PREFERENCE FOR FLEXIBILITY AND RANDOM CHOICE: AN EXPERIMENTAL ANALYSIS*

Mark Dean and John McNeill†

Abstract

People may be uncertain about future preferences, leading to both a preference for flexibility in choice between menus and stochastic choice from menus. We describe an experimental test of preference uncertainty in a real-effort task. We observe subjects' preferences over menus of work contracts, and choice of effort from those contracts. We find that preference uncertainty is important: 48% of subjects exhibited preference for flexibility. A model of preference uncertainty accurately describes the relationship between choice of and from menus. Introducing a stochastic element to work contracts increased preference for flexibility, suggesting a causal role uncertainty in menu preferences.[1][2]

JEL codes: D81, C91

Preferences may be both unstable and unpredictable. Some days you may be energetic and willing to put in a lot of work in return for more pay. On others, you may feel lazy and sluggish and be much less willing to work hard. Moreover, it may be difficult to predict in advance in which of these states you will find yourself on a particular day. Such preference uncertainty may lead to a 'preference for flexibility': a decision maker (DM) who is uncertain about their future cost of

*The authors gratefully acknowledge support from NSF Grant NSF 1156090, and thank Marina Agranov, Pedro dal Bó, Pietro Ortoleva and participants at the Brown Theory Lunch for helpful suggestions.

†*Dean: Department of Economics, Brown University, Providence RI 02912. Telephone: 401-863-2097. Fax: 401-863-1970. Email: mark_dean@brown.edu*

[1]This experiment received IRB approval from Brown University (Protocol #1301000749) on March 05, 2013.

[2]Word count: 10795

effort may want to have a number of different output/payment combinations available to choose from when the uncertainty resolves itself. Preference uncertainty has been heavily studied in the decision theory literature,[3] and is of significant practical importance, for example in the design of labor contracts. Recent work[4] has also highlighted the role of preference uncertainty in offsetting any preference for commitment arising from time inconsistency due to problems of temptation.

This paper experimentally explores the extent to which preference uncertainty is an important component of behavior in a real-effort task. We first measure the extent to which workers exhibit preference for flexibility in choice of labor contracts. Next, we ask whether observed preference for flexibility can be explained as a rational response to preference uncertainty, by testing whether preference over contracts is linked to choice from those contracts in the manner predicted by Ahn and Sarver [2013].[5] Finally, we test the comparative statics of the preference uncertainty model by introducing an explicit stochastic element to the labor contract, which should lead to an increase in preference for flexibility.

Subjects in our experiment were offered the chance to perform simple arithmetic tasks for money. The experiment was run using Amazon's Mechanical Turk platform, an Internet marketplace through which people are hired to perform small tasks that are not easily automated, such as podcast transcribing or image tagging. A significant advantage of using the Mechanical Turk platform is the close relationship between the tasks involved in our experiment and the tasks that these workers usually perform, meaning that we classify our study as a 'quasi-field' experiment.

Subjects' payments were governed by an employment contract which described how much they would earn conditional on achieving a target number of tasks. A 'Low' contract had a low target and low payment, a 'High' contract a high target and high payment, and a 'Flex' contract both low and high targets, with the corresponding payments. Subjects first worked under exogenously specified contracts, before being offered the chance to choose between different contracts for subsequent work sessions.

[3] See for example Kreps [1979], Dekel, Lipman and Rustichini [2001] and Ahn and Sarver [2013].

[4] Amador, Werning and Angeletos [2006].

[5] This model combines the preference for flexibility model of Dekel et al. [2007] with the stochastic choice model of Gul and Pesendorfer [2006].

Our first result is that subjects exhibited a significant degree of preference for flexibility. 35% of well-behaved subjects[6] exhibited a strict preference for flexibility (i.e. were prepared to pay for the Flex contract over both the Low and High contracts) in our baseline specification, and 48% exhibited such preferences for some parameter values. Moreover, 94% of subjects demonstrated preferences in line with set monotonicity (i.e. never strictly prefer a smaller set), a necessary condition for preference uncertainty to explain menu preferences (Dekel, Lipman and Rustichini [2001]; Ahn and Sarver [2013]).

Our second result is that the subjects' effort choices from contracts were largely consistent with their preferences between contracts, in line with the model of Ahn and Sarver [2013]. This provides evidence that the preference for flexibility we observe is related to preference uncertainty rather than, for example, a heuristic preference for larger choice sets, or a failure of rationality. We estimated stochastic choice functions for our population as a whole, and for subsets of subjects categorized by their preference between menus. Ahn and Sarver [2013] provides two conditions that are necessary for preference for flexibility and stochastic choice being driven by the same underlying preference uncertainty. The first is that a DM who is prepared to pay for an alternative to be included in his or her menu must use that alternative some of the time. The second is that a DM who uses an alternative must be prepared to pay to have it in his or her menu (except in the case of indifference). We find evidence for both of those effects: subjects who paid for the Flex contract over the Low contract were significantly more likely to use the high target, while those who paid for the Flex contract over the High contract were significantly more likely to use the low target. As a result, subjects with preference for flexibility made substantial use of both contracts. Moreover, 96% of non-indifferent subjects who used the low target strictly preferred the Flex contract to the High contract, while 83% of such subjects who used the high target strictly preferred the Flex contract to the Low contract.

Our final result is that an exogenous increase in uncertainty leads to an increase in preference

[6] We describe what we mean by well-behaved in section 3.1 - essentially it means subjects who avoided making dominated menu choices. Due to the noisy nature of Mechanical Turk we have a significant number of subjects - 48% - who failed this test. However, including these subjects actually increases the proportion who exhibited preference for flexibility in the baseline specification to 43%.

for flexibility. Preferences over contracts therefore respond to changes in the environment as predicted by a model of preference uncertainty. We demonstrate this effect in a further experiment which introduced a stochastic component to the number of tasks that needed to be completed for the High target. This uncertainty was resolved after the contract was chosen but before it was implemented. Thus, subjects did not know how many tasks they would have to complete for the High target at the time that they chose the contract, but did know when the contract was in effect. We find an increased preference for flexibility under such a 'random' contract, indicating that subjects take the degree of uncertainty into account when making menu choices.

It is of interest to try to understand what it is that our subjects were uncertain about in the baseline (deterministic) experiment. Our favored interpretation is that it was their relative valuation of effort and monetary reward at the time they must perform the task. In Section 5.2 we argue against an alternative explanation - that subjects were uncertain about their ability to complete the task within the allotted time. Only a small number of subjects were in danger of running out of time, and those subjects were somewhat *less* likely to exhibit preference for flexibility than those who had plenty of time.

To our knowledge our paper is the first to experimentally document and study preference for flexibility in menu choice.[7] We are also the first to link such preferences to subsequent choice from menus. This link helps to identify preference uncertainty as a cause of preference for larger menus, a mechanism that has been posited since at least Kreps [1979], but never tested. In doing so, we also provide an explicit test of the model of Ahn and Sarver [2013].

Our work also has relevance for contract design. The large fraction of subjects in our experiment who express preferences for flexibility in their choice of work arrangements are likely to dislike restrictive labor contracts which (for example) include fixed quotas or hard deadlines. This is of particular interest in light of the large theoretical and practical literature relating to temptation and self-control. Sophisticated agents who have self-control problems may exhibit preference for

[7] By 'flexibility' we mean that the union of two menus is preferred to either of the underlying menus. Other work (e.g. Sonsino and Mandelbaum [2001]) has documented the fact that people may pay for larger menus in line with standard theory.

commitment (i.e. a preference for smaller choice sets), yet there is limited evidence of demand for commitment devices.[8] Preference uncertainty could be one possible explanation for this. As pointed out in Amador, Werning and Angeletos [2006], preference uncertainty provides an offsetting preference for flexibility, making commitment costly. Our results suggest that it may be important to take into account preference uncertainty when designing commitment contracts. We discuss this issue further in Section 5.1.

The remainder of the paper is organized as follows. Section 1 describes the theoretical background, 2 outlines the experimental design, 3 our data and identification strategies, and 4 our results. Section 5 discusses the implications of our findings and the source of subject uncertainty. Section 6 describes the related literature, and 7 concludes.

1 Model

Our analysis is based around a model of behavior in which a decision maker faces uncertainty about his or her future preferences. This can lead to both a preference for flexibility (in the manner of Dekel et al. [2007], henceforth DLRS) and stochastic choice (in the manner of Gul and Pesendorfer [2006], henceforth GP). Ahn and Sarver [2013] (henceforth AS) provides conditions under which the same uncertainty can be seen as driving both preference for flexibility and stochastic choice.

