income shock. To do this, we estimate Equation (9) without income group fixed effects while restricting our sample to the lowest income quartile. We show the results of this estimation in Table 18.

Table 16 provides suggests evidence that single households, regardless of income group, reduce food consumption by 23.01 pp ($p < .05$) more than married households following a negative income shock. Further, they reduce non-durable consumption by 13.54 pp ($p < .1$) and total consumption by 11.37 pp ($p < .1$) more than married households following a negative income shock. Finally, the result for the consumption of durable goods is qualitatively consistent, but not significant.

Table 18 reports our results when focusing on the first income quartile, which is the one income quartile where we have enough instances of negative income shocks to isolate the within-group effect of a negative income shock on single and married households. Single households in this income category reduce food consumption by 26.61 pp ($p < .05$) and non-durable consumption by 17.02 pp ($p < .1$) more than married households in the same income category following a negative income shock. Finally, the results for the consumption of durable goods and total consumption are qualitatively consistent but neither is significant.

Interestingly, we show that when we evaluate the effect of a negative income shock on the first income group in isolation, or on the entire sample, there is a significant, negative response for the consumption of food and non-durable goods. This suggests that difference in the way single and married households smooth their consumption of these goods is not restricted to the lowest income quartile in America.

Variables	Food (1)	Non-durable (2)	Durable (3)	Total (4)
Single	-0.2661**	-0.1702*	-0.0832	-0.0538
	(0.1217)	(0.0896)	(0.1103)	(0.0791)
Year FE	✓	✓	✓	✓
Controls	✓	✓	✓	✓
Observations	314	316	303	316
R^2	0.0812	0.0999	0.0411	0.0637

Table 18: This table shows estimations from a feasible generalized least squares regression on four consumption measures where we differentiate between single and married households in the first income group alone. Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

There is no clear way to say whether this is due to structural differences between single and married households or due to better decision-making by married households. The fact that we see this result so significantly for the lowest income quartile may suggest that it is in fact due to structural frictions. However, this result is robust to making comparisons within income class and controlling for available wealth balances. Additionally, the fact that this result holds in the aggregate may suggest that forming joint decisions could benefit all households rather than just the lowest-income households.

However, we can neither claim these are causal estimates nor cleanly distinguish the underlying mechanism. These seeming differences in how single and married households respond to negative income shocks could result from either cognitive and structural frictions.

7.5 Instructions for *Individuals*

Overview:

Welcome! You are here today to participate in an economic experiment involving the experimental simulation of an economy. If you read these instructions carefully and make appropriate decisions, you may earn a considerable amount of money that will be paid to you in cash immediately following the experiment.

We will pay each participant $10 for attending this experimental session. Throughout the experiment you can accrue additional earnings based on the decisions and predictions you make. You will earn points for each decision you make. Every 50 points you earn is worth an additional $1.

You are not allowed to communicate with other participants during this experiment. If you have any questions, the experimenter will be glad to answer them privately. If you have not done so already, please turn off your cell phone now. If you do not comply with these instructions, you will be excluded from the experiment and deprived of all payments aside from the minimum payment of $10 for attending.

Today's experiment consists of 2 sections.

Section 1 Instructions:

The first section has two parts. The first part of section one requires you to choose among a set of possible gambles. We will implement whichever gamble you choose and pay you based on the outcome of this gamble. The second part of section 1 requires you to answer a series of questions. We will pay you $.25 for each question you answer correctly. We will provide further instructions for section 1 on your screen whenever necessary.

The second section of today's experiment involves two 'sequences' of decision making. Each sequence consists of 20 periods. You will make a new decision in each of these periods. You will make these decisions using an experimental program displayed on the screen at your terminal. Your goal during the second section of today's experiment is to convert income into consumption points. Your income in this game is valueless until you convert it into consumption points. We will convert these consumption points into cash and pay you at the end of this experiment.

Section 2 Instructions:

