PREFERENCE FOR FLEXIBILITY AND RANDOM CHOICE: AN EXPERIMENTAL ANALYSIS*

Mark Dean†*and John McNeill*‡

Abstract

Agents may be uncertain about future preferences, leading to a preference for flexibility in menu choice and stochastic choice from menus. Such uncertainty may be important for contract design, and may offset preference for commitment arising from temptation. We experimentally measure choice between and from menus in a real-effort task. We find a preference for flexibility in 61% of subjects. Demand for flexibility persists even when contract choices are implemented immediately after the contract choice is made. The choice of contracts is predictive of subsequent choice of effort level, suggesting that preference for flexibility is a rational response to uncertainty.[1]

JEL codes: D81, C91

Preferences may be both unstable and unpredictable. Some days a worker may be energetic and willing to put in a lot of work in return for more pay, while on other days, she may feel tired and sluggish and be much less willing to work hard. Moreover, it may be difficult to predict in advance

*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 also thanks the Columbia Experimental Laboratory for the Social Sciences (CELSS) for support. A previous version of this paper (Dean and McNeill [2014]) was circulated using results from a subset of these experiments. The updated methodology represents our response to the helpful suggestion of reviewers.

†*Department of Economics, Columbia University, New York NY 10027. Email: mark.dean@columbia.edu*

‡*Department of Economics, University of Toronto, Toronto ON M5S 3G7. Email: john.mcneill@utoronto.edu*

[1]This experiment received IRB approval from Brown University (Protocol #1301000749) on March 05, 2013 and from Columbia University (IRB-AAAQ7956) on April 8th 2016.

in which of these states she will find herself on a particular day. Such preference uncertainty may lead to a 'preference for flexibility': a decision maker (DM) who is uncertain about her future cost of 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 studied in the decision theory literature,[2] and is of significant practical importance, for example in the design of labor contracts.

Preference uncertainty is of particular importance in relation to the sizable recent behavioral literature on issues of temptation and self-control. A key behavioral implication of temptation is a *preference for commitment*, or a desire to reduce the options available to choose from in the future. For example, a dieter may prefer a restaurant which sells only salads, rather than one which sells salads and hamburgers, to avoid the future temptation. Preference for commitment has received much theoretical[3] and empirical[4] attention, including in labor market settings similar to that we study below (Kaur, Kremer and Mullainathan [2014]). Yet arguably there are surprisingly few examples of commitment contracts outside those designed by behavioral economists (see, for example, Laibson et al. [2015]). One possible reason for this is preference uncertainty: recent work[5] has highlighted the role of uncertainty in offsetting any preference for commitment arising from time inconsistency related to problems of temptation. However, as yet there exist no empirical measures of the importance of preference uncertainty for menu choice.

This paper reports the results of three experiments which explore the extent to which preference uncertainty is an important component of behavior in a real-effort task. We first document that subjects exhibit significant preference for flexibility in their choice of labor contracts. This true both when choosing contracts for work to be performed 10 days in the future and, more surprisingly, for work to be performed immediately. Introducing an explicit stochastic element to the labor contract increases preference for flexibility, as would be expected if the preference for flexibility is a rational response to uncertainty. Secondly, we show that preference over contracts

[2] See for example Kreps [1979], Dekel, Lipman and Rustichini [2001] and Ahn and Sarver [2013].
[3] for example Gul and Pesendorfer [2001]; Noor [2007], see Lipman and Pesendorfer [2011] for a review.
[4] for example Ashraf, Karlan and Yin [2006]; Toussaert [2018*a*], see Bryan, Karlan and Nelson [2010] for a review.
[5] for example Amador, Werning and Angeletos [2006] and Galperti [2015].

is related to choice from those contracts in a manner that is qualitatively in line with that predicted by Ahn and Sarver [2013].[6]

Subjects in our experiment were offered the chance to perform simple arithmetic tasks for money. The experiment was conducted 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 had a high target and high payment, and a 'Flex' contract offered both low and high targets, with the corresponding payments. Subjects were asked to make choices between pairs of contracts using a multiple price list (MPL) method, and subsequently worked under various different contract regimes - some provided exogenously and some based on their choices.

Using this basic framework we ran three distinct experiments. In Experiment 1, subjects had the opportunity to work for pay on 14 consecutive days. For the first 11 days, subjects were exogenously provided with the Flex contract. After working on day 1, subjects were asked to make a series of incentivized choices which determined the work contract they faced on days 12-14. Thus, in Experiment 1 there was a significant time gap between the choice of contract and the implementation of that contract. In order to explore the importance of this time gap, we also ran a second experiment (Experiment 2) in which the chosen contract was implemented immediately after it was chosen. In Experiments 1 and 2, any uncertainty was naturally occurring - i.e. due to uncertainty of the subject about the costs and benefits of completing the experiment. Experiment 3 explored the causal impact of uncertainty on menu choice by introducing an explicit stochastic

[6]This model combines the preference for flexibility model of Dekel, Lipman and Rustichini [2001] with the stochastic choice model of Gul and Pesendorfer [2006].

element to the labor contracts. Such uncertainty should act to increase preference for flexibility.

We find a significant degree of preference for flexibility in all our experiments. In Experiment 1, 42% of well-behaved subjects[7] 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 parametrization, and 61% exhibited such preferences for some parameter values. Moreover, in our baseline specification, 93% of subjects demonstrated preferences in line with set monotonicity (i.e. never strictly prefer a smaller set), as required in Dekel, Lipman and Rustichini [2001] and Ahn and Sarver [2013]. Across all specifications, 18% of subjects exhibit some violation of set monotonicity.

One possible reason for the preference for flexibility exhibited in Experiment 1 is the significant time gap between the choice of a contract and its implementation. Experiment 2 suggests that this is not the only, or perhaps even primary, motivator for the demand for flexibility. Even when subjects are choosing between contracts that they use immediately after making that choice we still find considerable preference for flexibility - 51% of the subjects in the baseline parametrization (and 78% for some parametrization) exhibit a preference for flexibility in Experiment 2.