The building blocks of the ‘preference uncertainty’ model described by AS are a set of alternatives Z, lotteries over these alternatives $\Delta(Z)$, and menus of such lotteries A, with $\mathscr{A}$ denoting the set of such menus. The DM has preferences over $\Delta(Z)$ which depend on the (unobserved by the researcher) realization of a subjective state $s \in S$, and are modeled via a state dependent utility function $u : S \times \Delta(Z) \to \mathbb{R}$. The DM has beliefs about the likelihood of states denoted by the probability distribution $\mu \in \Delta(S)$.

This model can be used to understand the DM’s preference over menus (which we represent with the complete preference relation $\succeq$ on $\mathscr{A}$). The assumption is that uncertainty about S resolves

[8] Although recent work by Kaur, Kremer and Mullainathan [2010] does find preference for commitment in a setting which shares some features with ours.

itself after the DM chooses a menu, but before he or she chooses from that menu. Thus menu preferences are represented by a utility function defined for every A by:

$$U(A) = \sum_{S} \mu(s) \cdot \max_{x \in A} u(s,x) \tag{1}$$

In words, the utility of a menu A is equal to the expected utility of the best option in A, with expectations taken over the different possible utility functions indexed by the state S.

The same model can also be used to describe the distribution of choices from a given menu. Using $\lambda : \mathscr{A} \to \Delta(\Delta(Z))$ to denote the probability of choosing each alternative from each menu, the model predicts that[9]

$$\lambda^A(x) = \sum_{s \in S} \mu(s) \mathbf{1}[x \in \arg\max_{x \in A} u(s,x)] \tag{2}$$

The probability that an option x is chosen from a set A is equal to the probability of the states in which x is the highest utility object in A, as captured by the indicator function in Equation 2.

AS provides conditions under which a DM's preferences over menus are consistent with his or her (stochastic) choice from menus: in other words, the same beliefs and state-dependent utility function can be used to represent a DM's preferences over menus (in the sense of Equation 1) and stochastic choice function (in the sense of Equation 2).

Of particular interest to our study are four behavioral implications of the preference uncertainty model. The first is related to choice over menus, the second to choice from menus, and the third and fourth to the relationship between these two behaviors:[10]

[9] Subject to the tie-breaking rule τ, which GP also address as another potential source of randomness.

[10] The AS representation requires all of the axioms used in the construction of the DLRS and GP representations, some of which we do not seek to test in our experiment. DLRS requires that preferences over menus be a weak order (DLRS1) and nontrivial (DLRS3) - we did not test these directly, but do discuss the fraction of subjects whose choices were intransitive and those who displayed strict preferences over menus. Implicitly we test the axioms of AS conditional on menu preferences being a weak order. The requirements of Lipschitz continuity on preferences over menus (DLRS2), and continuity of random choice rules in mixtures over decision problems (GP 'mixture continuity') have no empirical content in our data set. DLRS also requires 'weak independence' as an independence property on menu preferences, and although we do observe preferences over submenus, we did not mix them with other menus in a way that allows us to address failures of weak independence; similarly GP requires a form of independence ('linearity') that we did not address with this experimental framework, as we did not observe choices over lotteries - the uncertainty in our Random Contracts treatment was resolved before subjects chose from the menu. GP's 'extreme' property requires that the random choice rule selects a maximizer of some utility function with probability 1. In our set up, this is equivalent to never choosing a dominated number of tasks to complete. See Section 3.2.1 for a discussion

1. **Preference for Flexibility:** Consider two menus A and B such that $A \succeq B$. If future preferences are certain (for example if S is a singleton) then Equation 1 would imply $A \cup B \sim A$. Yet if S is not a singleton, it is possible that $A \cup B \succ A$. This is what is referred to as strict 'Preference for Flexibility', and results directly from preference uncertainty. Moreover, the preference uncertainty model rules out a strict preference for smaller choice sets, so $A \subseteq B \Longrightarrow B \succeq A$. This property distinguishes the preference uncertainty model from models of temptation and self-control (for example Gul and Pesendorfer [2001]) in which smaller choice sets may be preferred to avoid exposure to tempting options.

2. **Choice Monotonicity:** When applied to choice from menus, the preference uncertainty model is essentially a random utility model, and shares the implication of choice monotonicity: adding options to a choice set cannot increase the probability that an existing option will be chosen - i.e. $x \in A \subset A' \implies \lambda^{A'}(x) \leq \lambda^{A}(x)$. This is in contrast to attention-based models of stochastic choice (for example Caplin and Dean [2014]), in which new alternatives can alter attention in such a way to increase the likelihood of choosing a previously available option.

3. **Consequentialism:**[11] Consider a DM who strictly prefers adding an alternative x to menu A. According to Equation 1, this can only be the case if x has higher utility than all the elements of A in some state that occurs with non-zero probability. Thus if the same preference uncertainty is to explain stochastic choice, it must be that alternative x is chosen from the menu $A \cup \{x\}$ in some state. This is this empirical content of the consequentialism axiom of AS:
$$A \cup \{x\} \succ A \implies \lambda^{A \cup \{x\}}(x) > 0$$

4. **Responsive Menu Preferences:**[12] The natural counterpart to the consequentialism condition is that if x is sometimes chosen from the menu $A \cup \{x\}$, then this menu should be strictly

of this issue.

[11]This is Axiom 1 of AS.

[12]This is Axiom 2 of AS.

preferred to A. However, this is too strong, as it could be the case that the DM is indifferent between x and another element of A in the state in which it was chosen, and its selection was as a result of a tie-breaking rule. Thus, the fourth condition is that $\lambda^{A \cup \{x\}}(x) > 0$ implies $A \cup \{x\} \succ A$ except in cases of indifference. AS use continuity conditions to behaviorally rule out indifference. We discuss how we deal with this issue in section 4.2.3.

2 Experimental Design

Subjects were offered the opportunity to complete effort tasks for payment. Thus, final choices were made over effort/money pairs. Subjects chose from menus of such pairs - i.e. contracts that specified payment for the number of tasks completed. We observed subjects' choice of effort/money pairs from different contracts and their preferences over contracts, allowing us to test the predictions discussed in the previous section. Implicitly, preference uncertainty would relate to the relative value of time and effort to money at the point at which the task was to be completed.

The experiment was conducted using Amazon's Mechanical Turk (MT) platform, a digital marketplace for work. "Reqesters" post Human Intelligence Tasks, or "HITs", which are usually simple, repetitive jobs that typically pay very small sums for each completed task. Workers on MT view descriptions of the HITs, decide which to accept, and complete those HITs over the Internet. In our case, subjects who accepted the HIT followed a link to an external webpage, where they completed the experiment. Upon completion they were given a randomly generated code, which was used to pay them the appropriate amount given their choices in the experiment.[13]

Mechanical Turk is a relatively new environment in which to conduct economic experiments.[14]

[13]The MT worker pool is global and diverse. It is possible to filter who can accept and complete a HIT according to different criteria, including geographic location and the lifetime HIT approval rating for the worker (on all HITs completed, not just those related to this task). As we had no a priori reason to restrict our workers to a specific geographic subset, we imposed only the requirement that the worker's HIT approval rating must be above 90%, a common requirement. We did not target any particular demographics of MT workers for our recruitment. The HIT was described as a "decision-making experiment", and used the keywords "decision", "experiment", "study", "bonus", and "payment".

Some workers accepted the HIT but did not complete the experiment - in the vast majority of those cases (81%), the subject exited the experiment prior to completing the instructions.

[14]For more discussion of MT and its use in experimental social sciences, see Paolacci, Chandler and Stern [2010],

One key advantage for the current study is that it represents a 'quasi-field' setting: subjects had signed up through the MT platform to supply labor in exchange for money in precisely the way that was on offer in our experiment. Thus for this study population the experimental setting is arguably more natural than for the typical undergraduate pool.

Further advantages of MT are the ready availability of subjects and the low cost of collecting data. The prevailing wage rate on MT is extremely low. The payments involved in our experiment were therefore low by the standard of traditional on-campus laboratory experiments, but commensurate with the prevailing wages on MT. One cost of using MT is a reduction in experimenter control: because MT workers complete the experiment remotely, the experimental environment is not as tightly controlled as it would be in the laboratory. Accordingly it is important to pay particular attention to data quality, as subject are potentially less focused on the task at hand than they would be in the laboratory. We address this issue in Section 3.1.