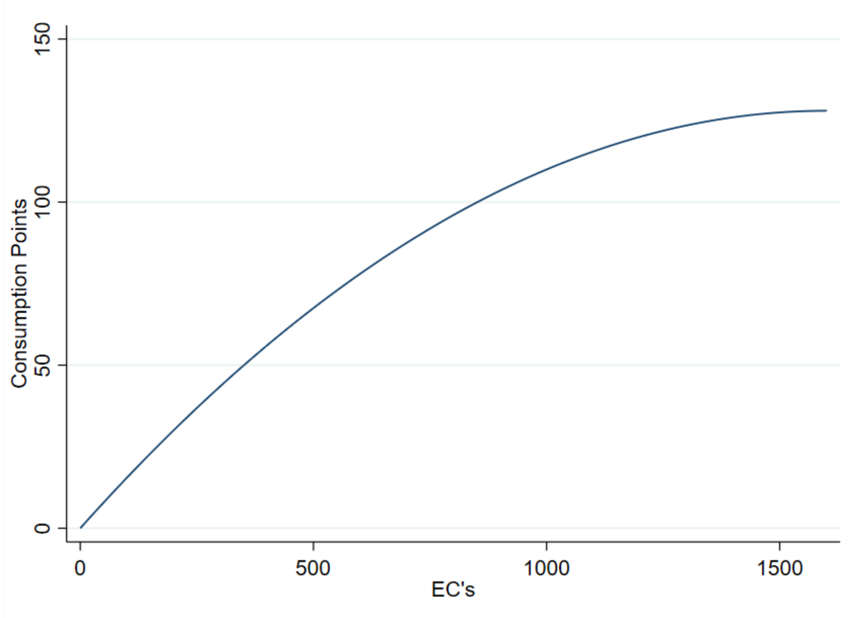
You are endowed with experimental credits (ECs) at the beginning of each period. We refer to these experimental credits as income. The amount of income you receive in each period is determined randomly and will always be an amount between 60 and 80, inclusively. You may receive as income any number of ECs between 60 and 80 with equal probability. Income in each period is independent of whatever income you received before.

After randomly determining your per-period income, the program will display this amount to you and deposit this money automatically into your bank account. The program will also display bank account balance (see Figure 1). This amount in your bank account represents your total wealth.

You must decide in each period how much of your total wealth to convert into consumption points for that period. You will earn points for consuming. Specifically, the number of points you earn in a single period is given by:

$$u(c_t) = [1600 * c_t - \frac{1}{2}c_t^2]\frac{1}{10,000}$$

Graphically:

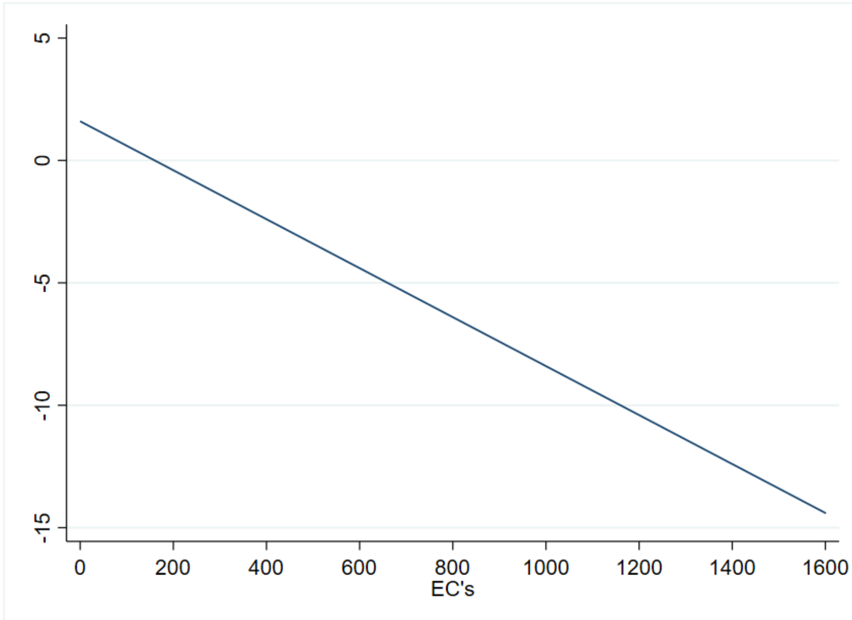


You can see from the graph above that each EC you spend on consumption (X-axis) earns a positive, but diminishing, number of consumption points (Y-axis). Each EC that you spend within a period will earn you less consumption points than the previous EC spent in the same period. This is known as diminishing marginal returns.

The rate at which you can convert wealth into consumption points is given by:

$$u'(c_t) = 1.6 - \frac{c_t}{100}$$

Graphically:



This graph shows you how many additional points you receive within a period (Y-axis) for spending a certain amount of wealth (X-axis).

ECs have no value in this experiment. Only consumption points have value. We convert consumption points to U.S. dollars at the rate of 50 points for $1.

Saving and borrow:

Saving:

You may save money in this experiment. Saving occurs automatically. If you spend an amount of ECs that is less than the amount of ECs in your bank account at the beginning of a period, this is called saving. Since we automatically deposit your per-period into your bank account and all of your available income is stored in your bank account, saving requires no additional actions.

Any wealth that you do not use in a period for consuming will remain in your bank account and will be available for consuming in later periods. Note that your bank balance does not earn interest. Any money left in your account at the end of the 20th period of a sequence becomes worthless.

Borrowing:

You may borrow up to 60 credits in all periods except the last period. You cannot borrow in the last period because you are not allowed to end this game with a negative balance. Borrowing is also straightforward. If you wish to borrow money for consumption, simply add the amount of money you wish to borrow for consumption to your consumption decision. The program will always allow you to spend (except in the final period) an amount equal to whatever is in your bank account at the beginning of a period plus 60 ECs.