We also find that an exogenous increase in uncertainty leads to an increase in preference 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 Experiment 3, 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.

We next study subjects' choice of effort when faced with a labor contract, and compare this to their preferences between different types of contract. We find that the two are related, in line with the model of Ahn and Sarver [2013]. Using the 11 repetitions of the exogenously provided 'Flex'

[7]We describe what we mean by well-behaved in section 3.1 - essentially it means subjects who avoided making dominated menu choices.

contract in Experiment 1 we estimate stochastic choice functions for each subject. Ahn and Sarver [2013] provides two conditions that are necessary for menu preferences and stochastic choice to derive from 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 with positive probability. 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 effort level, while those who paid for the Flex contract over the High contract were significantly more likely to use the low effort level. Moreover, 90% of non-indifferent subjects who used the low target strictly preferred the Flex contract to the High contract, while 89% of such subjects who used the high target strictly preferred the Flex contract to the Low contract. 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

It is of interest to try to understand what it is that our subjects were uncertain about, particularly as we find a considerable amount of preference for flexibility in Experiment 2 when there is a small temporal gap between the choice of contract and its implementation. Our preferred interpretation is that subjects are uncertain about 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.

To our knowledge our paper is the first to experimentally document and study preference for flexibility in menu choice.[8] 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 a preference for larger

[8]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. Also related is the work of Costa-Gomes, Cueva and Gerasimou [2019] which demonstrates a desire for choice deferral - see section 6.

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]. Other recent papers have documented a preference for smaller choice sets in their subject population. In some cases (for example Kaur, Kremer and Mullainathan [2014]; Toussaert [2018*b*]) this is explained as due to desire to avoid temptation. Le Lec and Tarroux [2020] interpret their results as stemming either from fears about the decisions a future self will make, or cognitive short cuts on the part of the subject. Identifying the environmental factors which lead people to prefer larger or smaller choice sets is clearly an important direction for future research.

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. 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.

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, Lipman and Rustichini [2001], 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 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

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

stochastic choice function (in the sense of Equation 2). Of particular interest to our study are three behavioral implications:

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. **Consequentialism:**[10] 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:

$$A \cup \{x\} \succ A \implies \lambda^{A \cup \{x\}}(x) > 0$$

3. **Responsive Menu Preferences:**[11] 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 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 4.2.2.

[10] This is Axiom 1 of AS.
[11] This is Axiom 2 of AS.

The AS representation requires all of the axioms used in the construction of the DLRS and GP representations, some of which we did not seek to test in our experiment. DLRS requires that preferences over menus be a weak order and nontrivial. Our data does not allow us to test for intransitivity (though a previous version of the experiment - reported in Dean and McNeill [2014] - was able to perform such a test and found little evidence for intransitive preference). We discuss the fraction of our subjects who exhibit nontrivial menu preferences in section 4. The requirements of Lipschitz continuity on preferences over menus, 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' of 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 of this issue. Finally, GP require stochastic choice to obey monotonicity: adding options to a choice set cannot increase the probability that an existing option will be chosen - i.e. $x \in A \subset A' \implies \lambda^{A'}(x) \leq \lambda^{A}(x)$. In this version of the experiment we do not observe subjects choosing from nested subsets of alternatives, and so cannot test this property. However, in Dean and McNeill [2014] we do find support for monotonicity.

2 Experimental Design

We consider behavior in three related experiments. In each experiment subjects were offered the opportunity to complete effort tasks for payment, so that final choices were made over effort/money pairs. Subjects made choices from menus of such pairs - i.e. contracts that specified the payment for the number of tasks completed. In addition to observing subjects' choices of effort/money pairs

from different contracts, we also elicited their preference over such 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. In Experiment 1 subjects made choices over contracts which were then implemented 11-14 days in the future. In Experiment 2 subjects also made choices over contracts, but had their choices of contract implemented immediately after they had been made. Experiment 3 modified the design of Experiment 2 to include explicit randomness that would be resolved after the choice of contract but before the contract was implemented.

All experiments were conducted using Amazon's Mechanical Turk (MT) platform, a digital marketplace for work. "Requesters" 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.[12]

One advantage of using MT 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

[12]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, the subject exited the experiment prior to completing the instructions.

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 Experimental Components

Participants in all experiments were offered the opportunity to complete effort tasks for payment. The four major components of the experiment were the tasks that 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 the 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.

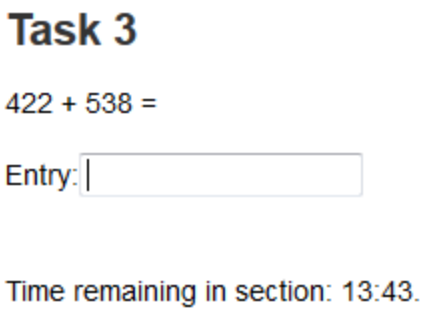


Figure I:
A typical task screen

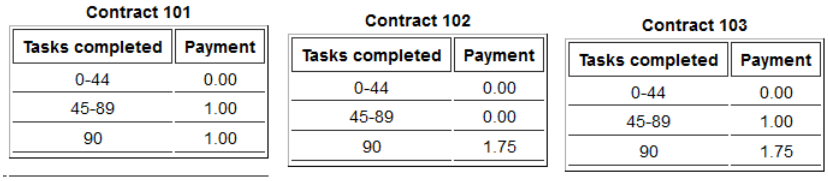
2.1.2 Task Sections

In each experiment subjects were presented with a number of 'task sections'. 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). In each section subjects could complete any number of tasks, but had to do so within a time limit.[13] Within the time limit they could complete as many tasks as they wished, although they could "retire" from the section at any time. Thus, a task section ended whenever the timer ran out, or the subject decided to stop. 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. Most subjects had ample time to complete the number of three-digit addition tasks required for maximal payments 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.