2.1 Main treatment

The four major components of the experiment were the tasks subjects could choose to complete to earn additional payment, the 'task sections' in which these tasks were performed, the contracts that specified how many tasks had to be completed in order to earn different payments in a task section, and multiple price list questions used to elicit subject preferences over contracts.

2.1.1 Tasks

The building block of the experiment was a real-effort task: a simple activity that the subject could choose to perform numerous times in order to earn additional payment. Each task was an arithmetic problem that required the subject to add together two three-digit numbers. After submitting their proposed solution to the problem the subject was told whether the submitted answer was correct or incorrect - correct answers counted towards his or her current total, while incorrect answers did not. An example of a typical task screenshot can be seen in Figure I.

Mason and Suri [2012], and Goodman, Cryder and Cheema [2013].

[Figure I approximately here]

2.1.2 Task Sections

Subjects completed several 'task sections': opportunities to perform tasks in exchange for money. Payment in each task section was governed by a contract which specified how many tasks needed to be completed in order to earn specified additional payments (see Section 2.1.3). Subjects could complete as many or as few tasks as they wished, but had to do so within a time limit.[15] Once they began a given task section, subjects had 15 minutes to complete as many tasks as they wished, although they could "retire" from the section and continue on with the experiment at any time. Thus, a task section ended whenever the timer ran out or the subject decided to stop. Subjects could not return to a previous task section once it had ended. For the contracts used in this experiment it did not benefit subjects to complete more than 50 tasks, although there was no strict upper limit on the number of tasks they could potentially complete. This means that subjects who wished to earn the highest possible reward needed to complete on average at least one task every 18 seconds. The intention of this time limit was to keep the subjects focused on the task at hand, rather than to make it difficult to complete the desired number of tasks. For most subjects, 18 seconds was ample time to complete a three-digit addition task with a moderate application of effort, but the time limit helped to discourage subjects from pausing in the middle of a section to do other things. In practice, subjects did not seem to encounter problems with the time limit, as we discuss in Section 5.2.

2.1.3 Contracts:

In each task section payment was governed by a contract which specified how many tasks must be completed in that section in order to receive an additional amount of payment, as shown in the example contracts shown in Figure II.

[Figure II approximately here]

[15] While completing tasks, subjects were shown the number of tasks they had completed correctly in that section, as well as the payment that that number of tasks had earned.

Under Contract 24 (Figure II), for example, completing fewer than 20 tasks would earn no additional compensation, completing between 20 and 49 tasks would earn \$0.20, and completing 50 or more tasks would earn a total of \$0.40 for that section. This was not a piece-rate arrangement, so the amount specified was the total payment earned for completing a given number of tasks. It would be possible to achieve something similar using a piece-rate wage, in the spirit of Kaur, Kremer and Mullainathan [2010]. However, this approach has the advantage of producing a relatively small set of sensible choices the subjects can make: given that effort is costly, subjects should either choose to do no tasks at all or to do enough tasks to reach a given payment level and then stop. This has the benefit of resulting (in principle) in data in which subjects' efforts are clustered at one of a small number of different completion levels, so that their actions can be condensed into one of a few types.

Each contract used in the experiment had either one or two levels at which the payment increased. In each task section, subjects could complete as many or as few tasks as they chose, which always included the option to complete no tasks for no additional payment. The other level(s) in a given contract were either a low number of tasks for a lower payment or a high number of tasks for a higher payment. The contracts used in the main experimental treatment were of three types. The "High" contract included only the option to do 50 tasks for \$0.40 (Contract 25 in of Figure II). The "Low" contracts offered only the option to receive \$0.20 for completing the low number of tasks. Low contracts are indexed by the number of tasks required to earn the \$0.20. Our main analysis focuses on the case in which 20 tasks were required to receive the low payment (Contract 11 of Figure II), which we will refer to as "L20"; the contracts that subjects experienced in the exogenous contracts sections required 20 tasks to earn the low payment. However we also study the effect of changing the number of tasks required to receive the low payment; a "Low" contract requiring X tasks will be referred to as "LX". The "Flex-X" (or "FX") contract combined the options available in the "High" contract and the "LX" contract, containing both the option to do X tasks for \$0.20 and the option to do the 50 tasks for \$0.40 (F20 is shown as Contract 24 in Figure II).

2.1.4 Contract Preference Questions

The fourth component of the experiment was a set of questions used to elicit subjects' preferences over different contracts. Subjects responded to a series of multiple price list (MPL) questions which asked them to choose between two different contracts for potential use in a future task section, as well as side payments. An example of such an MPL is shown in Figure III. Following these questions, one line from one question was selected at random, and the subject's choice on that line was implemented as the contract they used to complete a subsequent task section.

[Figure III approximately here]

In each MPL question the subject made a series of eleven pairwise choices between the two options on each line. On each subsequent line the option on the right became relatively more attractive, either because the side-payment associated with the left-hand option decreased or because the side payment associated with the right-hand option increased. Accordingly, if the subject chose the right-hand option on a given line, they should also have chosen the right-hand option on each subsequent line, which provides a check that the subject was paying attention to his or her choices. Any subject who switched from right-to-left either had preferences that are not monotonically increasing in money, or (more probably) was not paying attention to the question. In all cases, the side payments were as shown in Figure III, ranging from \$0.00 to \$0.50, with \$0.01 the smallest amount that the subject could pay for a contract.

Typical responses to the MPL questions begin on the left and at some point switch over to the right. The point at which the subject switched translates to the strength of preference for the contract on the left compared to the contract on the right - a subject who was willing to forgo an additional payment (for example, choosing "Contract 24" instead of "Contract 25 + \$0.01") must have strictly preferred the contract selected. A subject who was indifferent between the contract on the left-hand side and the contract on the right-hand side would always select the option with additional payment, and could choose either option in the line which has no additional payments.

2.1.5 Experimental Structure:

The experiment began with a set of instructions that introduced the tasks and contracts, including a comprehension quiz that required subjects to demonstrate that they understood how to read the contracts and determine how much payment would accompany a given number of tasks completed.[16] Following the instructions and comprehension quiz, subjects were required to complete four practice tasks, to ensure a baseline level of familiarity with the experimental interface before they began the main part of the experiment. Following this, they completed three task sections. In each of these sections the subject was exogenously assigned a different contract. The three contracts used were those shown in Figure II: L20, H and F20. The order of contracts was randomized between subjects.

After the three exogenous contract sections, subjects responded to nine questions eliciting their preferences over contracts to be used in a fourth and final task section. Subjects responded to MPLs covering bilateral comparisons between H, L20 and F20. These questions refer to the experienced contracts, and are the focus of our main analysis. In addition, subjects answered MPL questions between H and FX and between LX and FX for X=15, 10 and 5. The order of the nine questions was randomized across subjects.

After the subject completed the questions, one of the lines of one of the questions was randomly selected to be realized in a final task section. Note that the time lag between the choice of contract and the implementation of that contract was relatively short. We discuss this issue further in Section 5.2.

Following the final task section, the subject was paid according to his or her performance during the experiment through the MT payment interface. Subjects were paid for all four task sections, as well as for the selected line in the MPL questions, and received a participation fee of $0.25.

[16] A copy of the instructions is shown in Supplementary Appendix 1.

2.2 “Random Contracts” Treatment

Any preference for flexibility found in the main treatment is naturally-occurring. To test whether an increase in uncertainty would increase preference for flexibility we ran a further treatment which introduced exogenous uncertainty about some contract features, with the uncertainty resolved after the choice between menus but prior to the section in which the contract was implemented.

The random contract treatment worked as follows: for all Low and Flex contracts in this treatment, the “low” option was the same: $0.20 for completing 20 tasks. In High and Flex contracts, the “high” option paid $0.40, but the number of tasks required to earn the $0.40 varied between contracts: it was either “easy” (requiring 30 tasks), “hard” (requiring 70 tasks), or “random”. If the high option was ‘random’, then the contract involved a lottery which assigned the subject to either the ‘easy’ or the ‘hard’ parameters for the high option with equal probability. If the subject selected a random contract, then the number of tasks necessary for the high option was determined before it was implemented.

There were two kinds of random contracts used in this treatment: “High-random” which was a 50/50 lottery between the contract with only the ‘high-easy’ option and the contract with only the ‘high-hard’ option, and “Flex-random” which was a lottery between the Flex contract formed with the low option and the ‘high-easy’ option and the Flex contract composed of the low option and the ‘high-hard’ option.