Saving and Borrowing example:

Suppose you have 100 ECs in your bank account at the beginning of period 2:

1. Suppose you spend 75 ECs on consumption. Then your bank account balance at the end of period 2 will be 25 ECs. Your bank account balance at the start of period 3 will be 25 ECs plus whatever endowment you receive for period 3.

2. Suppose you decide you want to spend 130 ECs. To do this, simply submit 130 ECs as your consumption decision (we discuss how to do this later in instructions). The program will allow you to spend the 130ECs and your bank account balance at the end of period 2 will be -30 ECs. Your bank account balance at the beginning of period 3 will be -30 ECs plus whatever endowment you receive for period 3.

Making a consumption decision:

We discuss two things in this section of the instructions. First, we discuss a tool available to you that will aid your consumption decision. We call this tool the consumption calculator. Second, we discuss how to submit a consumption decision.

Consumption Calculator:

We provide you with a consumption calculator to assist you when making a consumption decision. This is shown in Figure 1 below.

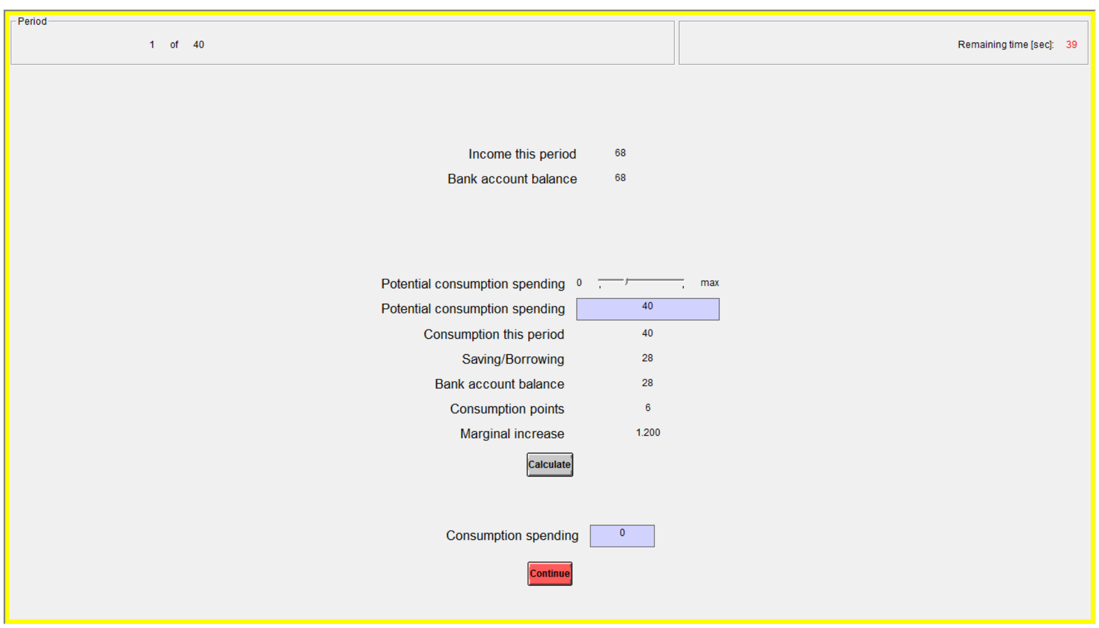


Figure 15: Decision screen for *Individuals* treatment.

The consumption calculator allows you to select a potential level of income you'd like to spend on consumption and shows you how much money you would save or borrow based on that decision, your resulting bank account balance, and the number of consumption points you would earn for spending that amount of income on consumption in that period.

You can choose a potential level of consumption income in two ways. First, you can move the slider (top line of the middle section of the screen in Figure 1) to some potential level of consumption spending. Doing this will cause all information to update automatically. Second, you can type a level of potential consumption spending into the box in the same section. Next, clicking the 'calculate' button in this section will cause all information to update based on whatever number you entered into the box.

Additionally, this calculator will show you the additional amount of consumption points you would earn if you decided to spend an additional EC in that period. This is called the marginal return to consumption. Recall, Each EC that you spend within a period will earn you less consumption points than the previous EC spent in the same period.

Information:

As shown in Figure 1 above, you will always have information about your current period endowment and bank account balance whenever making a consumption decision. Furthermore, you will always have the consumption calculator available to help you understand how a potential level of consumption spending would impact your earnings and change your available bank account balance for spending in future periods.

Additionally, we will complete each period (after you make a consumption decision) by providing a review screen that reminds you of how much income you spent on consumption in that period, your bank account balance at the end of that period, the amount of consumption points you earned in that period, and your total earnings. This is shown in Figure 2.