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.

[13] 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.



Contract 101

Tasks completed	Payment
0-44	0.00
45-89	1.00
90	1.00

Contract 102

Tasks completed	Payment
0-44	0.00
45-89	0.00
90	1.75

Contract 103

Tasks completed	Payment
0-44	0.00
45-89	1.00
90	1.75

Figure II:
Low, High, and Flex contracts

Under Contract 103 (Figure II), for example, completing fewer than 45 tasks would earn no additional compensation, completing between 45 and 90 tasks would earn $1.00, and completing 90 or more tasks would earn a total of $1.75 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 [2014]. However, our 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 important level(s) in a given contract were a low number of tasks for a lower payment (in Experiment 1 this was 45 tasks) or a high number of tasks for a higher payment (90 tasks in Experiment 1). The contracts used in all experiments were of three types. The "Flex" contract included the options to complete the low number tasks for a lower payment or the high number of tasks for a higher payment (as in Contract 103 of Figure II). The "High" contract included only the option to do the high tasks for the higher payment (as in Contract 102 in of Figure II). The "Low" contracts offered only the option to complete the low number of tasks for a lower payment, as in Contract 101.

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 future task sections, 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 subsequent task sections.

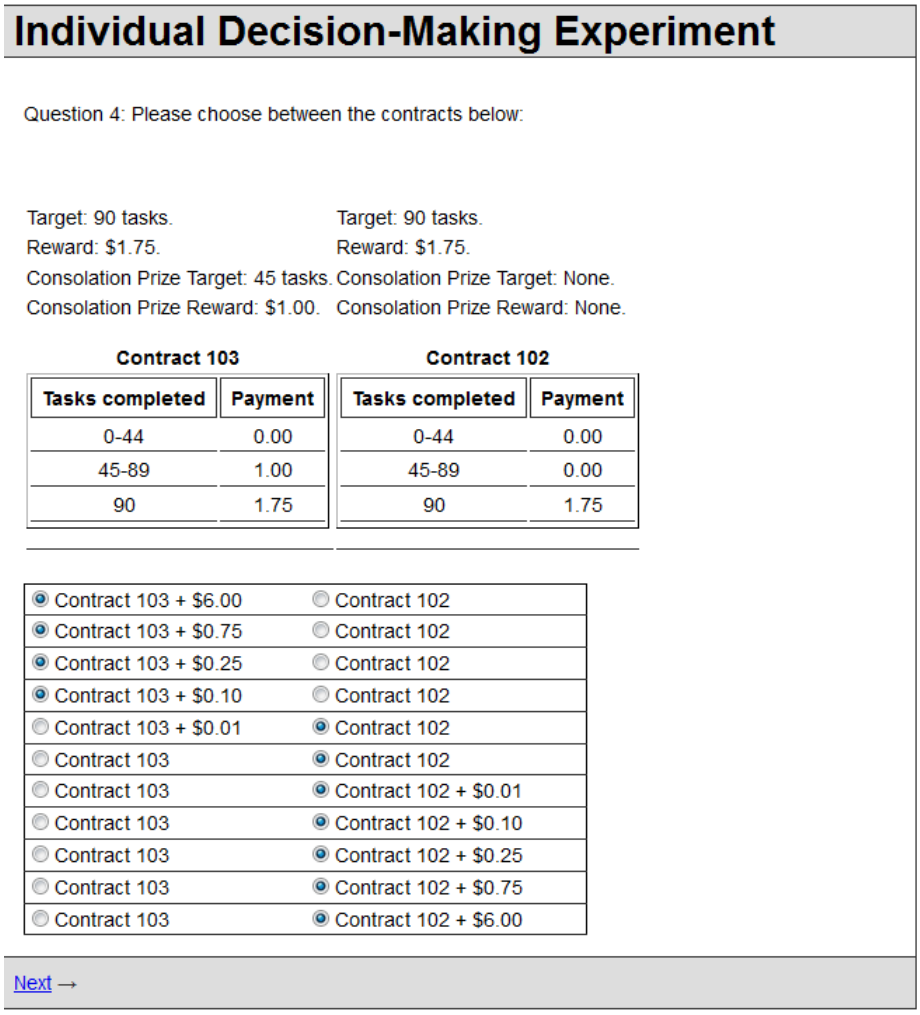


Figure III:
A typical MPL question (H vs F20)

In each MPL question the subject made a series of 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.

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 102" instead of "Contract 103 + $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.

2.2 Experiment 1

A subject's participation in Experiment 1 lasted for two weeks: on the first day the subject was introduced to the experiment and completed a task section with an exogenously specified contract, then responded to questions eliciting their preference over different contracts. On each of the subsequent 13 days, the subject could return to the experiment and complete one additional task section for payment. On 10 of those 13 days, the task section available to the subject was governed by the same contract that they used in the initial task section; on the remaining three days the contract in effect was determined by the subject's choices in the contract preference questions.

Participation, performance, and payment in each day after the first was independent. That is, subjects were not required to complete tasks in order to be eligible to participate in future tasks sections, nor would their performance on a given day impact their contracts in future periods. Subjects were paid for each day of the experiment, so earnings from one day were independent

of earnings from other days. On any day that the subject wished to return to the experiment and complete additional task sections, they could return to the same URL and use the same login information as on the first day of the experiment. Additionally, on the first day of the experiment, subjects were given the option to submit an email address that we used to send them an email reminder with their login details. They were under no obligation to share an email address with us; they were told the experimental URL and reminded of their login details so that they did not need to allow us any communication with them in order to continue to participate. Nonetheless, a substantial number of subjects did choose to share a contract email address with us; for details regarding participation in days 2-14 of the experiment, see Section 3.1.1.

The first day of 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.[14] 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. The subjects then completed a task section using Contract 103: a Flex contract with a low payment of $1.00 for 45 tasks and a high payment of $1.75 for 90 tasks.