As in the baseline, subjects completed three exogenous task sections, again in a random order. Two of these involved degenerate contracts: ‘Flex-easy’ and ‘Flex-hard’. The remaining task section used the the ‘Flex-random’ contract. For this section, the subject was informed that the contract they were using in this section involved randomization, and reminded that the computer would generate a random number which would determine which of the two subcontracts they would use, either the Flex-easy or the Flex-hard. On the next screen they were shown the random number generated by the computer, and reminded which contract they would be using in the task section.

[Figure IV approximately here]

Subjects then responded to MPL questions asking for their preferences between Flex and Low contracts and between Flex and High contracts for the three different types of Flex (and High) menus. Figure IV shows the choice between the High-random and the Flex-random contracts.[17] If an increase in uncertainty does lead to a preference for flexibility, we would expect to see more such behavior for the random contracts than for the easy or hard contracts.

As in the main treatment, one of these choices was actualized, and the subject completed the fourth and final task section using this contract. If the realized contract was one involving randomization, then the subject was reminded of their choice and that the computer would generate a random number to determine the contract to be used.

3 Data Overview

In total, 239 subjects participated in the main experimental treatment and 149 in the random contracts treatment, between November 9, 2013 and July 12, 2014. All subjects were recruited through Mechanical Turk, and completed the experiment over the Internet. Subjects earned an average of $1.35, including a participation fee of $0.25.

3.1 Data Quality

While all experimental data contains a certain amount of noise, the use of Mechanical Turk makes our data particularly likely to contain uninformative responses. The prevailing norm on MT is to perform small and repetitive tasks for small sums of money, and while this prepares the subjects well for the tasks used in this experiment, it also potentially rewards a meta-strategy of not exerting too much effort on any given HIT, instead focusing on completing a larger volume of HITs in a given time. This accordingly increases the concern that the subjects were clicking through the MPL questions without serious consideration of the choices being made. Because of this, we

[17] That is, the MPLs compared the Low/Flex and High/Flex contracts for each of the easy, hard, and random types. The order of presentation was randomized between subjects.

are particularly interested in identifying subjects who were not paying attention when choosing between contracts, so their data can be excluded from further analysis.

To screen these subjects from our data, the MPL method provides two potential tools. The first, as previously mentioned, is that the right-hand side option becomes more attractive relative to the left-hand side option as one goes down the list. This means that subjects' choices should never cross from right to left. The second is that the side payments used on the first and last lines of each question are sufficiently large to overwhelm any marginal earnings from having one contract over another. The additional payments on the top and bottom lines of each question are \$0.50, which is more than can be made with any given contract under consideration, and so the \$0.50 payment should always dominate the contracts in question. Accordingly, subjects should always select the left-hand option on the first line, and the right-hand option on the last line of the question, with exactly one crossover from left to right.

In the main experimental treatment we used these checks to exclude subjects based on their responses to the three MPL questions related to the contracts they experienced during the exogenous contracts section (L20/H, L20/F20, H/F20). In these questions 89% of subjects never switched from right to left, while 56% always selected the left option on the first line and the the right option in the last line. Excluding subjects that fail one of these checks left 124 subjects or 52% of the original. While this rate of exclusion is high relative to most laboratory experiments, it is unsurprising given the small stakes and noise inherent in the MT environment. Using equivalent criteria[18] we retained 61 subjects (41%) from the random contracts treatment. It should also be noted that our screening procedure in fact reduces the amount of preference for flexibility we observe: 43% of all subjects exhibited strict preference for flexibility amongst H, F20 and L20 contracts, compared to 35% of subjects who survived screening on these three questions.

[18]Because the analysis of the random contracts treatments require answers to all six questions asked, these subjects are screened on six rather than three questions, leading to a somewhat higher exclusion rate.

3.2 Identifying Menu Preferences and Stochastic Choice

We now discuss how we use the data generated by the experiment to identify the menu preferences and stochastic choice functions of our subjects.

3.2.1 Menu Preference

It is useful to interpret the contracts in our experiment as being menus of options defined by the tuple (n, p) where n refers to number of tasks completed and p to the monetary payments received. All menus contain the option $(0, 0)$: i.e. to do zero tasks for no payment. LX and FX menus additionally contain the option $(X, 0.20)$. H and FX menus additionally contain the option $(50, 0.40)$. In this sense, the FX contract is the union of the LX and H menus.[19]

We can identify preferences between menus using the MPL questions. We say that Contract A is weakly preferred to Contract B if it is chosen in the case of no side payments. Similarly, we identify strict preferences using the assumption that a subject who is prepared to pay a positive amount to have Contract A rather than Contract B must strictly prefer A to B.

Note that our MPL allows us only to identify strict preferences in which the difference in value between the two options is greater than \$0.01. Preferences that are weaker than this will be classified as indifference. Thus, our estimates provide a lower bound on strict preference for flexibility.

3.2.2 Identification of Stochastic Choice

In a given task section the menu of possible task/payment options was governed by the contract in effect in that task section. Subject choice from a menu was measured by the number of tasks they choose to complete.[20] We estimated choices from menus using data from the task sections in

[19]This is true if one considers non-dominated options. The LX menu contains the option to do the high number of tasks for the low reward (i.e. (50, \$0.20)), whereas the FX menu does not. However all options other than the three described above are strictly dominated, in the sense that the subject could exert less effort for the same reward. We will restrict our attention to the three non-dominated options described above.

[20]Choices were measured as the highest non-dominated number of tasks completed. So, for example, in an F20 contract, subjects who completed 25 tasks were classed as choosing to complete 20. Over all subjects and all task sections, subjects completed an average of 4.6 ‘unnecessary’ tasks per section. In many cases, these ‘unnecessary’

which the contracts are exogenously determined.

In practical terms, it was not possible to collect enough data to estimate subject-level stochastic choice functions. Completing the real-effort task is costly in terms of both time and effort, and so in our design we only observe one choice from a given contract per subject. We therefore estimated stochastic choice functions by pooling data across subjects. To do so we grouped subjects into equivalence classes based on their expressed preferences over menus. Details on how the subjects were pooled are given in Section 4. Notice that this means we cannot discuss whether an individual subject satisfies choice monotonicity, consequentialism and responsive menu preferences - only whether a class of subjects does so.

4 Results

4.1 Preference for Flexibility

Uncertainty about the relative cost of time and effort in future task sections could lead our subjects to exhibit a strict preference for flexibility, as discussed in Section 1. In our experiment a subject exhibited strict preference for flexibility if they strictly preferred the FX contract to both the LX contract and the H contract. The preference uncertainty model also requires set monotonicity, which implies that the FX contract must be weakly preferred to both the LX and the H contracts.

We used the data from the pairwise choices between F20, L20 and H to group subjects into five categories based on their revealed preferences:

1. Preference for Flexibility: $F20 \succ L20$ and $F20 \succ H$

2. Standard: Either $F20 \succ H$ and $L20 \succ H$ with no strict preference between $F20$ and $L20$, or

tasks occurred in the Flex contracts, with subjects performing more tasks than were necessary to achieve the lower target but not enough to achieve the high target. This is rationalizable if the subject had a change in preference during the course of the task. If instead we focus on truly unnecessary tasks, where the subject performed more tasks than was necessary to obtain maximum payment, we see an average of 3.2 tasks per subject. However, the distribution is heavily right-skewed: the median average number of excess tasks is 0.5. Excluding subjects who do many excess tasks does not materially change our results: preference for flexibility amongst those below the 75th percentile of excess tasks (3.1) is 39%, while amongst those that never perform an excess task (40 subjects) it is 33%.

$F20 \succ L20$ and $H \succ L20$ and no strict preference between $F20$ and H

3. Indifferent: No strict preference between any contracts

4. Commitment: Either $H \succ F20$ or $L20 \succ F20$, with no intransitivity

5. Intransitive: ($X \succ or \succeq Y$ and $Y \succ Z$) or ($X \succ Y$ and $Y \succ or \succeq Z$) but not $X \succ Z$, for some combination of menus X, Y, and Z

Appendix A provides a table which lists the categorization for all non-intransitive preference profiles.

