Do Decision Makers Have Subjective Probabilities? An Experimental Test

David Ronayne, Roberto Veneziani and William R. Zame

Working Paper No. 940 June 2022 ISSN 1473-0278

School of Economics and Finance

Queen Mary
University of London

Do Decision Makers have Subjective Probabilities? An Experimental Test*

David Ronayne†, Roberto Veneziani‡ and William R. Zame§

This version: June 21, 2022

Abstract: Anscombe and Aumann (1963) offer a definition of subjective probability in terms of comparisons with objective probabilities. That definition – which has provided the basis for much of the succeeding work on subjective probability – presumes that the subjective probability of an event is independent of the prize consequences of that event, a property we term Prize Independence. We design experiments to test Prize Independence and find that a large fraction of our subjects violate it; thus, they do not have subjective probabilities. These findings raise questions about the empirical relevance of much of the literature on subjective probability.

JEL: D01 (Microeconomic Behavior: Underlying Principles), D81 (Criteria for Decision-Making under Risk and Uncertainty), D84 (Expectations, Speculations), C09 (Design of Experiments)

Keywords: subjective probability, choice under uncertainty, online experiments.

*We are grateful for helpful discussions with Alexander Cappelen, Matt Elliott, Duarte Gonçalves, Shachar Kariv, Marco Mariotti, and Peter Matthews. We thank audiences at the Royal Economic Society conference, the Analytical Political Economy workshop, and UCL for comments and suggestions. This research was begun while Ronayne was a Postdoctoral Fellow and Zame was an Academic Visitor at Nuffield College, University of Oxford. Ronayne is grateful for financial support from the University of Oxford's George Webb Medley Fund; Zame is grateful for support and hospitality from the Geneva Finance Research Institute, University College London, and the Einaudi Institute for Economics and Finance. Ethical approval was granted by the University of Oxford's Economics Departmental Research Ethics Committee (ECONCIA17-18-011) and ESMT's Ethics Committee (21-001-E). The experiments are registered at the AEA RCT registry `https://doi.org/10.1257/rct.6174-2.0`.

†European School of Management and Technology (ESMT Berlin). `david.ronayne@esmt.org`

‡Queen Mary University of London. `r.veneziani@qmul.ac.uk`

§University of California at Los Angeles. `zame@econ.ucla.edu`

1 Introduction

If a Decision Maker in New York assigns probability 1/2 to the outcome of a fair coin toss being Heads, we think we know what the Decision Maker means – and we think that all rational Decision Makers would assign the same probability. Indeed, we usually view a fair coin toss as a canonical example of an *objective* probability. But if the same Decision Maker assigns *subjective* probability 1/2 to rain at high noon, six months from today, in Zame (a small village in Burkina Faso), what do we think that means? It seems natural to think it means that this Decision Maker would be indifferent between a lottery which pays out if it rains in Zame and a lottery with the same prizes which pays out if the coin lands Heads.

Indeed, this method of interpreting *subjective probabilities* in terms of *objective probabilities* is precisely the way in which Anscombe and Aumann (1963) *define* subjective probabilities. This seems compelling to us, and to most of the large literature on subjective probability.[1] But for this definition to be really convincing, it would seem that the Decision Maker's indifference should be *independent of the prize* offered in the two lotteries. Anscombe and Aumann (1963, p.203) show that this property, which we dub "Prize Independence", is a consequence of the axioms they put forward but Prize Independence is not at all unique to their approach. It is Assumption P4 in Savage (1972 [1954]), which precedes their work; it is derived in Machina and Schmeidler (1992), which follows (and generalizes) their work; and it is derived or assumed in almost all of the large literature on subjective probability.[2] It seems to us, and to much of the literature, that *any* reasonable notion of subjective probability *must* satisfy Prize Independence.

This paper presents the results of the first experiments explicitly designed to ask whether people have subjective probabilities by testing the empirical validity of the property of Prize Independence. We offer our subjects the *same* objective and subjective lotteries for a range of *different* prizes. Our experiments build on the approach pioneered by Holt and Laury (2002), and feature design choices inspired by Anscombe and Aumann (1963), whose framework lends itself well to our objective. We present subjects with three sets of choices. In each, subjects must choose between a single fixed lottery with an *unknown* probability of success and a sequence of lotteries with *known* probabilities of success. Within each set of choices, the prize for success is *the*

[1]That literature is too large to review here; we discuss some of the most relevant contributions in Section 4 and we refer to Karni (2014) and Karni et al. (2015) for broader surveys.

[2]The assumption of state-dependent utility (Karni and Safra, 1995) can provide a notable exception, but not one relevant to our experiments (as we discuss in Section 4.3).

same for all lotteries, but across sets of choices the prizes are *different*.