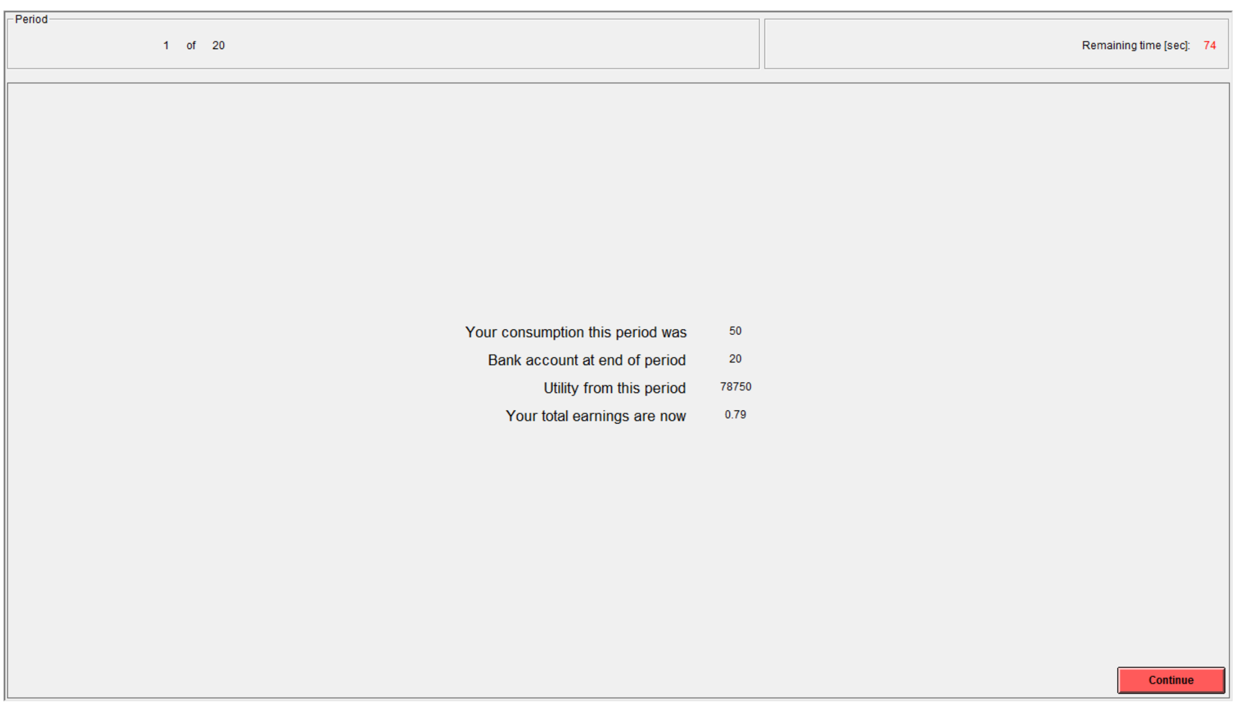


Figure 16: Review screen.

Once all subjects complete the first 20-period sequence, we will begin another 20-period sequence. The only difference between the first and second 20-period sequence is that the sequence of endowments (the income you receive at the beginning of each period) will be different. This is because the sequence is randomly drawn with equal probability from the closed interval of [60,80].

Payment:

Your payment today will consist of your $10 show-up fee, your earnings from the initial questionnaire (where you earn $.25 for each correct question), whatever you earn from your randomly implemented gamble, and your earnings from the two, 20-period sequences of decisions.

Questions?

Now is the time for questions. If you have a question, please raise your hand and the experimenter will answer your question in private.

Quiz:

Before continuing on to the experiment, we ask that you complete the following quiz. You can use the instructions to help answer these questions. *Your performance on this quiz does not affect your payoff.* Write or circle your answers to the quiz questions as indicated. Do not put your name on this quiz. If any questions are answered incorrectly, we will go over the relevant part of the instructions again.

1. In part one you will earn ______________ for each correct answer in the quiz. 2. Suppose it is period 5. Does the endowment you receive in period 5 depend on the endowment you received in period 4? ______________.

Does it instead depend on an endowment received in some earlier period (1, 2 or 3)? ______________.

3. Suppose you have 100 ECs in your bank account at the beginning of a period. Does this include your endowment for that period? ______________.

4. Suppose you have 100 ECs in your bank account at the beginning of a period. What is the maximum amount you can spend on consumption this period? ______________. What will be your bank account balance at the end of the period if you spend this maximum amount? ______________.

5. True or False: We will pay you for the decisions you make in both sequences? ______________

6. Suppose you earn 200 consumption points total. How much money do you earn? ______________.

7.6 Instructions for Pairs

<u>**Overview:**</u>

Welcome! You are here today to participate in an economic experiment involving the experimental simulation of an economy. If you read these instructions carefully and make appropriate decisions, you may earn a considerable amount of money that will be paid to you in cash immediately following the experiment.