The preferences of 'Standard' and 'Indifferent' subjects could be explained by a model of menu choice in which the DM has no preference uncertainty - i.e. by a model of the form of Equation 1 in which there is only one state (see for example Kreps [1979]). The 'Preference for Flexibility', or 'PFF' subjects can only be explained by this model if there is preference uncertainty. 'Commitment' subjects cannot be explained in the framework of Equation 1, although they could be explained by a model of temptation and self-control (for example Gul and Pesendorfer [2001]).[21] 'Intransitive' subjects cannot be explained by any utility-based model of choice between menus.

Table I shows the breakdown of subjects across these five categories. Preference for flexibility was common in our task, with 35% of subjects falling into this category. Overall, the large majority of subjects (85%) behaved in line with the preference for flexibility model of Equation 1. Of those that did not, 9% violated the transitivity condition. Only 6% of subjects fall into the commitment category, violating the set monotonicity condition.

[Table I approximately here]

A natural concern is that the preference for flexibility we observe is due to noise in the subject responses: if subjects were answering at random then some of them would exhibit preference for flexibility 'by chance'. This is a particular worry in our data, as data from Mechanical Turk is generally noisy. To rule out this possibility, we benchmark our results in two ways in the spirit of

[21]Another class of models that could capture these subjects is that which includes models of regret, for example Sarver [2008].

Bronars [1987] and Beatty and Crawford [2011]. Benchmark I in Table I is the fraction of subjects that would fall into each category if preference profiles were drawn uniformly at random from the possible transitive preference profiles. Benchmark II is the distribution of preference types that would result from subjects randomizing between the four possible preferences ($\succ, \succeq, \preceq, \prec$) in each set of pairwise choices. In both cases, the p-values report the result of the binomial test that the observed frequency is equal to the benchmark frequency. In both cases the fraction of subjects who exhibited preference for flexibility is significantly higher than in the benchmark, while the fraction of subject who exhibited preference for commitment is significantly lower.

Subjects were prepared to pay a significant amount for flexibility, given the context of the experiment. We can measure a subject's willingness to pay for flexibility as the minimum of their willingness to pay for the Flex contract over the Low contract and their willingness to pay for the Flex over the High contract. On average, the 43 Preference for Flexibility subjects had a willingness to pay for flexibility of \$0.07, a significant fraction of the amount that could be earned in the task section (a maximum of \$0.50).

It is notable how little preference for commitment we observe in our subjects. Following Gul and Pesendorfer [2001], much of the theoretical work on menu preference has been related to issues of temptation and self-control, and how these can generate a strict preference for smaller choice sets. Indeed, Kaur, Kremer and Mullainathan [2010] demonstrates a preference for commitment in a task similar to ours in a study of data entry personnel in India. As we discuss below, our work suggests that preference uncertainty may act as a powerful offsetting force to preference for commitment.

Preference for flexibility is also present for other values of the low threshold. For the F15/L15 and F10/L10 contracts, rates are similar to that of the F20/L20 case described above, with 34% and 35% of well-behaved subjects exhibiting strict preference for flexibility. At the F5/L5 level the rate is somewhat lower at 30%. Overall, 48% of well-behaved subjects exhibited preference for flexibility at some value of the low threshold.

4.2 Stochastic Choice

We next examine whether the stochastic choice data generated by our subjects is consistent with the preference uncertainty model. This first requires us to ensure that the data is consistent with some underlying preference uncertainty, then second to ensure that stochastic choice and menu preferences can be rationalized by the same underlying preference uncertainty.

Because we do not observe enough choices from the same menu to meaningfully identify individual-level stochastic choice functions, we instead constructed equivalence classes of subjects based on their responses to the menu preference questions. We make use of three different partitions on our set of subjects in the following analysis. First, we grouped subjects into the PFF, Standard and Indifferent categories defined above (we do not have enough subjects in the Commitment and Intransitive categories to construct meaningful stochastic choice functions). Next we divided subjects into those who strictly preferred F20 to H and those who did not. Finally we grouped subjects into those that strictly preferred F20 to L20 and those who did not.

[Table II approximately here]

4.2.1 Choice Monotonicity

As discussed in Section 1 the first key test is choice monotonicity: adding additional options to a set can only weakly diminish the probability of choosing the options previously in that set. For this experiment, both the L20 and H menus are subsets of the F20 menu, so the probability of choosing to do 0 tasks in the F20 menu must be lower than the probability of doing so in either the L20 or the H menus, and the probability of doing 20 (50) tasks must also be less in the F20 menu than it is in the L20 (H) menu. Table II displays the relevant frequencies with which subjects chose elements from those menus for all subjects. Table B.1 in Appendix B displays this information for Flex, Standard, and Indifferent subjects. Reported p-values reflect the sign test.

Table II and Appendix B show that behavior was broadly in line with choice monotonicity. For almost all groups and options, the probability of choosing that option is either significantly lower in the larger menu (for example, Table II shows that, for all subjects the probability of doing 20 tasks was significantly lower in the F20 contract than in the L20 contract) or there is no significant difference (more subjects performed 0 tasks in the F20 menu than in the L20 menu (17 vs 10) but the related sign test has a significance level of 0.14).

4.2.2 Consequentialism

[Table III approximately here]

Consequentialism requires that DMs only strictly prefer a larger choice set if the additional options are at least sometimes chosen from the resulting menu. In our experiment, this means that if a subject strictly preferred the F20 menu to either the L20 or the H menus, then he or she must exercise the additional option this grants them at least some of the time.

Because this prediction involves the stochastic choice function, we must utilize the equivalence classes in order to test it. The top panel of Table III divides subjects into those who exhibited a strict preference for F20 over H and those that did not. For each group it shows the fraction of subjects who did 20 questions when exogenously given the F20 contract. The second panel divides subjects into those who exhibited a strict preference for F20 over L20 and those that did not, reporting the fraction of subjects who performed 50 tasks in the exogenously provided F20 contract.

It is clear from Table III that the (rather moderate) requirement of the Consequentialism condition is satisfied. Subjects who strictly preferred F20 to H did 20 tasks 37% of the time in the F20 contract, while those that that strictly preferred F20 to L20 did 50 tasks 77% of the time in the F20 contract.

A more stringent condition is that subjects who are prepared to pay to have an option included in their choice set should use it more than those that are not. While this is not a requirement of the model characterized by AS, it is implied if the number of subjects who are indifferent between the different elements of the Flex contract is small.[22] Table III shows that subjects who paid to add an element to a menu were significantly more likely to exercise that option, with a Mann-Whitney p-value that is significant at the 0.01 level.

[22]The number of indifferent subjects is an upper bound on the number of subjects who do not strictly prefer F20 over H (L20) but complete 20 (50) tasks. For a further discussion see Section 4.2.3.

A corollary to the Consequentialism condition is that subjects who exhibit preference for flexibility should make use of both the 20 and 50 question levels. Table IV shows that this is the case. The first two columns show the distribution of choices made in the F20 contract by PFF subjects. The second two columns show the same distribution for Standard/Indifferent subjects who did not have a strict preference for F20 over L20. The last two show the distribution for such subjects who did not have a strict preference for F20 over H. Subjects who exhibited preference for flexibility made significant use of both the 20 and 50 task levels (28% and 65% respectively). Moreover, they made use of the 50 task level more than subjects who showed no strict preference for the F20 over the L20 contract (p=0.067) and made use of the 20 task level more that those who showed no strict preference for F20 over the H contract (p=0.026).

[Table IV approximately here]

4.2.3 Responsive Menu Preferences

The converse of the Consequentialism condition would be that $\lambda^{A\cup\{x\}}(x) > 0$ implies $A \cup \{x\} \succ A$. In our experiment, this would imply that subjects who chose to do 20 tasks when the F20 contract was exogenously given should strictly prefer F20 over H, while subjects who did 50 tasks should strictly prefer F20 to L20. Table V shows the menu preferences of these two groups of subjects. It shows that a large majority (83%) of subjects who did 20 tasks did indeed strictly prefer F20 to H. As a benchmark, only 47% of subjects who did 50 tasks prefer F20 to H - significantly different at the 0.1% level, p=0.001. Thus, subjects who actually used the 20 task effort level were more likely to pay to have it included in their contract. While this is not a direct implication of the AS model, it is what we would expect to see if there is heterogeneity in the distribution of effort costs across the population. Similarly, 71% of subjects who did 50 tasks preferred F20 to L20 (relative to 40% of the subjects who did 20 tasks - significantly different at the 1% level, p=0.003).