Prize Independence predicts that the choices between the fixed lottery with an unknown probability of success and the various lotteries with known probabilities of success should be *the same* across the three sets of choices. However, we find that the choices of two thirds of subjects violate Prize Independence. Remarkably, even when we restrict attention to subjects who appear attentive, whose choices conform with standard requirements of rationality, who understand the instructions well, and who display a good understanding of basic notions of probability, we find that the choices of about half of our subjects violate Prize Independence. We even find the same fraction of violations among the choices of subjects who are classified by the subject pool operator to be highly skilled and attentive. We conclude that, even abstracting from inattentive and irrational behavior, *about half of our subjects do not have subjective probabilities*.

We emphasize that we do *not* test the specific axiomatic structure put forward by Anscombe and Aumann (1963), but rather the property of Prize Independence. As we have noted above, virtually all of the main approaches to subjective probability taken in the literature either assume or derive Prize Independence. Thus, our findings would seem to cast doubt on the empirical relevance of this entire literature and not only of Anscombe and Aumann (1963) specifically.

In the experimental literature on subjective probabilities, the focus is on methods to elicit them. Classic examples include the proper scoring rule method (Savage, 1971), the promissory notes method (De Finetti, 1974), and the lotteries method (Kadane and Winkler, 1988). Most related, the famous Becker-DeGroot-Marshak mechanism (Becker et al., 1964) aims to elicit an objective probability equivalent of a subjective probability.[3] Although our design is not intended to elicit subjective probabilities, it could be used for that purpose.[4] However, as we shall see, the choices of many of our subjects are *not* consistent across prize levels; at least for these subjects, providing such estimates would seem to be a pointless exercise.

The rest of the paper is structured as follows. In Section 2 we describe the experimental designs and in Section 3 we present the data. In Section 4 we discuss interpretations and explanations of our data, relating our findings to relevant literature.

[3]See also Armantier and Treich (2013), Hao and Houser (2012), Holt and Smith (2009), and Karni (2009).

[4]We present each subject with three sets of choices between one, fixed lottery with an unknown probability of success, and a series of lotteries with known (and increasing) probabilities of success. *If* a subject's choices are both consistent across prize levels and rational, then the number of times they choose the lottery with the unknown chance of payout provides an interval estimate of their beliefs of that chance.

2 Experiments

2.1 Overview

The Anscombe and Aumann (1963) definition of an individual's subjective probability is organized around an individual's preferences over and between lotteries with known and unknown probabilities. We tie our design closely to their theory.

We give each subject a task in which they have to make 21 choices. Each choice is between two boxes. Each box contains 20 balls; some Red and the remainder Yellow. After the subject has made all 21 choices, one choice is drawn at random and a ball from the chosen box is drawn at random; if the ball is Red the subject wins a prize (either \$1, \$5, or \$20); if the ball is Yellow, the subject wins nothing. This is explained to subjects in the instructions before the main page of choices. In addition, a summary is provided to them at the top of the page of choices: *"To maximize your chance of winning a bonus: in each choice, choose the box you think has the highest number of red balls."*

A key element of our design is the use of two types of box: *open* and *half-open*.[5] Open boxes display all 20 balls they contain so subjects see how many are Red and how many are Yellow. There are seven open boxes, which vary in how many Red balls they contain: 2, 4, 6, 8, 10, 12, and 14, corresponding to payout probabilities: 0.1, 0.2, ..., 0.7, respectively.[6]

There is *one* half-open box, which is fixed throughout (the number of Red and Yellow balls in it does not change, a point emphasized to subjects), and subjects can see 10 of its 20 balls.[7] Of the 10 visible balls, three are Red and seven are Yellow, but we provide no information about the remaining 10. As such, from the subjects' perspective the half-open box could have any number of Red balls between 3 and 13 (corresponding to payout probabilities between 0.15 and 0.65). We present the 10 unseen balls with a question mark on them.[8] Each of the 21 choices is a choice between one of the open boxes and the half-open box.

[5] Open boxes with a known probability of winning correspond to Anscombe and Aumann's (1963) *roulette lotteries*, whereas the half-open box corresponds to one of their *horse lotteries*.

[6] We decided to have 20 balls per box and to differentiate the open boxes by increments of two Red balls. We chose this rather than, say, 10 balls per box and increments of one to make the differences clearer.

[7] We made *some* balls visible to make it clear that there are balls inside, in the same way we did for the open boxes.

[8] The actual number of Red balls in the half-open box (unknown to subjects of course) was seven (i.e., the three visible Red balls plus four others). This was determined at random by us once and for all, before the experiments.

2.2 Primary Design – Experiments 1 and 1_M

In our Primary Design, the 21 choices faced by subjects are organized in three sets of seven choices. The three sets are identical except for the prize offered; within each set the prize is \$1, \$5, or \$20; each subject faces one set at each of these prize levels. The amounts are salient numbers, each noticeably different from the others, and all at least as large as the participation fee of \$1. We expect the prizes to be perceived as substantial for most of our subjects, who regularly respond to smaller stakes. We chose to have three prize levels rather than two because, aside from exploring more of the domain of prize levels, three allows us to detect richer patterns of choices across levels, should there be any such variation.