After the first exogenous task section, subjects responded to eight questions eliciting their preferences over different contracts to be used for the last three days of the two-week period. Subjects responded to MPL questions eliciting their preferences between Flex contracts and High contracts, and between Flex contracts and Low contracts, all of which used the low payment of $1.00 for 45 tasks and high payment of $1.75 for 90 tasks. Because these questions refer to the payment levels that the subject had already actually experienced at the time of this choice, they are the focus of our main analysis. Additionally, subjects answered MPL questions between Flex and High contracts and between Flex and Low contracts for the following payment parameters: $1.00 for 45 tasks and $1.50 for 90 tasks, $1.25 for 45 tasks and $1.75 for 90 tasks, and $1.25 for 45

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

tasks and $2.00 for 90 tasks. The order of presentation was randomized between subjects. After the subject completed the questions, one of the lines of one of the questions was randomly selected to be realized during the three endogenous contracts days.[15]

After the final day of the two week period, each subject was paid according to his or her performance during the experiment through the MT payment interface. Because the subjects were free to return to the experiment on any of the fourteen days of the experiment (regardless of participation on any previous days), payment could not be made until the two weeks were complete, and subjects were aware that they could not influence the date of their payment through their actions in the experiment. Subjects were paid their earnings from each of the 14 tasks sections, as well as for the selected line in the MPL questions, and received a participation fee of $1.00.

2.3 Experiment 2

If future preferences are generally uncertain, preferences in the immediate future might be less uncertain than preferences in the more distant future. In Experiment 2 we test whether we still observe a preference for flexibility when there is effectively no gap between the choice of contract and its implementation. Subjects again completed effort tasks for money and made choices over which contract to use in the future, but then worked under that contract immediately after making the contract choice.

Unlike Experiment 1, all the elements in Experiment 2 took place on the same day. Otherwise, the basic structure remained the same: subjects initially completed a task section under an exogenously selected contract, then made choices between the same set of contracts as in Experiment 1, then completed three further task sections with a contract based on their choices. The key difference between the two experiments is that in Experiment 2 each subject's contract choice was implemented immediately after choosing between contracts.

[15] Due to a database error, a small fraction of participants actually experienced their endogenously-selected contract for four days, rather than three. Because they did not know this at the time at which they were making decisions (and in fact could not have known this until the final day of the experiment), they are included in this sample.

2.4 Experiment 3

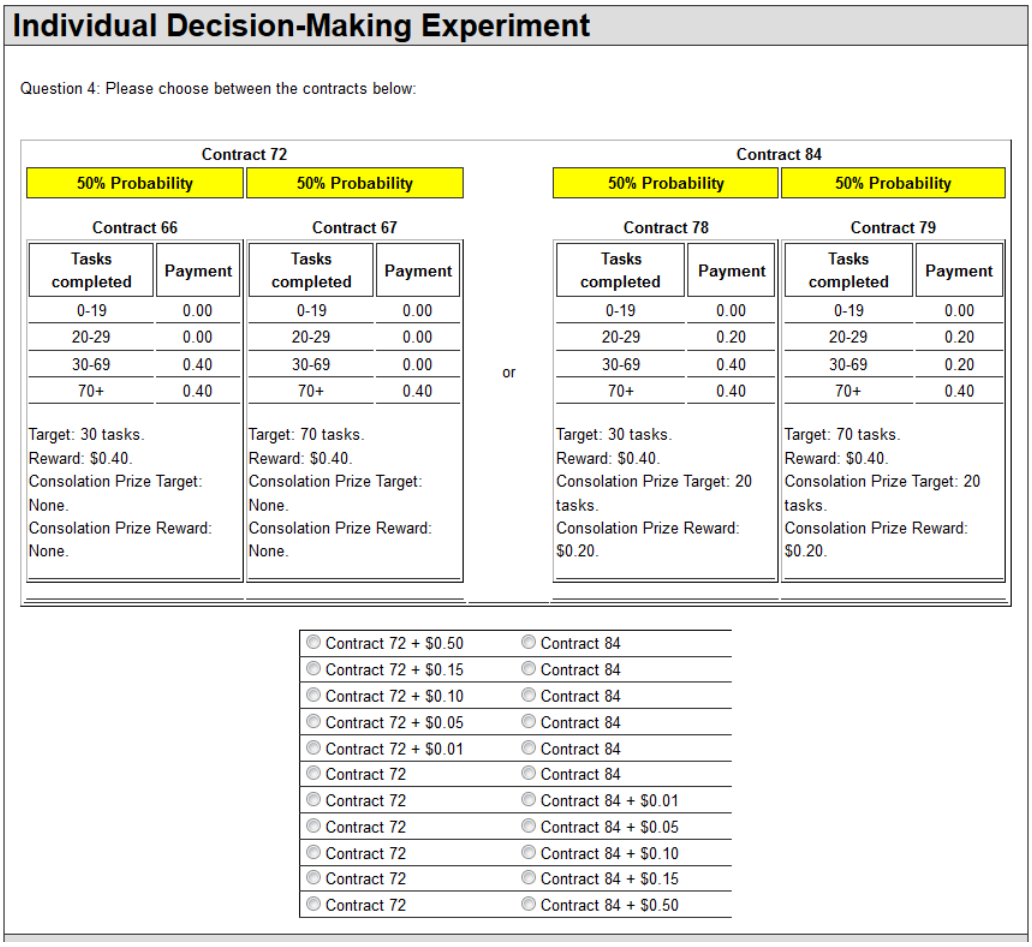
Any preference uncertainty found in the previous treatments is naturally-occurring. To test whether an increase in uncertainty would increase preference for flexibility, we ran another experiment which introduced exogenous uncertainty about some contract features, with the uncertainty resolved after the choice between contracts but prior to the section in which the contract was implemented. As with Experiment 2, all parts of Experiment 3 were completed in the same day. For Experiment 3, subjects experienced contracts with lower payments and fewer tasks required to earn each payment than in Experiment 1. Subjects completed three tasks sections with exogenously specified contracts (described below), then responded to contract-choice questions for a fourth and final task section completed immediately after the contract-selection questions.