As AS point out this condition is overly restrictive: it could be the case that the subject was indifferent between (for example), performing 20 tasks for $0.20 or performing 50 tasks for $0.40, but preferred both to performing 0 tasks for $0.00. Such a subject would choose to do one of 20 or

50 tasks, but would not pay for that effort level to be included in their contract. Thus, the fact that some subjects who did 20 tasks (50 tasks) were not prepared to pay for F20 over H (F20 over L20) may be due to indifference, or a strength of preference of less than \$0.01, the smallest increment measurable in our multiple price lists. Such subjects should display no strict preferences between contracts - i.e. they should fall into the 'Indifferent' category. The final column of Table V shows that dropping these subjects increases compliance with the Responsive Menu Preference condition - to 96% for subjects who did 20 tasks from the F20 contract and to 83% for subjects who did 50 tasks.

[Table V approximately here]

4.3 Exogenous Uncertainty

Our final experiment tested whether exogenously introducing uncertainty about the contract increases preference for flexibility, and whether subjects respond predictably to changes in uncertainty. If a subject expresses preference for flexibility in the Easy or the Hard contracts in the 'Random Contracts' treatment, then they should also have a preference for flexibility for the Random contract. Furthermore, there may be subjects who exhibit preference for flexibility for the Random, but not for either the Easy or Hard contracts. This may include subjects who have no intrinsic preference uncertainty - for example a subject who always prefers (30 tasks, \$0.40) over (20 tasks, \$0.20), but always prefers (20 tasks, \$0.20) over (70 tasks, \$0.40) should exhibit preference for flexibility only for the Random contract. This means the fraction of subjects expressing preference for flexibility should be (weakly) higher in the Random contracts than in either the Easy or the Hard contracts.

Of the 61 well-behaved subjects in this experiment, 41% exhibited preference for flexibility for the Random contract, significantly higher than for the Easy (23%) and Hard (16%) contracts (McNemar p-values of 0.04, 0.01). At the individual level, 84% of subjects behaved in line with the prediction that preference for flexibility in either the Easy or Hard contracts should lead to preference for flexibility in the Random contract. 28% of subjects exhibited preference for flexibility only for the Random contract.

5 Discussion

5.1 Modeling Implications

Our results demonstrate that many of our subjects exhibit a strict preference for flexibility. We also demonstrate that preferences over work contracts are related to choices *from* those contracts, and respond in a predictable manner to exogenous changes in uncertainty. The latter two findings provide supporting evidence that the preference for flexibility we find is a rational response to preference uncertainty, rather than either a heuristic preference for larger choice sets, or a failure of rationality that leads subjects to ignore the worthlessness of flexibility when preferences are known. If subjects were irrationally demanding contracts with more options, we would not expect to see this flexibility being used, nor would we expect to see demand increase as flexibility becomes more valuable.

One possible alternative cause of preference for flexibility is a *preference for hedging*. As discussed by Epstein, Marinacci and Seo [2007] and Saito [2014], a DM with uncertainty averse preferences prefers to hedge between uncertain prospects by randomizing between them. If DMs can perform such randomizations on their own, then they may strictly prefer larger choice sets as this allows them to hedge using their own devices. Whether our data can be explained as resulting from a preference for a hedging depends on whether a particular effort/payment pair is viewed as an uncertain prospect at the time at which it is chosen from a given menu. If so, then a preference for hedging could indeed lead to a preference for flexibility, and would also imply monotonicity, consequentialism and responsive menu preferences. However, as the Certainty Strategic Rationality axiom of Saito [2014] makes clear, prospects the outcome of which are certain at the time of choice do not need to be hedged, and therefore do not lead to a preference for flexibility. Arguably, in our experiment, uncertainty has resolved itself by the time that choice of effort occurs, making preference for hedging a less compelling explanation for observed menu preferences.

Our results have important implications for models of temptation and self-control. Since the work of Gul and Pesendorfer [2001], it has been understood that a key implication of temptation is a preference for commitment - i.e. a preference for *smaller* menus. Such models often rule out preference for flexibility of the type observed in our experiment. Yet, despite the apparent ubiquity of temptation, commitment devices are relatively rare. One possible reason for this, suggested by Amador, Werning and Angeletos [2006], is the offsetting impact of preference uncertainty. For example, committing to performing a certain amount of work on a given day may be a bad idea if there are 'effort shocks' which may make providing that work very costly. Our results suggest that such preference uncertainty is important in a work contract setting. However, Kaur, Kremer and Mullainathan [2010] find a significant preference for commitment in a closely related context. This suggests that there may be significant welfare gains from designing 'smart' commitment contracts of the type discussed in Amador, Werning and Angeletos [2006] by incorporating a degree of flexibility into commitment devices. Indeed there already exist services which offer more flexible forms of commitment.[23]

5.2 What Is the Source of Subject Uncertainty?

In this experiment we find a considerable amount of preference for flexibility, despite the fact that there was a relatively small temporal gap between the choice of contract and its implementation. This begs the question: what is it that subjects were uncertain about? Our preferred interpretation is that it is the cost of supplying the effort necessary at the point at which the contract must be implemented. Possible sources include uncertainty over the difficulty of the particular questions they will be asked, changes in the opportunity cost of time (for example due to receiving a phone call), or the effect of fatigue and boredom. These sources of uncertainty can be important even across small time periods. Importantly, they all have natural counterparts in a wide variety of work settings. It is, however, plausible that uncertainty over the cost of effort grows with time, implying that our estimates may be a lower bound on the extent of preference for flexibility.

[23] For example, https://www.beeminder.com/.

One source of uncertainty that does not seem to be important for our subjects is about their *ability* to complete the requisite number of tasks withing the time limit. Of the 248 exogenous tasks sections in which subjects could have potentially completed level H, there were only 13 instances (5.2%) in which a subject used more than 85% of their 15 minute time limit without successfully completing level H. It is also possible that subjects realized that they would have been unable to complete level H at their current rate, and so quit the section before spending too much time on it. However, this also seems to have happened infrequently, as there were only 29 instances (11.7%) in which subjects would not have completed at least 50 tasks at their current rate.[24] Moreover, subjects who experienced a task section in which they would not have completed 50 tasks at their current rate actually expressed lower preference for flexibility than subjects who did not (0.17 vs 0.39, Fisher exact p=0.087).

6 Related Literature

There is a relatively small but growing literature on menu preference in laboratory and field experiments. In the laboratory, Sonsino and Mandelbaum [2001] document a preference for increased menu size and experimentally examine the tradeoff between a desire for larger choice sets and decision complexity by eliciting subjects' values for stochastic asset portfolios that vary in the number of options they contain. In their across-subjects design, they find that subjects placed a higher value on larger menus compared to strict subsets of those menus, which they interpret as a form of flexibility-seeking, based on the nature of the elements added. This differs from the definition of flexibility we consider, which requires a DM to value the union of two menus strictly higher than either of the submenus.

The related (though opposing) phenomenon of preference for smaller menus has also been experimentally documented in several environments, ranging from field experiments (Ashraf, Karlan and Yin [2006] and Giné, Karlan and Zinman [2010]) to laboratory experiments with real-effort

[24]Computed for subjects who completed at least five tasks.

tasks (Houser et al. [2010] and Augenblick, Niederle and Sprenger [2013]). Most closely related to our experiment is Kaur, Kremer and Mullainathan [2010], which studies commitment-seeking behavior in a field experiment involving data-entry workers in India who similarly choose between different contract structures.

Stochastic choice has been experimentally documented at least as far back as Tversky [1969]. Numerous experiments, often intended to test Expected Utility and relaxations thereof, have found that subjects often change their answer to a question when it is repeated, and have treated this data as a form of 'reliability check' as in Camerer [1989], or used this data to estimate a model with 'mistakes' or white noise variation, as in Hey and Orme [1994]. Later work has treated the stochastic choice as an explicit object of study, as in Hey [2001] which investigates whether the choices have reduced variability with increased repetition. More recently, Regenwetter and Davis-Stober [2012] test whether it is possible for seemingly intransitive choice data to come from a collection of underlying weak orders (as in GP), and finds that the data in their sample is consistent with that possibility. On the other hand, Agranov and Ortoleva [2013] also investigates stochastic choice behavior, but finds that subjects actually paid to use a costly randomization device, which they argue is a type of stochastic choice that is not explained using random utility, but fits better with models of hedging.