The order in which the seven pairs of choices within each set are presented is fixed to be ascending in the number of Red balls in the open boxes, while the order in which the three sets are presented is randomized. All 21 choices are presented together in a list, on a single webpage. This allows subjects to scroll up and down, easily compare their choices, and change their mind if they wish to. Figure 1 is a screenshot showing one set of choices.[9]

We test the Primary Design first on a large, unrestricted sample of subjects (Experiment 1) and then on a smaller sample of subjects classified by the subject pool operator to be more skilled and attentive (Experiment 1_M); see Section 2.7 for details.

2.3 Rational Choices and Prize Independence

The assumption of rationality makes significant predictions about choices within each set. The half-open box (which contains at least 3 Red balls) dominates the open box with 2 Red balls, and so the half-open box should be strictly preferred to this open box. At the other extreme, the open box with 14 Red balls dominates the half-open box (which contains at most 13 Red balls), and so this open box should be strictly preferred to the half-open box. Moreover, because a larger number of Red balls means a higher probability of winning the prize, boxes with more Red balls should be strictly preferred to boxes with fewer Red balls; transitivity implies that if an open box with s Red balls is preferred to the half-open box, then any open box with *more than* s Red balls should also be preferred to the half-open box. Putting these predictions together, we

[9]Subjects can scroll up and down if their screen does not show all 21 choices at once. The pages of this manuscript, however, are too small to show all 21; Figure 1 displays the first seven.

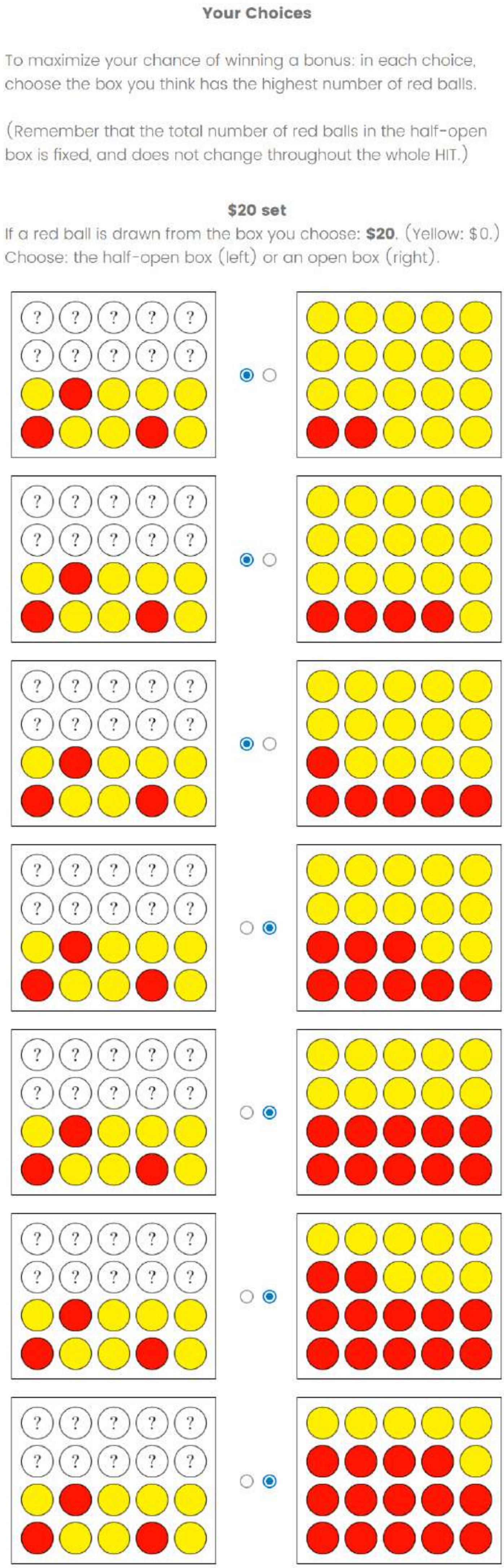


Figure 1: Primary Design. Subjects see a summary of the instructions followed by 21 choices, organized in three sets of seven. The three sets (which differed only by prize level: $1, $5, $20) are shown in a random order. In this screenshot (cropped from below), the $20 set is shown first (for subjects, the other two sets would follow immediately below, on the same webpage). The seven choices *within* a set are ordered by the number of Red balls in the open boxes. By scrolling, subjects can easily check, compare, and amend any of their 21 choices. Subjects are required to make all 21 choices by selecting one of the two options and there is no time limit. The screenshot shows an example in which the half-open box is chosen in the first three choices and the open boxes in the fourth to seventh. In this example, the subject chooses the half-open box three times and then switches to the open boxes; we say that the subject's switch point is 3. If the subject's switch point is also 3 in the $1 and $5 prize sets, their choices are consistent with Prize Independence.

see that the assumption of rationality implies that, within each set (i.e., at a given prize level), there should be some *unique* number $s \in \{1, 2, \ldots, 6\}$ such that the subject prefers the half-open box to every open box containing $m \leq 2s$ Red balls, and prefers every open box containing $m > 2s$ Red balls to the half open box. Put differently: a rational subject should choose the half-open box rather than the open box with 2 Red balls, and *should switch* from choosing the half-open box to choosing an open box *exactly once*. The result is that rational subjects should choose the half-open box s times. We refer to the unique number s as the *switch point*.