Experiment 3 was structured 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 'easy' version of the high option and the contract with only the 'hard' version of the high option, and "Flex-random" which was a lottery between the Flex contract composed of the low option and the 'easy' version of the high option and the Flex contract composed of the low option and the 'hard' version of the high option.

After the instructions, subjects completed three exogenous task sections in a random order. Two of these involved nonrandom 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.



Individual Decision-Making Experiment

Question 4: Please choose between the contracts below:

Contract 72

50% Probability	50% Probability

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.

or

Contract 84

50% Probability	50% Probability

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

After the three exogenous tasks sections, subjects responded to MPL questions eliciting their preferences between Flex and Low contracts and between Flex and High contracts for the three different types of Flex (and High) contracts: easy, hard and random. The order of presentation was randomized between subjects. Figure IV shows the choice between the High-random and the Flex-random contracts.

As in Experiment 2, 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, 136 subjects participated in Experiment 1, 101 in Experiment 2, and 149 in Experiment 3 between November 9, 2013 and September 26, 2016. All subjects were recruited through Mechanical Turk, and completed the experiment over the Internet. Subjects in Experiment 1 and Experiment 2 earned an average of $11.89 and $6.64 respectively, including a participation fee of $1.00. Subjects in Experiments 3 earned $1.55, including a $0.25 participation fee.

3.1 Data Quality

While all experimental data contains a certain amount of noise, our use of Mechanical Turk suggests that our data requires a high level of quality control. 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 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. In Experiments 1and 2, the additional payments on the top and bottom

lines of each question are $6.00, which is more than three times what can be made with any given contract under consideration, and so the $6.00 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. For Experiment 3 the maximal side-payments were $0.50, still significantly more than the subject should be prepared to pay.

In Experiment 1 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. In these questions 93% of subjects never switched from right to left, while 98% 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 125 subjects or 92% of the original. For the 101 individuals recruited for Experiment 2, similar exclusion criteria yields 91 subjects (90%). For Experiment 3, these criteria retain 61 of the 149 subjects recruited (41%); this drastic difference in exclusion rate is due primarily to the fact that the endpoints of the MPL in Experiment 3 are significantly lower than in Experiments 1 and 2 - the exclusions are the results of subjects who select all of the options on one side of the MPL question.

3.1.1 Participation

Experiment 1 took place over two weeks. All subjects participated in a task section and contract-preference questions on the first day, and were invited to return to the experiment to complete an additional task section on as many of the rest of the days of the two week period as they chose. At the end of the two-week period they were paid for their performance on each of the days, as well as for their contract-selection question and participation fee. Participation on any given day of the two week period was not contingent on previous participation (with the exception of the first day), nor did performance on any given day impact the contracts applied to future days (with the exception of the contract-selection questions of the first day). Just over 70% of the subjects returned for at least one additional section, including 20% who returned for every section, while

the remaining 30% did not return to complete any additional sections. Table I shows the fraction of subjects who returned to complete additional tasks on each of the later days of the experiment.

Table I:
Participation: Percent of subjects completing at least 45 tasks in Days 2-14 in Experiment 1

Day number	2	3	4	5	6	7	8	9	10	11	12	13	14
Participation	.52	.50	.52	.44	.43	.50	.48	.50	.42	.43	.42	.42	.42

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

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.

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. Low and Flex contracts additionally contain the option (l_t, l_p), where l_t is a low number of tasks and l_p is the low payment. High and Flex contracts additionally contain the option (h_t, h_p), where h_t is the high number of tasks and h_p is the high payment. In this sense, the Flex contract is the union of the Low and High menus.[16]

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

[16]This is true if one considers non-dominated options. The low menu contains the option to do the high number of tasks for the low reward (i.e. (h_t, l_p)), whereas the Flex 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.

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. We estimated choices from menus using data from the task sections in which the contracts were exogenously determined.

Using the data from Experiment 1 we can estimate individual-level stochastic choice functions. Specifically, we look at the task sections from days 1-11, in which the subjects were presented with the Flex contract in which they could perform 45 tasks for $1.00 or 90 tasks for $1.75 (or 0 tasks for $0). This allows us to identify, for each day, which of three possible choices the subject made from the menu: 0, 45 or 90 tasks.[17] Pooling these choices across the 11 days gives an (admittedly coarse) estimate of the stochastic choice function for this Flex contract, based on the frequencies of each choice.

For Experiment 2, which took place in a single day, it was not feasible to collect enough data to estimate subject-level stochastic choice functions. 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. Due to the relatively small number of well-behaved subjects, and the complications induced by the Random contract, we do not estimate stochastic choice functions from Experiment 3.

[17] Subjects who did not log in on a given day were interpreted to select 0 tasks. Choices are measured as the highest level of contract completed. Thus, the Flex contracts contain three possible choices: subjects who completed between 0 and 44 tasks are included in the group that completed zero tasks, subjects who completed between 45 and 89 tasks are classified as choosing to complete 45 tasks, and subjects who completed 90 tasks are classified as such. In that way, it is possible for subjects to complete 'unnecessary' tasks, if they completed more than the minimum number required to earn a payment amount. This would happen, for instance, if they began to complete tasks to earn the next payment tier but changed their mind before completing that amount, or if they simply completed some additional tasks before realizing that they had achieved their intended target. On average, subjects completed 0.39 extra tasks per Flex contract; the median number of extra tasks per contract is 0, and the 75th percentile is 0.20.

4 Results

4.1 Choice Between Contracts

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 all experiments a subject exhibited a strict preference for flexibility if they strictly preferred the Flex contract to both the Low contract and the High contract. The preference uncertainty model also requires set monotonicity, which implies that the Flex contract must be weakly preferred to both the Low and the High contracts.