We will pay each participant $10 for attending this experimental session. Throughout the experiment you can accrue additional earnings based on the decisions and predictions you make. You will earn points for each decision you make. Every 25 points you earn is worth an additional $1.

You are not allowed to communicate with other participants during this experiment. If you have any questions, the experimenter will be glad to answer them privately. If you have not done so already, please turn off your cell phone now. If you do not comply with these instructions, you will be excluded from the experiment and deprived of all payments aside from the minimum payment of $10 for attending.

Today's experiment consists of 3 sections.

Section 1 Instructions:

The first section has two parts. The first part of section one requires you to choose among a set of possible gambles. We will implement whichever gamble you choose and pay you based on the outcome of this gamble. The second part of section 1 will require you to answer a series of questions. We will pay you $.25 for each question you answer correctly. We will provide further instructions for section 1 on your screen whenever necessary.

The second section of today's experiment involves two 'sequences' of decision making. Each sequence consists of 20 periods. You will make a new decision in each of these periods. You will make these decisions using an experimental program displayed on the screen at your terminal. Your goal during the second section of today's experiment is to convert income into consumption points. Your income in this game is valueless until you convert it into consumption points. We will convert these consumption points into cash and pay you at the end of this experiment.

You will make your consumption decisions in each period with a partner. We will randomly assign you a partner during this experiment. You will be able to communicate with your partner using a chat feature. Your partners are fixed for the entirety of this experiment. That is, you will work with the same partner for both 20-period sequences.

The third section again requires you to choose among a set of possible gambles. However, you will be working with the same partner to make this decision. You will be able to communicate with your partner using a chat feature. We will implement whichever gamble you and your partner choose and pay you based on the outcome of this gamble. We will provide further instructions for section 3 on your screen whenever necessary.

Section 2 Instructions:

You and your partner are jointly endowed with experimental credits (ECs) at the beginning of each period. We refer to these ECs as income. The amount of income you and your partner receive in each period is determined randomly and will always be an amount between 60 and 80, inclusively. You may receive as income any number of ECs between 60 and 80 with equal probability. Income in each period is independent of whatever income you received before.

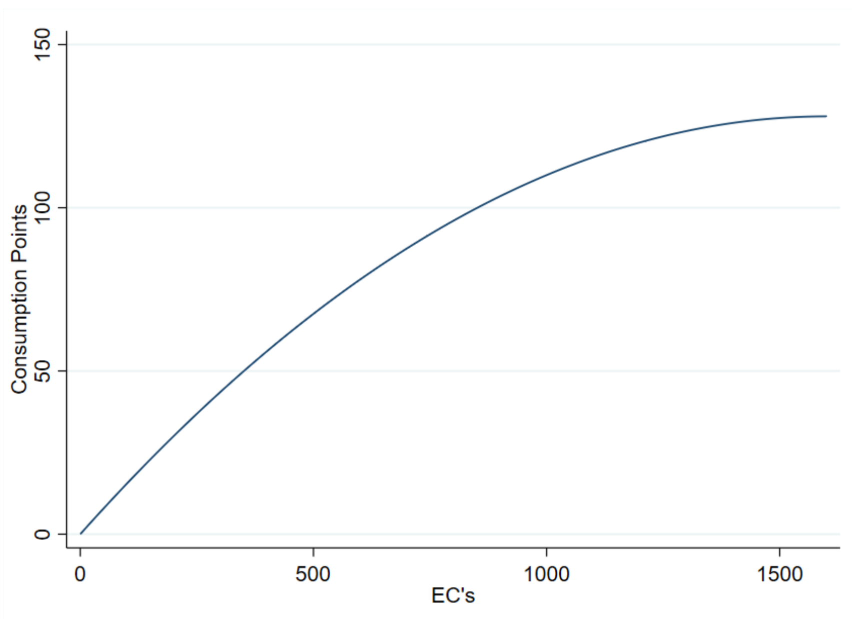
After randomly determining you and your partner's joint per-period income, the program will display this amount to you both and deposit this money automatically into your joint bank account. The program will also display the joint bank account balance (see Figure 1). This amount in your bank account represents your total wealth.

For example, suppose your joint endowment for a period is 70 ECs. You and your partner will both see this number. This means that together you must decide how to spend use these 70 ECs. To be clear, this would not mean that you have jointly gained 140 ECs.