We stress that, although the assumption of rationality predicts that a subject has a single switch point *within* each prize level, it makes no prediction about switch points *between* prize levels. Choices of different switch points at different prize levels would not be violations of rationality – but they would be violations of Prize Independence. However, in conjunction with rationality, Prize Independence requires that a subject's switch points be the same at each prize level.

2.4 Secondary Design – Experiment 2

The Primary Design's presentation format is inspired by widely adopted methods (e.g., Holt and Laury's (2002) risk preference elicitation) and offers a natural way to test Prize Independence. But it is of course not the only way to present the choices. We intend our Secondary Design to be a tough check, in that it requires subjects to make the same 21 choices, but framed in such a way as to make it *easier* for subjects to make decisions consistent with Prize Independence.

In our Secondary Design, we reveal one pair of lotteries at a time and ask subjects to choose between them three times, once for each prize level. The webpage initially displays only the first two boxes, called "The Half-Open Box" and "Open Box 1" (as shown in Figure 2), and the first question *"For a prize of $1, which box do you choose?"* Once the subject chooses, the second question appears: *"For a prize of $5, which box do you choose?"* And, again, once the subject chooses, the third question appears: *"For a prize of $20, which box do you choose?"* After the first three choices are made, and below the last ($20) choice for the first pair of boxes, the next pair of boxes (The Half-Open Box and Open Box 2) is revealed along with the corresponding $1 question. This process of gradual revelation (chosen to enhance engagement) continues until all three choices are made for all seven pairs of boxes. Thus, as in the Primary Design, the subject is left with all 21 choices on one page. At any point, they may scroll up or down to check or amend their prior answers; this does not make any content disappear or change in any way.