We used the data from the pairwise choices between Flex (F), Low (L) and High (H) to group subjects into four categories based on their revealed preferences:

1. Preference for Flexibility: $F \succ L$ and $F \succ H$

2. Standard: Either $F \succ H$ and $F \sim L$, or $F \succ L$ and $F \sim H$,

3. Indifferent: No strict preference between any contracts

4. Commitment: Either $H \succ F$ or $L \succ F$

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]).[18]

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

4.1.1 Experiment 1

Table II shows the breakdown of subjects across these four categories for the four different parameter levels in Experiment 1. Preference for flexibility was common in our experiment, with between 42% and 51% of subjects falling into this category across the different parameter values. Overall, 61% of subjects exhibited preference for flexibility at some parameter levels. At each parameter level, only a small number of subjects (between 6% and 10%) violate the predictions of the preference for flexibility model by exhibiting a strict preference for commitment. Overall, 18% of subjects exhibit preference for commitment at some parameter level.

Table II:
Breakdown of subject types: Experiment 1.
Columns 1-4 show proportion of subjects of each type for each of the 4 different permutations of low and high payment levels. Columns 5-6 show the proportion of simulated subjects of each type for two benchmarks.

Type	\$1.00, \$1.75	\$1.00, \$1,50	\$1.25, \$1.75	\$1.25,\$2.00	Benchmark I	Benchmark II
Flexibility	44%	48%	51%	46%	11%	16%
Standard	30%	26%	26%	26%	22%	16%
Indifferent	18%	19%	17%	18%	11%	4%
Commitment	7%	7%	6%	10%	55%	64%

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 experiment, as data from Mechanical Turk is generally noisy. To consider this possibility, we benchmark our results in two ways in the spirit of Bronars [1987] and Beatty and Crawford [2011]. Benchmark I is the distribution of preference types that would result from subjects randomizing between the three possible preferences $(\succ, \sim, \prec)$ in each set of pairwise choices. Benchmark II is generated by uniform randomly selecting switch points from the MPL for each contract preference question,[19] and generating the resulting preference relations. It is possible to reject the hypothesis that the fraction of Preference for Flexibility or Commitment subjects in our baseline specification is the same as in either Benchmark I or Bench-

[19]Excluding the possibility of selecting every option on one side, behavior that would be excluded on the basis of our data quality controls.

mark II, with too many of the former type and not enough of the latter (p<0.001%). The proportion of Standard and Indifferent types is rejected for being too high relative to both Benchmark I (p<0.05 level) and Benchmark II (p<.0.0001).

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 Preference for Flexibility subjects had a willingness to pay for flexibility of \$0.38 in the $l = \$1.00$, $h = \$1.75$ treatment, a significant fraction of the amount that could be earned in the task section.

We do not observe much preference for commitment 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 [2014] 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.

4.1.2 Experiment 2

We use Experiment 2 to ask whether the preference for flexibility in Experiment 1was driven solely by the time difference between contract choice and implementation. In Experiment 1, subjects had a considerable delay between the time of their choice between contracts and their later choice of effort under that contract, whereas in Experiment 2 subjects completed the final task section immediately after their contract choice questions. If the only source of preference for flexibility in Experiment 1 was the time between the contract choice and implementation, then in Experiment 2 we might expect to find very little preference for flexibility. We find this not to be the case: subjects in Experiment 2 also exhibit a large amount of preference for flexibility, with 51% of subjects expressing a strict preference for flexibility. Again we find that the proportion of Preference for

Flexibility subjects is rejected for being too high relative to both Benchmarks I and II, while the proportion of Commitment subjects is too low ($p<0.0001$ in all cases).

Table III:
Breakdown of subject types: Experiment 2

Type	N	Percent	Benchmark I	Benchmark II
Flexibility	46	51%	11%	16%
Standard	28	31%	22%	16%
Indifferent	11	12%	11%	4%
Commitment	6	7%	55%	64%

4.1.3 Experiment 3

We use Experiment 3 to examine 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 Experiment 3 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 contract, 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. In our sample 28% of subjects exhibited preference for flexibility only for the Random contract.

4.2 Choice from Contracts

Having established that preference for flexibility is a nontrivial phenomenon in all three of our experimental treatments, we turn now to the question of how closely choices from menus seem to be related to choices between menus. As discussed in Ahn and Sarver [2013], there are two primary criteria: that agents who pay to have an option available must at least sometimes make use of that option, which we refer to as "consequentialism", and that agents who use an option as the result of strict preference rather than indifference must be willing to pay to have that option included in their menu, which we refer to as "responsive menu preferences".

4.2.1 Consequentialism

Table IV:
Proportions of subjects in Experiment 1 who pay to add an element to a menu who choose that element in the exogenously imposed Flex contract

Subjects who:	N	Percentage Using 45	p-value	Percentage 45 Choices	p-value
Do not strictly prefer Flex to High	51	35%		7%	
Strictly prefer Flex to High	74	54%	.040	14%	.030

Subjects who:	N	Percentage Using 90	p-value	Percentage 90 Choices	p-value
Do not strictly prefer Flex to Low	44	50%		25%	
Strictly prefer Flex to Low	81	75%	.004	38%	.001

Consequentialism requires that subjects 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 Flex menu to either the Low or the High menus, then he or she must exercise the additional option this grants them at least some of the time.

We first test Consequentialism using the data from Experiment 1. The top panel of Table IV divides subjects into those who exhibited a strict preference for Flex over High contracts for the $l_p = \$1.00$, $h_p = \$1.75$ treatment, and those that did not. For each group it shows, in the column "Percentage Using 45", the fraction of subjects in each group who performed 45 tasks in at least one of the 11 occasions in which they were exogenously provided with the Flex contract (i.e. on days 1-11 of the experiment). It also shows the fraction of all days in which the 45 task level was

used ("Percentage 45 Choices"). The second panel divides subjects into those who exhibited a strict preference for Flex over Low and those that did not, reporting the fraction of subjects who performed 90 tasks in the exogenously provided Flex contract.