7 Conclusion

The objective of this project was to gather evidence of preference for flexibility in a controlled setting, determine whether preference uncertainty could offer a unified description of choice between and from menus, and provide evidence of a causal role for preference uncertainty in determining menu preferences. We argue that the experimental data we gather from a quasi-field setting achieves all three of these aims: 48% of our subjects exhibited strict preference for flexibility, subject behavior fits well with the unified AS model, and introducing exogenous uncertainty increases preference for flexibility.[25]

[25]Dean: Brown University; McNeill: Brown University

A Table of Preference Profiles

Type of Preference	Allowable Preferences										
Standard-L	F	$\succ$	H		F	$\succeq$	L		H	$\prec$	L
	F	$\succ$	H		F	$\preceq$	L		H	$\prec$	L
Standard-H	F	$\succeq$	H		F	$\succ$	L		H	$\succ$	L
	F	$\preceq$	H		F	$\succ$	L		H	$\succ$	L
PFF	F	$\succ$	H		F	$\succ$	L		H	$\preceq$	L
	F	$\succ$	H		F	$\succ$	L		H	$\prec$	L
	F	$\succ$	H		F	$\succ$	L		H	$\succeq$	L
	F	$\succ$	H		F	$\succ$	L		H	$\succ$	L
Commitment-L	F	$\succeq$	H		F	$\prec$	L		H	$\prec$	L
	F	$\succ$	H		F	$\prec$	L		H	$\prec$	L
	F	$\preceq$	H		F	$\prec$	L		H	$\prec$	L
Commitment-H	F	$\prec$	H		F	$\succeq$	L		H	$\succ$	L
	F	$\prec$	H		F	$\succ$	L		H	$\succ$	L
	F	$\prec$	H		F	$\preceq$	L		H	$\succ$	L
Commitment-Either	F	$\prec$	H		F	$\prec$	L		H	$\preceq$	L
	F	$\prec$	H		F	$\prec$	L		H	$\prec$	L
	F	$\prec$	H		F	$\prec$	L		H	$\preceq$	L
	F	$\prec$	H		F	$\prec$	L		H	$\succ$	L
Indifferent	F	$\succeq$	H		F	$\succeq$	L		H	$\succeq$	L
	F	$\succeq$	H		F	$\succeq$	L		H	$\preceq$	L
	F	$\succeq$	H		F	$\preceq$	L		H	$\succeq$	L
	F	$\succeq$	H		F	$\preceq$	L		H	$\preceq$	L
	F	$\preceq$	H		F	$\succeq$	L		H	$\succeq$	L
	F	$\preceq$	H		F	$\succeq$	L		H	$\preceq$	L
	F	$\preceq$	H		F	$\preceq$	L		H	$\succeq$	L
	F	$\preceq$	H		F	$\preceq$	L		H	$\preceq$	L

B Frequency of choice by contract, extended

Table B.1:
Frequency of choice by contract

Number of subjects who choose to do:			
	20 tasks in L20	20 tasks in F20	
All (N=124)	114	30	p= 0.00
PFF (N=43)	39	12	p= 0.00
Standard (N=40)	38	10	p= 0.00
Indiff (N=23)	21	4	p= 0.00
	50 tasks in H	50 tasks in F20	
All (N=124)	79	77	p= 0.83
PFF (N=43)	31	28	p= 0.58
Standard (N=40)	27	29	p= 0.63
Indiff (N=23)	11	11	p= 1.00
	0 tasks in H	0 tasks in F20	
All (N=124)	45	17	p= 0.00
PFF (N=43)	12	3	p= 0.04
Standard (N=40)	13	1	p= 0.00
Indiff (N=23)	12	8	p= .22
	0 tasks in L20	0 tasks in F20	
All (N=124)	10	17	p= 0.14
PFF (N=43)	4	3	p= 1.00
Standard (N=40)	2	1	p= 1.00
Indiff (N=23)	2	8	p= 0.07

References

Agranov, Marina and Pietro Ortoleva. 2013. “Stochastic Choice and Hedging*.”.

Ahn, David and Todd Sarver. 2013. “Preference for Flexibility and Random Choice.” *Econometrica* 81(1):341–361.

Amador, Manuel, Ivan Werning and George-Marios Angeletos. 2006. “Commitment vs. Flexibility.” *Econometrica* 74(2):365–396.

Ashraf, Nava, Dean Karlan and Wesley Yin. 2006. “Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines.” *The Quarterly Journal of Economics* 101(2):635–672.

Augenblick, Ned, Muriel Niederle and Charles Sprenger. 2013. “Working Over Time: Dynamic Inconsistency in Real Effort Tasks*.”.

Beatty, Timothy and Ian A Crawford. 2011. “How Demanding Is the Revealed Preference Approach to Demand?” *American Economic Review* 101(6):2782–2795.

Bronars, Stephen. 1987. “The Power of Nonparametric Tests of Preference Maximization.” *Econometrica* 55(3):693–698.

Camerer, Colin F. 1989. “An Experimental Test of Several Generalized Utility Theories.” *Journal of Risk and Uncertainty* 2(1):61–104.

Caplin, Andrew and Mark Dean. 2014. “Revealed Preference, Rational Inattention, and Costly Information Acquisition*.”.

Dekel, Eddie, Barton Lipman and Aldo Rustichini. 2001. “Representing Preferences with a Unique Subjective State Space.” *Econometrica* 69(4):891–934.

Dekel, Eddie, Barton Lipman, Aldo Rustichini and Todd Sarver. 2007. “Representing Preferences with a Unique Subjective State Space: A Corrigendum.” *Econometrica* 75(2):591–600.

Epstein, Larry G., Massimo Marinacci and Kyoungwon Seo. 2007. "Coarse Contingencies and Ambiguity." *Theoretical Economics* 2(4):355–394.

Giné, Xavier, Dean Karlan and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics* 2(4):213–235.

Goodman, Joseph, Cynthia Cryder and Amar Cheema. 2013. "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples." *Journal of Behavioral Decision Making* 26:213–224.

Gul, Faruk and Wolfgang Pesendorfer. 2001. "Temptation and Self-control." *Econometrica* 69(6):1403–1435.

Gul, Faruk and Wolfgang Pesendorfer. 2006. "Random Expected Utility." *Econometrica* 74(1):121–146.

Hey, John and Chris Orme. 1994. "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica* 62(6):1291–1326.

Hey, John D. 2001. "Does Repetition Improve Consistency?" *Experimental Economics* 4:5–54.

Houser, Daniel, Daniel Schunk, Joachim Winter and Erte Xiao. 2010. "Temptation and Commitment in the Laboratory*.".

Kaur, S, M Kremer and S Mullainathan. 2010. "Self-Control at Work: Evidence from a Field Experiment*.".

Kreps, DM. 1979. "A Representation Theorem for "Preference for Flexibility"." *Econometrica* 47(3):565–577.

Mason, Winter and Siddharth Suri. 2012. "Conducting Behavioral Research on Amazon's Mechanical Turk." *Behavior Research Methods* 44(1):1–23.

Paolacci, Gabriele, Jesse Chandler and Leonard N Stern. 2010. “Running Experiments on Amazon Mechanical Turk.” *Judgment and Decision Making* 5(5):411–419.

Regenwetter, Michel and Clintin Davis-Stober. 2012. “Behavioral Variability of Choices Versus Structural Inconsistency of Preferences.” *Psychological Review* 119(2):408–416.

Saito, Kota. 2014. “Preferences for Flexibility and Randomization under Uncertainty*.”.

Sarver, Todd. 2008. “Anticipating Regret: Why Fewer Options May Be Better.” *Econometrica* 76(2):263–305.

Sonsino, Doron and Marvin Mandelbaum. 2001. “On Preference for Flexibility and Complexity Aversion: Experimental Evidence.” *Theory and Decision* 51:197–216.

Tversky, Amos. 1969. “Intransitivity of Preferences.” *Psychological Review* 76(1):31–48.