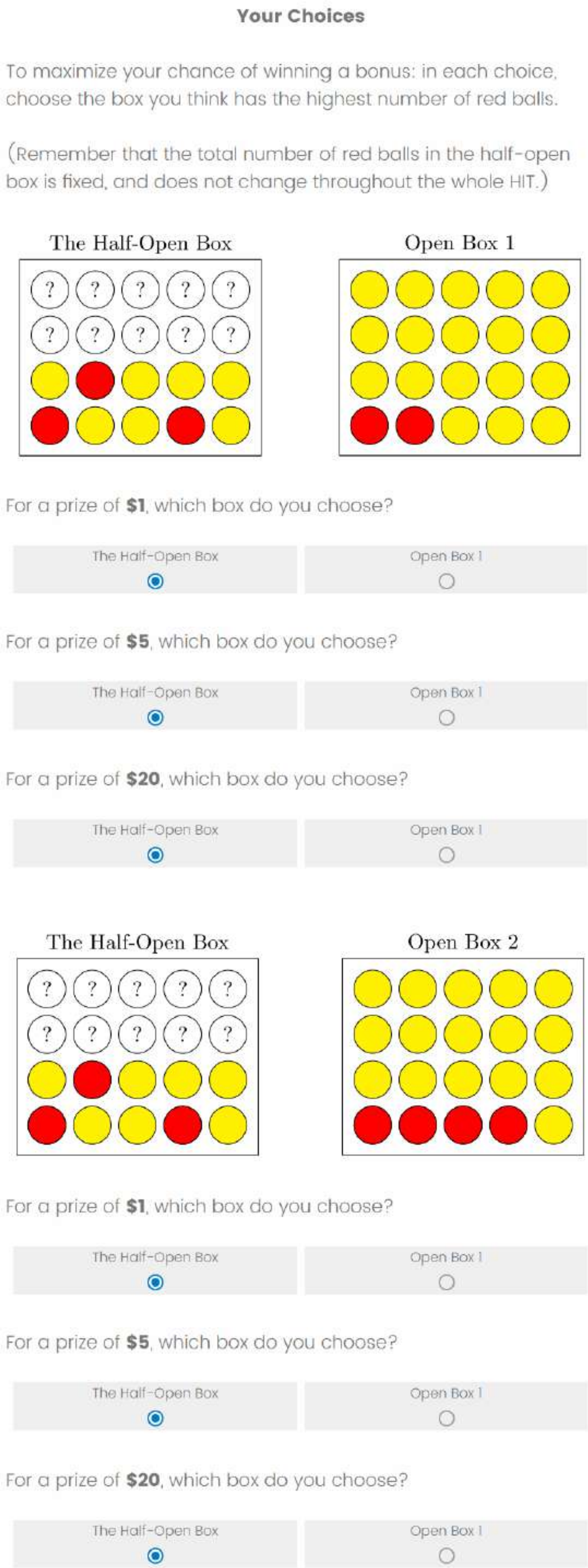


Figure 2: Secondary design. Initially, only the first pair of boxes (The Half-Open Box and Open Box 1) and the first question ("For a prize of $1, which box do you choose?") appear. Once answered, the second question ("For a prize of $5...") appears, and once that is answered, the third ("For a prize of $20...") appears. Once the subject has chosen between that pair of boxes for all three prize levels, the next pair of boxes and the associated $1 question appear. This process continues until all seven pairs of boxes and all 21 questions are on the page. By scrolling, subjects can easily check, compare, and amend any prior choice. Subjects are required to make all 21 choices by selecting one of the two options and there is no time limit. The screenshot (cropped from below) shows an example in which The Half-Open Box has been chosen in preference to both Open Box 1 and Open Box 2, for all prize levels. If, for each pair of boxes, a subject chooses the same box (be it the Half-Open Box or the respective Open Box) for all three prize levels (as the example subject has for the first two pairs), their choices are consistent with Prize Independence.

We present the seven pairs of boxes in order of the number of Red balls in the open boxes, as in our Primary Design. However, we considered it confusing to randomize the order in which the prize levels appear under each pair of boxes so we fixed it for all subjects to the ascending order (as seen for the first two pairs of boxes in Figure 2).

For a given pair of boxes, a subject's choices conform with Prize Independence if they are the same across prize levels. Conversely, they violate Prize Independence if, for a given pair of boxes, the subject chooses the half open box at one prize level and the open box at another. Facing our Secondary Design, consistent choices require a subject to simply select either the box on the left all three times, or the box on the right all three times, for each pair of boxes.

By presenting one pair of boxes at a time, asking subjects to choose between them at all three price levels sequentially, *and* aligning the choices in two columns, the Secondary Design aims to minimize the level of cognitive processing required from subjects, while maintaining engagement with the task. The layout may also play on a number of biases, well known in the psychology literature, all of which work to make choices consistent with Prize Independence. For example, survey respondents often show a desire to appear rational and consistent in their replies (Podsakoff et al., 2003): by prompting them to evaluate each choice in isolation at all three prize levels, the Secondary Design may make this *consistency motif* more salient. Further, respondents often select a point on response scales that appears to be reasonable for the first object, and then rate all of the remaining objects at that point (*common scale format* and *common scale anchor* effects; see, e.g., Podsakoff et al., 2003; Tourangeau et al., 2000). Overall, making choices consistent with Prize Independence is the path of least resistance as it involves selecting three options that are aligned one on top of the next.

In addition, and as in Experiment 1_M, for Experiment 2 we recruited only subjects classified by the subject pool operator to be more skilled and attentive; see Section 2.7 for details.

For all those reasons, we rely on the more neutral Primary Design with the Secondary Design serving as a stress test on the existence of any violations. We believe the percentage of subjects whose choices violate Prize Independence in the Secondary Design may form a *lower bound* of the fraction who do not have subjective probabilities.

2.5 Secondary Outcome Variables

Following the main choice task in each of our experiments, subjects answer five probability competence questions, which we adapted from the "expanded numeracy scale items" of Lipkus et al. (2001, Table 2). These questions allow us to learn whether a subject understands and can work with basic notions of probability. If violations of Prize Independence and poor probability comprehension are highly correlated, it might suggest that they are the result of a failure to understand probabilities. We also include standard demographic questions. For more details, see the transcript in our Appendix.

2.6 Instructions and Comprehension

On the first page of the experiments (after the participation agreement) we ask subjects to *"Please read the following instructions carefully, as they are followed by some comprehension questions."* We explain to them that if *"You get those [comprehension] questions wrong you will have to read the instructions again and re-take the questions."* To progress, subjects have to report they understand these initial messages. The threat of reading the instructions again provides a time incentive (on top of any intrinsic incentive) to pay attention to them.

We split the instructions into five short pages (each containing between 38-143 words; see the transcript in the Appendix for details). This makes the instructions easier to digest and avoids the potential to overwhelm subjects with one long page of text. We then ask them three comprehension questions.[10] If a subject answers exactly one of these questions incorrectly, we take them back to the page of the instructions relevant to that question, and then require them to answer all three questions again. If they answer more than one question incorrectly, we take them through all five pages of instructions again, before requiring them to re-take all three questions. However they answer in the second attempt at the comprehension questions, we allow them to progress (but this is not known to them).

Understanding of the instructions seemed good; 90% of 1,441 subjects answered all questions correctly on their first or second attempt.

[10]We did not include a practice because the scope for learning seems narrow in our experiments. Indeed, pilots with practice choices returned Prize Independence violation rates around 50%, as in our main experiment.