We can use this data to put a lower bound on the fraction of subjects who satisfy consequentialism. In our subject pool, we find that 60% of subjects utilize every option that they pay to include in their choice set (i.e use the 45 task level if they prefer Flex to High, and the 90 level if they prefer Flex to Low, and use both if they prefer Flex to both Low and High). These subjects satisfy Consequentialism. For the remaining 40% of subjects, we do not know whether they violate the Consequentialism axiom, or whether they have a positive probability of using the relevant task level, but happened, by chance, not to do so in our sample of 11 tasks. However, we can make probabilistic statements of the likelihood of such an event. For example, a subject who had a 50% probability of using the 45 task effort level in any task section has approximately a 0.05% chance of not doing so in any of the 11 trials. As a useful yardstick, the largest probability of using a given task that we can reject at the 5% level if we observe no such choices in 11 trials is 24%.

What is clear from Table IV is that there is a link between subject's menu preferences and their subsequent choice of effort level. 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.[20] Table IV shows that subjects who paid to add an element to a menu were significantly more likely to exercise that option: for example 75% of subjects who were prepared to pay for the Flex over the Low contract completed 90 questions, compared to 50% of those that were not prepared to pay. Subjects who paid for an option also made use of it more frequently: for example those that were prepared to pay for the high option used it in 38% of task sections compared to 25% for those that were not. All such differences are significant at the 5% level (Mann Whitney test).

[20]The number of indifferent subjects is an upper bound on the number of subjects who do not strictly prefer Flex over High (Low) but complete 45 (90) tasks. For a further discussion see Section 4.2.2.

Table V:
Proportions of subjects Experiment 2 who pay to add an element to a menu who choose that element in the exogenously imposed Flex contract

Subjects who:	Fraction doing 20 tasks in Flex	N	p-value
Do not strictly prefer Flex to High	0.08	26	p=0.02
Strictly prefer Flex to High	0.31	65	

Subjects who:	Fraction doing 50 tasks in Flex	N	p-value
Do not strictly prefer Flex to Low	0.45	33	p=0.004
Strictly prefer Flex to Low	0.67	58	

We see similar support for Consequentialism in Experiment 2 as can be seen in Table V. Because subjects in this treatment only make one choice from the Flex contract, we cannot construct meaningful random choice rules at the individual level. Instead, we pool subjects into an equivalence class based on whether or not they preferred to add a given option (either High or Low) to the contract, and ask whether the group of individuals who paid to add the High (Low) option were more likely to choose to complete the High (Low) number of tasks. Again we find that subjects who paid to include an option are much more likely to choose that option than are people who did not pay to include it in their contract.

4.2.2 Responsive Menu Preferences

The converse of the Consequentialism condition is that if a subject has some probability of choosing an alternative then they should be prepared to pay to include it in a menu: i.e. $\lambda^{A\cup\{x\}}(x) > 0$ implies $A \cup \{x\} \succ A$. In Experiment 1, this implies that subjects who made use of the 45 tasks under the Flex contract in some task section should strictly prefer Flex to High, while subjects who did 90 tasks should strictly prefer Flex to Low. Table VI shows the menu preferences of these two groups of subjects. It shows that a considerable majority (69%) of subjects who at some point did 45 tasks did indeed strictly prefer Flex to High. As a benchmark, only 51% of subjects who never use the 45 task level prefer Flex to High - significantly different at the 5% level (p=0.040). Thus, subjects who actually used the 45 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, 73% of subjects who did 90 tasks preferred Flex to Low (relative to 48% of the subjects who never used the 90 task level - significantly different at the 1% level, p=0.004).

As AS point out, requiring the converse of Consequentialism is overly restrictive: it could be the case that the subject was indifferent between (for example), performing 45 tasks for $1.00 or performing 90 tasks for $1.75, but preferred both to performing 0 tasks for $0.00. Such a subject would choose to do one of 45 or 90 tasks, but would not pay for that effort level to be included in their contract. Thus, the fact that some subjects who did 45 tasks (90 tasks) were not prepared to pay for Flex over High (Flex over Low) 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 VI shows that dropping these subjects increases compliance with the Responsive Menu Preference condition - to 87% for subjects who did 45 tasks from the Flex contract and to 91% for subjects who did 90 tasks.

Table VI:
Percentage of subjects in Experiment 1 preferring to add Low (High), for subjects that do and do not use the 45 (90) tasks in Flex contract

Subjects doing	Menu Preference:	All Subjects	Non-Indifferent
45 tasks in Flex	$F \succ H$	69%	87%
NOT 45 tasks in Flex	$F \succ H$	51%	61%
Subjects doing	Menu Preference:	All Subjects	Non-Indifferent
90 tasks in Flex	$F \succ L$	73%	91%
NOT 90 tasks in Flex	$F \succ L$	48%	57%

Similarly, subjects in Experiment 2 behave in line with the Responsive Menu Preferences condition. Of the subjects who completed the Low (High) number of tasks in the exogenously given Flex contract, 91% (73%) paid to have that option included in their menu, that is they paid to have the Flex contract rather than the High (Low) contract. Again, dropping those subjects who never expressed any strict preference between menus increases compliance to 91% for the Low option

and 85% for the High option.

Table VII:
Percentage of subjects in Experiment 2 preferring to add Low (High), for subjects that do and do not use the Low (High) number of tasks in Flex contract

Subjects doing	Menu Preference:	All Subjects	Non-Indifferent
45 tasks in Flex	$F \succ H$	91%	91%
NOT 45 tasks in Flex	$F \succ H$	65%	78%
Subjects doing	Menu Preference:	All Subjects	Non-Indifferent
90 tasks in Flex	$F \succ L$	72%	85%
NOT 90 tasks in Flex	$F \succ L$	51%	56%

5 Discussion

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 that flexibility has no value 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.