You and your partner must decide in each period how much of your total wealth to convert into consumption points that period. Specifically, the number of points you and your partner earn in a single period is given by:

$$u(c_t) = [1600 * c_t - \frac{1}{2}c_t^2]\frac{1}{10,000}$$

Graphically:

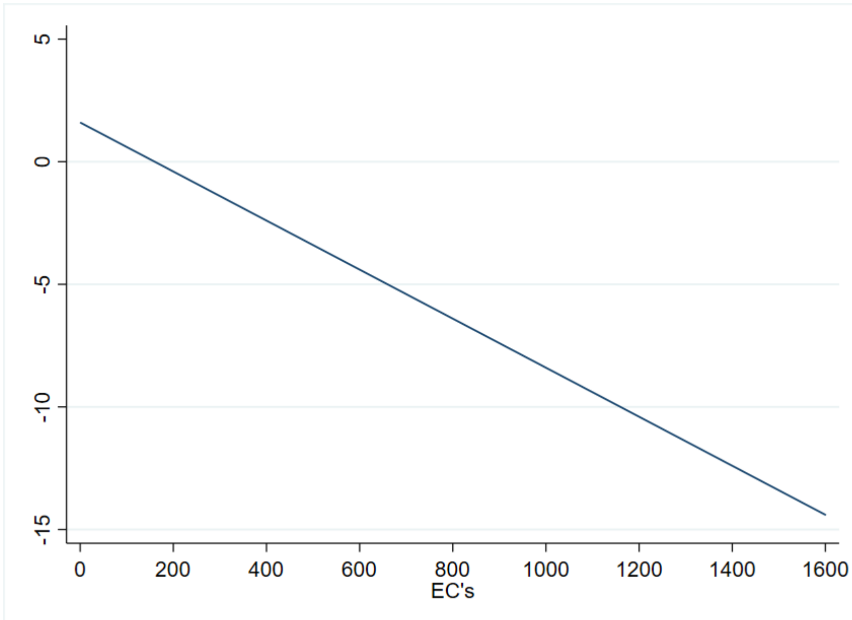


You can see from the graph above that each EC spent on consumption (X-axis) earns a positive, but diminishing, number of consumption points (Y-axis). Each EC that you spend within a period will earn you less consumption points than the previous EC spent in the same period. This is known as diminishing marginal returns.

Specifically, the rate at which you can convert wealth into consumption points is given by:

$$u'(c_t) = 1.6 - \frac{c_t}{100}$$

Graphically:



This graph shows you how many additional points you receive within a period (Y-axis) for spending a certain amount of wealth (X-axis).

ECs have no value in this experiment. Only consumption points have value. We convert consumption points to U.S. dollars at the rate of 25 points for $1.

You and your partner will splint income evenly. For example, if your joint consumption decisions lead to a payoff of $25 total, then you both receive $12.50.

Saving and borrow:

Saving:

You may save money in this experiment. Saving occurs automatically. If you spend an amount of ECs that is less than the amount of ECs in your bank account at the beginning of a period, this is called saving. Since we automatically deposit your per-period income into your bank account and all of your available income is stored in your bank account, saving requires no additional actions.

Any wealth that you do not use in a period for consuming will remain in your bank account and will be available for consuming in later periods. Note that your bank balance does not earn interest. Any money left in your account at the end of the 20th period of a sequence becomes worthless.

Borrowing:

You may borrow up to 60 credits in all periods except the last period. You cannot borrow in the last period because you are not allowed to end this game with a negative bank account balance.

Borrowing is also straightforward. If you wish to borrow money for consumption, simply add the amount of money you wish to borrow for consumption to your consumption decision. The program will always allow you to spend (except in the final period) an amount equal to whatever is in your bank account at the beginning of a period plus 60 ECs.

Saving and Borrowing example:

Suppose you have 100 ECs in your bank account at the beginning of period 2:

1. Suppose you spend 75 ECs on consumption. Then your bank account balance at the end of period 2 will be 25 ECs. Your bank account balance at the start of period 3 will be 25 ECs plus whatever endowment you receive for period 3. 2. Suppose you decide you want to spend 130 ECs. To do this, simply submit 130 ECs as your consumption decision (we discuss how to do this later in instructions). The program will allow you to spend the 130ECs and your bank account balance at the end of period 2 will be -30 ECs. Your bank account balance at the beginning of period 3 will be -30 ECs plus whatever endowment you receive for period 3.

Making a consumption decision:

We discuss two things in this section of the instructions. First, we discuss a tool available to you and your partner that will aid your consumption decision. We call this tool the consumption calculator. Second, we discuss how to submit a consumption decision.

Consumption Calculator:

We provide you with a consumption calculator to assist you when making a consumption decision. This is shown in Figure 1 below.