2.7 Subjects

Subjects were recruited via the Amazon Mechanical Turk (MTurk) online platform, commonly used throughout the social sciences including economics and finance (e.g., DellaVigna and Pope, 2017; Kuziemko et al., 2015; Lian et al., 2019). We restricted participation to those in the US, with a good track record (at least 95% of jobs approved), at least some experience (at least 100 jobs completed), and no participation in our pilots. Our employer account on the platform had a favorable reputation at the time of recruitment.[11]

We collected our data over the second half of 2020. In Experiment 1, 1,200 subjects completed the Primary Design; in Experiment 1_M, 120 subjects with MTurk's "Master" qualification completed the Primary Design; and in Experiment 2, 121 Master workers completed the Secondary Design. Master workers make up a small fraction of the total MTurk participant pool, and are generally understood to be an elite group of more attentive and more able workers. No subject was allowed to participate more than once in total. The overall average completion time was 9m 11s. Subjects received $1.00 for participating plus an average of $3.50 in incentive payments based on their choices, corresponding to an average hourly wage of $29.41.[12,13]

Many studies have shown the MTurk population to be more demographically diverse than, and to produce data of a comparable quality to, more traditional participant pools.[14] Demographic information on our subjects is presented in Table A1 in the Appendix. Snowberg and Yariv (2021) compare behavioral measures across representative, MTurk, and student samples. They find levels of some behaviors vary across subject pools with those from their MTurk sample generally sitting between those of their representative and student samples (and more often closer to the representative sample). Perhaps more importantly, they find correlations between behav-

[11]Many MTurk workers use review sites to check employers' reputations (rates of pay, treatment of workers, etc.). Our employer account has favorable reviews and summary statistics from many subjects from various studies prior to the present one. For example, the site TurkerView reports the account pays quickly and "generously", their top category, with no negative reports. Also importantly, workers can see that many subjects completing prior experiments reported they received incentive ("bonus") payments from the account. In addition to the stated commitment to pay in the participation agreement, this should have reduced any concerns surrounding trust.

[12]A task description posted on MTurk for potential subjects shows its participation fee. We added to ours that it takes 5-10 minutes and that bonus payments are available, without a description of the nature of those payments.

[13]This wage is high for MTurk, e.g., the community site TurkerView gives MTurk employer accounts paying average wages above all US minimum wage levels a "green light" in their traffic light system of wage reporting.

[14]See, for example, Chandler et al. (2014); Crump et al. (2013); Paolacci and Chandler (2014). See also the many studies replicating classic experiments across various domains including cognitive psychology (e.g., Goodman et al., 2013; Paolacci et al., 2010) and economics (e.g., Horton et al., 2011).

iors to be similar across all samples. A disadvantage of MTurk subjects' data they highlight is that they can be noisier than data from other subject populations. This suggests that researchers using MTurk subjects may wish to collect larger samples to detect behaviors. In view of that, to provide a reasonably precise estimate of any Prize Independence violation rate, and to ensure we would have a large sample even when focusing on subsets of subjects satisfying increasingly stringent criteria, we collected data from 1,441 subjects.

The COVID-19 pandemic presents challenges for conducting social sciences research experiments. In addition to a widespread suspension of the operations of physical laboratories, various factors including falling employment might have led to an uptick in the use of online subject platforms. In terms of the effects on MTurk, Arechar and Rand (2021) report such an increase in workers on the platform and investigate the characteristics of those who signed up after the pandemic's effects were felt in the US. They find the pool of newcomers to be more diverse and representative but also less attentive than those who joined before the pandemic. We anticipated that the pandemic may have an impact on subject characteristics and included a question on whether they joined pre or post February 1, 2020.[15] In line with Arechar and Rand (2021), we find evidence that workers are indeed more likely to provide "noisier" data (e.g., inattentive or dominated choices) if they joined after that date.[16] However, after applying our most basic competence criteria, we find no effect of subjects' sign-up date on our main outcome measure.[17]

2.8 Pre-Trial Registration

The experiments were registered in advance in the AEA RCT Registry (Ronayne et al., 2020). There, we provided an overview of the experimental design and sample size, and described our primary measure to be the proportion of subjects whose choices violate Prize Independence. We also stated our secondary measures: some of the probability competence questions of Lipkus et al. (2001) and various demographics. All these measures are reported or included in our analyses and no other measures were registered.

[15]The effects of the pandemic had hardly been felt in the US prior to that date (e.g., there had been no restrictions within the US) and the World Health Organization only issued a Global Health Emergency the day before. February 1 is also before the surge in MTurk participation identified by Arechar and Rand (2021), which began mid-March.

[16]For example, among those joining after February 1, 2020, 42% make a dominated choice in our experiments, while among those joining before that date 28% make a dominated choice (d = 0.14; p < 0.001).

[17]After applying criteria A-C (see Table 1) across all our subjects, 44.8% of those who joined before that date make choices that violated Prize Independence vs 44.7% of those who joined after that date (d = 0.001, p = 0.988). Also see the insignificant coefficients on the join date dummy variable in the estimates of Table A6 in the Appendix.

3 Data

3.1 Primary Design

Experiment 1. We find 67% of subjects have switch points that differed across prize levels and thus make choices that violate Prize Independence. *Prima facie*, this casts substantial doubt on the empirical validity of Anscombe and Aumann (1963) and most of the literature on subjective probabilities. However, these violations are not particularly meaningful if due to inattention or basic irrationality, as suggested by the following (non mutually exclusive) behaviors:

- 46/1,200 subjects made all their choices in less than 30 seconds.
- 458/1,200 subjects made at least one dominated choice.
- 489/1,200 subjects have more than one switch point *within* a prize level.