It is perhaps more surprising that we find as much preference for flexibility as we do in Experiment 2, when the time between the choice of contract and the utilization of that contract is collapsed. Finding considerable amounts of preference for flexibility in even the very short term is indicative that preference uncertainty may be important over even the comparatively short term. An alternate explanation would be that subjects in the Experiment 2 were uncertain about their ability to complete tasks due to the relatively greater time pressure. However, this explanation

seems unlikely given the subjects' performance in the task section they completed prior to making choices between contracts - it seems as though the time constraint was rarely a factor for these subjects. Subjects rarely completed tasks at a rate that would be incommensurate with being able to complete the High number of tasks: in only 12% of tasks sections did subjects complete tasks at a rate that would not have finished the High level of tasks within the time limit. In fact, subjects who experienced a task section in which they would not have completed 50 tasks at their current rate expressed lower preference for flexibility than subjects who did not (0.17 vs 0.39, Fisher exact p=0.087).

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 (see Laibson et al. [2015] particularly). One possible reason for this, suggested by Amador, Werning and Angeletos [2006], is the offsetting impact of preference uncertainty. For example, committing to perform 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 show that such preference uncertainty can be important in a work contract setting. However, Kaur, Kremer and Mullainathan [2014] 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.[21]

One possible alternative cause of a preference for flexibility is a *preference for hedging.* As discussed by Epstein, Marinacci and Seo [2007] and Saito [2015], 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

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

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 [2015] 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.

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 trade-off 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], Giné, Karlan and Zinman [2010], and Kaur, Kremer and Mullainathan [2014]) to laboratory experiments with real-effort tasks (Houser et al. [2010], Augenblick, Niederle and Sprenger [2013] and Toussaert [2018*a*]). Most closely related to our experiment is Kaur, Kremer and Mullainathan [2014], which studies commitment-seeking behavior in a field experiment

involving data-entry workers in India who similarly choose between different contract structures. An important aspect of future work is therefore to identify the interplay between preference for commitment and flexibility.

A recent paper by Le Lec and Tarroux [2020] experimentally studies menu preferences. The objects of choice are websites that the subjects can look at in the second half of the experiment. Unlike our results, the authors find strong evidence for choice aversion - i.e. the value of larger choices is below what would be expected given the valuation of smaller choice sets. Interestingly, rather than the possibility of temptation, the preferred explanation is a fear of bad decisions at the second stage. Combined with our results this suggest that choice set attitudes are dependent on the underlying object of choice. Understanding the nature of this relationship is an important avenue for future research. Indeed a recent working paper, Costa-Gomes, Cueva and Gerasimou [2019] finds evidence that subjects will pay to defer choice between consumer goods, which is operationally similar to our preference for flexibility.

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

In this paper we ask whether there is evidence that preference for flexibility is a potentially important consideration in menu choice, particularly with respect to the design of contracts, and if so, whether that flexibility-seeking behavior is consistent with the preference uncertainty interpretation. We gather data from a quasi-field experimental labor market setting, in which we observe subjects' choices between contracts with varying degrees of flexibility as well as their behavior when working under those contracts. We find that for our subjects, flexibility is an important consideration in the choice of labor contracts: 61% of our subjects expressed a strict preference for flexibility for some set of parameters. Moreover, the subjects' choices between contracts were predictive of their behavior when choosing effort levels from contracts, in a way consistent with the predictions of Ahn and Sarver [2013] which relate preference uncertainty to preference for flexibility. Further, we find that flexibility-seeking behavior responds predictably to the introduction of exogenous uncertainty, but also persists even when the plausible degree of non-preference uncertainty is diminished by implementing the chosen contracts immediately after the choice between menus. In sum, the behavior of the subjects in our experiment is consistent with agents who compensate for preference uncertainty by demanding flexibility. As we discuss in Section 5, preference for flexibility motivated by preference uncertainty may play an important role in understanding the observed menu choices of agents, particularly those who also experience issues related to temptation or self-control.[22]

[22]Dean: Columbia University; McNeill: University of Toronto

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.

Bryan, Gharad, Dean Karlan and Scott Nelson. 2010. "Commitment devices." *Annu. Rev. Econ.* 2(1):671–698.

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

Costa-Gomes, Miguel, Carlos Cueva and Georgios Gerasimou. 2019. *Choice, deferral and consistency*.

Dean, Mark and John McNeill. 2014. Preference for Flexibility and Random Choice: an Experimental Analysis. Working Papers 2014-10 Brown University, Department of Economics.

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

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

Galperti, Simone. 2015. "Commitment, Flexibility, and Optimal Screening of Time Inconsistency." *Econometrica* 83(4):1425–1465.

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.

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, Supreet, Michael Kremer and Sendhil Mullainathan. 2014. "Self-Control at Work." *Journal of Political Economy, forthcoming* .

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

Laibson, David et al. 2015. "Why Don't Present-Biased Agents Make Commitments?" *American Economic Review* 105(5):267–72.

Le Lec, Fabrice and Benoît Tarroux. 2020. "On attitudes to choice: some experimental evidence on choice aversion." *Journal of the European Economic Association* 18(5):2108–2134.

Lipman, Barton L and Wolfgang Pesendorfer. 2011. "Temptation." *Working Paper* .

Noor, Jawwad. 2007. "Commitment and Self-control." 135:1–34.

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

Saito, Kota. 2015. "Preferences for Flexibility and Randomization under Uncertainty." *American Economic Review* 105(3):1246–1271.

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.

Toussaert, Séverine. 2018*a*. "Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment." *Econometrica* 86(3):859–889.

Toussaert, Séverine. 2018*b*. "Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment." *Econometrica* 86(3):859–889.

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