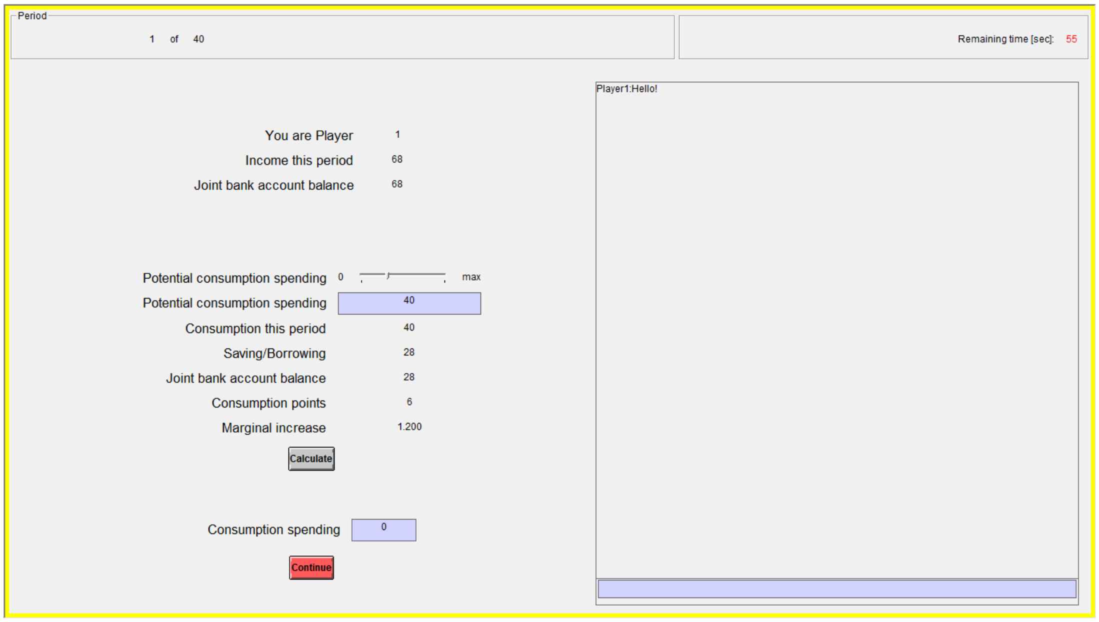


Figure 17: Decision screen for Pairs treatment.

The consumption calculator allows you to select a potential level of income you'd like to spend on consumption and shows you how much money you would save or borrow based on that decision, your resulting bank account balance, and the number of consumption points you would earn for spending that amount of income on consumption in that period.

You can choose a potential level of consumption income in two ways. First, you can move the slider (top line of the middle section of the screen in Figure 1) to some potential level of consumption spending. Doing this will cause all information to update automatically. Second, you can type a level of potential consumption spending into the box in the same section. Next, clicking the 'calculate' button in this section will cause all information to update based on whatever number you entered into the box.

Additionally, this calculator will show you the additional amount of consumption points you would earn if you decided to spend an additional EC in that period. This is called the marginal return to consumption. Recall, Each EC that you spend within a period will earn you less consumption points than the previous EC spent in the same period.

Both you and your partner have *independent* consumption calculators. This means that your partner does not automatically see information for potential levels of consumption spending that you check using your calculator and vice versa.

You and your partner can chat freely using the chat box pictured on the right side of Figure 1. You should use this chat box to jointly agree upon a decision about how much of your joint income you should spend on consumption in each period.

Once you have reached an agreement using the chat box, you should both input your consumption spending decision and click continue. If you both input the same number, the program will proceed and you will jointly earn whatever amount of consumption points corresponds to your joint decision. If the numbers do not match, the program will not continue forward. You will receive a notification from the program whenever you input a number that does not match your partners.

Information:

As shown in Figure 1 above, you will always have information about your current period endowment and bank account balance whenever making a consumption decision. Furthermore, you will always have the consumption calculator available to help you understand how a potential level of consumption spending would impact your earnings and change your available bank account balance for spending in future periods. This is shown in Figure 2.