Subjects who make all their choices in less than 30 seconds seem inattentive. Rational subjects should not make dominated choices or have more than one switch point within a prize level (Section 2.3). While these subjects do make choices that violate Prize Independence, these violations are perhaps not very interesting. In the rest of our analysis we therefore focus on subjects who seem attentive and whose choices are consistent with basic principles of rationality.

We proceed to assess the empirical validity of Prize Independence by focusing on subjects that satisfy increasingly stringent competence criteria, as shown in Table 1.

Table 1: Competence Criteria

Key	Criterion
A	Spend at least 30 seconds on choices
B	Make no dominated choices
C	Have exactly one switch point within each prize level
D	Answer all comprehension questions correctly on second attempt
E	Answer all comprehension questions correctly on first attempt
F	Answer all probability competence questions correctly
G	Have a post-graduate (Master's, Professional, or Doctorate) degree

The top graph in Figure 3 shows the percent of subjects in Experiment 1 whose choices violate Prize Independence. We find that 47% of subjects who satisfy the basic attentiveness and rationality criteria, A-C, make choices that violate Prize Independence. Remarkably, this percentage is *virtually unchanged* even when additional competence criteria D, E, F are applied. Indeed, even when we restrict attention to those who satisfy the competence criteria A-F *and* have a post-graduate degree (criterion G), the violation rate of Prize Independence is 40%.[18]

We take these data to imply that about half of subjects do not have subjective probabilities.

Experiment 1_M. The 120 subjects in Experiment 1_M also faced the Primary Design. Because they are all Master workers, we expect larger fractions to pass the competence criteria of Table 1, especially up to F, and that is indeed the case, for example, and most notably, 116/120 pass A-C.

Should violations of Prize Independence by the choices of those in Experiment 1 who meet our most basic criteria (A-C) still be due to inattention, we would expect to see lower violation rates among the choices of the Master workers of Experiment 1_M. But we do not. Comparing the two histograms of Figure 3, once A-C are applied we see only negligible differences in the percent of subjects whose choices violate Prize Independence, at every level of competence.[19]

Overall, we find that a large proportion of subjects make choices that violate Prize Independence. While some of these violations may be ascribed to inattention or irrationality, approximately half of subjects who appear attentive, make no irrational choices, display a good comprehension of the task, and possess a good understanding of basic notions of probability make choices that violate Prize Independence, even if they have achieved Master worker status. There is also no evidence that having a post-graduate degree makes a difference.

3.2 Secondary Design

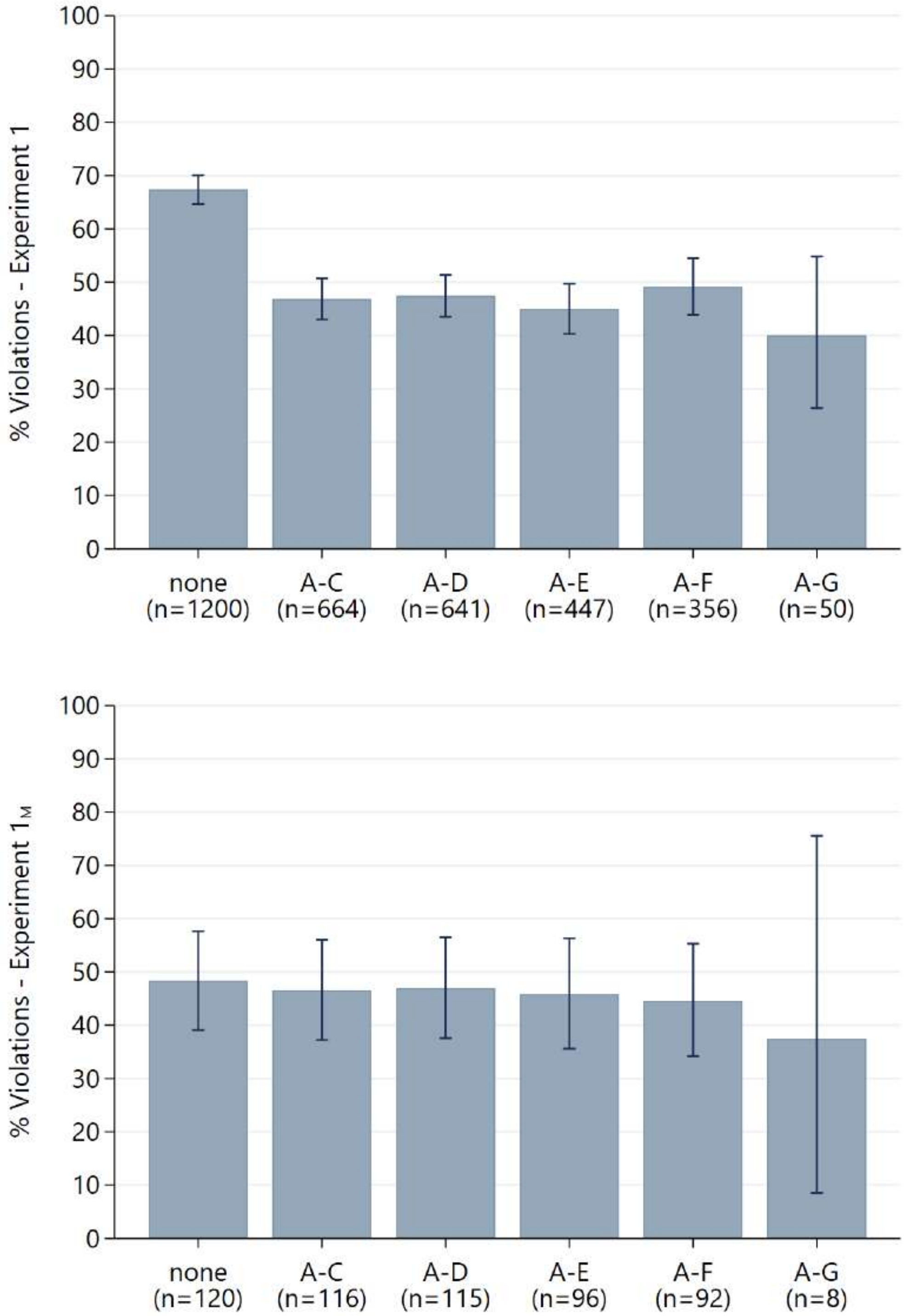
Experiment 2. We set up this design to make it easier for subjects to make choices consistent with Prize Independence (see Figure 2 and Section 2.4). To make consistent choices a subject selects either the box on the left all three times (once for each prize level), or the box on the right all three times (choices are aligned, one on top of the other), for each pair of boxes.