Table I:
Breakdown of subject types in main treatment

Type	N	Percent	Benchmark I	p-value	Benchmark II	p-value
Flexibility	43	35%	17%	0.00	6%	0.00
Standard	40	32%	17%	0.00	6%	0.00
Indifferent	23	19%	25%	0.12	13%	0.06
Commitment	7	6%	42%	0.00	16%	0.00
Intransitive	11	9%	-	-	59%	0.00

Table II:
Frequency of choice by contract

Number of subjects who choose to do		N=124
20 tasks in L20	20 tasks in F20	
114	30	p= 0.00
50 tasks in H	50 tasks in F20	
79	77	p= 0.83
0 tasks in H	0 tasks in F20	
45	17	p= 0.00
0 tasks in L20	0 tasks in F20	
10	17	p= 0.14

Table III:
Proportions of subjects who pay to add an element to a menu to choose that element in the exogenously imposed F20 contract

Subjects who:	Fraction doing 20 tasks in F20	N	p-value
Do not strictly prefer F20 to H	0.09	57	p=0.00
Strictly prefer F20 to H	0.37	67	

Subjects who:	Fraction doing 50 tasks in F20	N	p-value
Do not strictly prefer F20 to L20	0.42	53	p=0.00
Strictly prefer F20 to L20	0.77	71	

Table IV:
Breakdown of subject choices in exogenous F20 contract

	PFF		L20/Indiff		H/Indiff	
	F20≻L20, F20≻H		Not F20≻L20		Not F20≻H	
N=	44 subjects		40 subjects		46 subjects	
Complete at least	Number	Percent	Number	Percent	Number	Percent
0 tasks	3	.07	8	.20	9	.20
20 tasks	12	.28	14	.35	4	.09
50 tasks	28	.65	18	.45	33	.72

Table V:
Frequency of preferring to add L20 (H), for subjects that do L20 (H) in F20

Subjects doing	Menu Preference:	All Subjects	Non-Indifferent
20 tasks in F20	$F20 \succ H$	0.83	0.96
	$F20 \succ L20$	0.40	0.46
50 tasks in F20	$F20 \succ H$	0.47	0.55
	$F20 \succ L20$	0.71	0.83

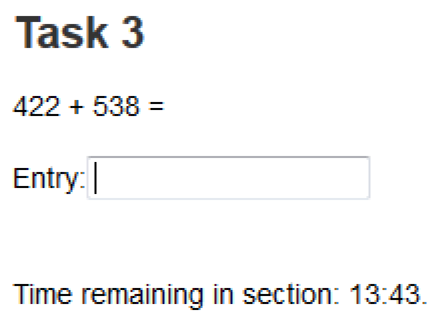


Figure I:
A typical task screen

Contract 11

Tasks completed	Payment
0-4	0.00
5-9	0.00
10-14	0.00
15-19	0.00
20-49	0.20
50+	0.20

Contract 25

Tasks completed	Payment
0-4	0.00
5-9	0.00
10-14	0.00
15-19	0.00
20-49	0.00
50+	0.40

Contract 24

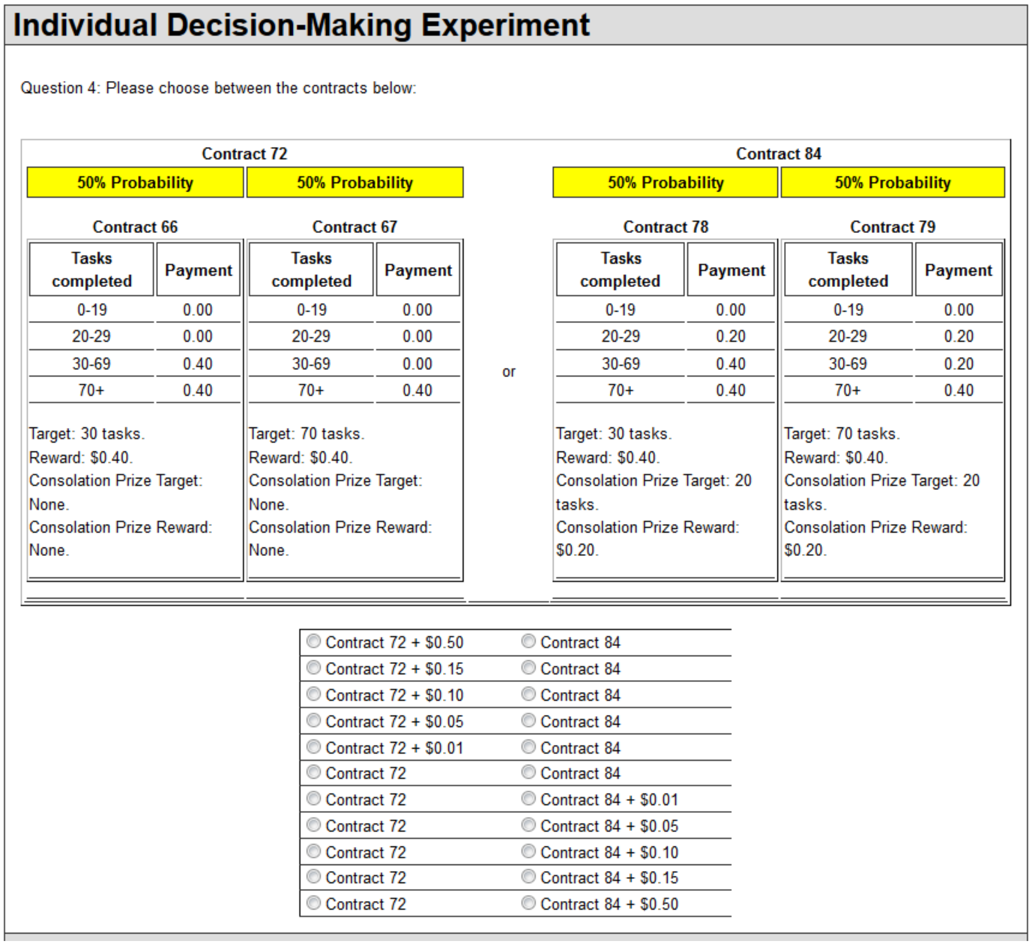
Tasks completed	Payment
0-4	0.00
5-9	0.00
10-14	0.00
15-19	0.00
20-49	0.20
50+	0.40

Figure II:
Low, High, and Flex contracts

Contract 25		Contract 24	
Tasks completed	**Payment**	**Tasks completed**	**Payment**
0-4	0.00	0-4	0.00
5-9	0.00	5-9	0.00
10-14	0.00	10-14	0.00
15-19	0.00	15-19	0.00
20-49	0.00	20-49	0.20
50+	0.40	50+	0.40

Contract 25 + $0.50	Contract 24
Contract 25 + $0.15	Contract 24
Contract 25 + $0.10	Contract 24
Contract 25 + $0.05	Contract 24
Contract 25 + $0.01	Contract 24
Contract 25	Contract 24
Contract 25	Contract 24 + $0.01
Contract 25	Contract 24 + $0.05
Contract 25	Contract 24 + $0.10
Contract 25	Contract 24 + $0.15
Contract 25	Contract 24 + $0.50

Figure III:
A typical MPL question (H vs F20)



Individual Decision-Making Experiment

Question 4: Please choose between the contracts below:

Contract 72 (50% Probability: Contract 66; 50% Probability: Contract 67)

or

Contract 84 (50% Probability: Contract 78; 50% Probability: Contract 79)

Contract 66

Tasks completed	Payment
0-19	0.00
20-29	0.00
30-69	0.40
70+	0.40

Target: 30 tasks.
Reward: $0.40.
Consolation Prize Target: None.
Consolation Prize Reward: None.

Contract 67

Tasks completed	Payment
0-19	0.00
20-29	0.00
30-69	0.00
70+	0.40

Target: 70 tasks.
Reward: $0.40.
Consolation Prize Target: None.
Consolation Prize Reward: None.

Contract 78

Tasks completed	Payment
0-19	0.00
20-29	0.20
30-69	0.40
70+	0.40

Target: 30 tasks.
Reward: $0.40.
Consolation Prize Target: 20 tasks.
Consolation Prize Reward: $0.20.

Contract 79

Tasks completed	Payment
0-19	0.00
20-29	0.20
30-69	0.20
70+	0.40

Target: 70 tasks.
Reward: $0.40.
Consolation Prize Target: 20 tasks.
Consolation Prize Reward: $0.20.

○ Contract 72 + $0.50	○ Contract 84
○ Contract 72 + $0.15	○ Contract 84
○ Contract 72 + $0.10	○ Contract 84
○ Contract 72 + $0.05	○ Contract 84
○ Contract 72 + $0.01	○ Contract 84
○ Contract 72	○ Contract 84
○ Contract 72	○ Contract 84 + $0.01
○ Contract 72	○ Contract 84 + $0.05
○ Contract 72	○ Contract 84 + $0.10
○ Contract 72	○ Contract 84 + $0.15
○ Contract 72	○ Contract 84 + $0.50

Figure IV:
A typical random contracts question