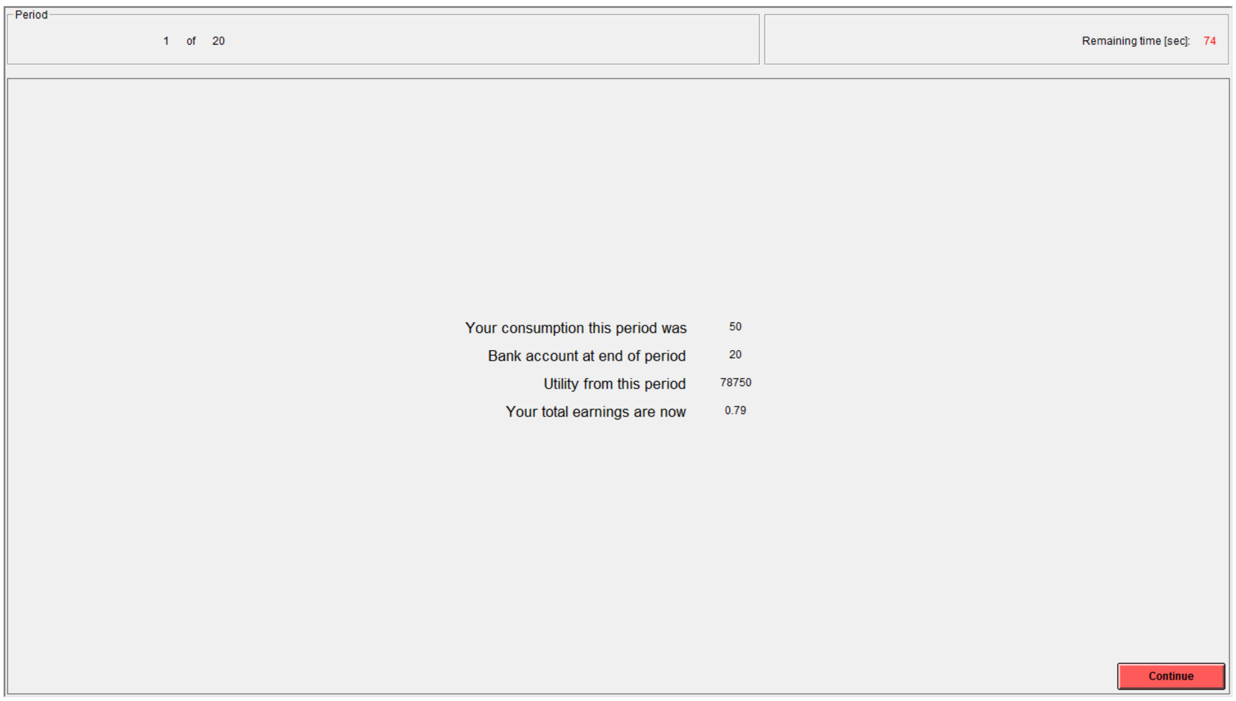


Figure 18: Review screen.

Additionally, we will complete each period (after you make a consumption decision) by providing a review screen that reminds you of how much income you spent on consumption in that period, your bank account balance at the end of that period, the amount of consumption points you earned in that period, and your total consumption points.

Once all subjects complete the first 20-period sequence, we will begin another 20-period sequence. The only difference between the first and second 20-period sequence is that the sequence of endowments (the income you receive at the beginning of each period) will be different. This is because the sequence is randomly drawn with equal probability from the closed interval of [60,80].

Payment:

Your payment today will consist of your $10 show-up fee, your earnings from the initial questionnaire (where you earn $.25 for each correct question), whatever you earn from both of your randomly implemented gamble, and your earnings from the two, 20-period sequences of decisions.

Questions?

Now is the time for questions. If you have a question, please raise your hand and the experimenter will answer your question in private.

Quiz:

Before continuing on to the experiment, we ask that you complete the following quiz. You can use the instructions to help answer these questions. Your performance on this quiz does not affect your payoff. Write or circle your answers to the quiz questions as indicated. Do not put your name on this quiz. If any questions are answered incorrectly, we will go over the relevant part of the instructions again.

1. In part one you will earn ______________ for each correct answer in the quiz.

2. Suppose it is period 5. Does the endowment you receive in period 5 depend on the endowment you received in period 4? ______________. Does it instead depend on an endowment received in some earlier period (1, 2 or 3)? ______________.

3. Suppose you have 100 ECs in your bank account at the beginning of a period. Does this include your endowment for that period? ______________.

4. Suppose you have 100 ECs in your bank account at the beginning of a period. What is the maximum amount you can spend on consumption this period? ______________. What will be your bank account balance at the end of the period if you spend this maximum amount? ______________.

5. True or False: We will pay you for the decisions you make in both sequences? True False.

6. If you and your partner together earn $30 for your joint consumption decisions, then you will personally earn how much? ______________.

7. Suppose you earn 200 consumption points total. How much money do you and your partner earn? ______________.

8. Does the marginal increase from an EC spent within a period earn you more or less consumption points than the previous EC spent in the same period?

7.7 Other Materials

CRT Questions:

1. A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost in cents?

2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets, in minutes?

3. In a lake there is a patch of lily pads. Every day, the patch doubles in size. If it takes

48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake, in days?

4. In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?

5. If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?

6. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class?

Demographics Survey:

1. Select your gender. (Male, Female, Other?)

2. What is your age?

3. Which year in school are you? (Freshman, Sophomore, Junior, Senior, Graduate)

4. What is your major?

5. To the best of your knowledge, what is your GPA?

6. Approximately how much student loan debt do you have?

7. Approximately how much other debt do you have?

8. What income class were you in growing up, e.g. lower, middle, upper?

9. What is your current political affiliation?

Survey of Decisions:

1. What information did you use in making your consumption decisions?

2. Did you have a decision rule, if so, what was it?

3. Did you feel like you had enough time to make your decisions?

4. Do you believe it is good or bad to have debt?

5. Do you believe it is good or bad to have savings?

6. How well do you believe you performed on the consumption task? 25th percentile? 50th percentile? 75th percentile? 99th percentile?

Extra survey of decisons questions for Pairs treatment:

7. What was your communication strategy with your partner?

8. Did you tend to agree or disagree with your partner?

9. What was your strategy for overcoming disagreements?