Our main conclusions rely on the more neutral Primary Design. However, by stacking the odds *against* finding any violations, Experiment 2 provides a more stringent test, almost a sanity

[18]The criteria of Table 1 are ordered in the way we think best represents an increasing level of competence. For *ceteris paribus* interpretations of those criteria and other variables, we refer the reader to the estimates in Table A6 and related text in the Appendix.

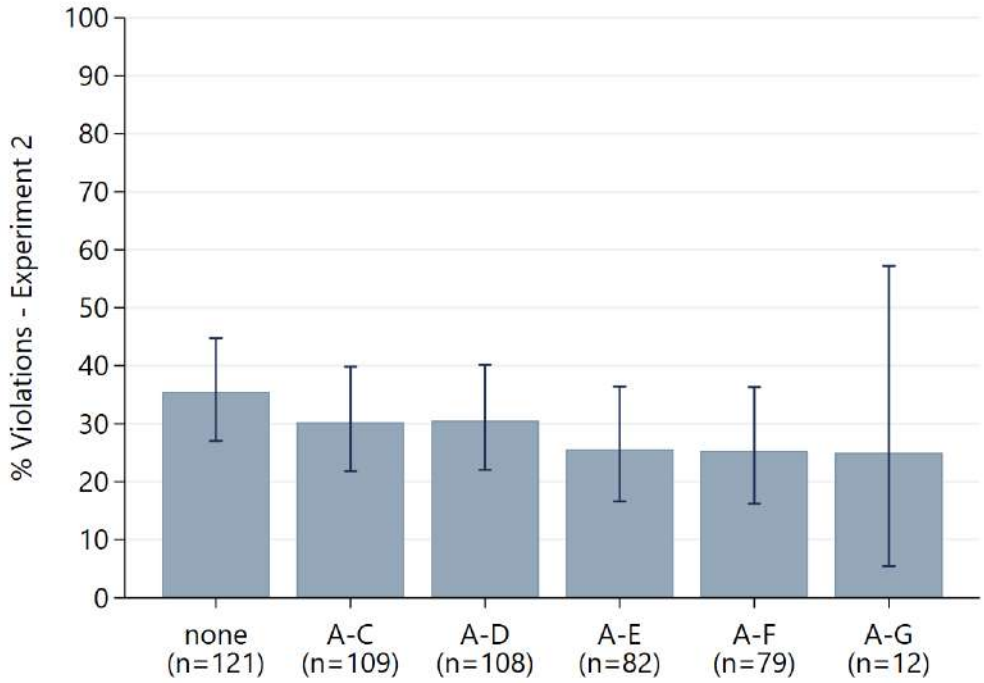
[19]The five corresponding tests for a difference in proportions, bar by bar, across the two samples are insignificant (Experiment 1 % – Experiment 1_M% = $0.3, 0.5, -0.9, 4.6, 2.5$; $p = 0.95, 0.93, 0.88, 0.43, 0.89$).

Figure 3: Violations of Prize Independence by Competence Criteria - Primary Design



Each bar has a 95% confidence interval. These data are presented in tabular form in Tables A2 and A3 in the Appendix. We detail the increasingly restrictive criteria, A to G, in Table 1. We require subjects in Experiment 1_M to have earnt Amazon MTurk's Master worker qualification. We impose no such restriction in Experiment 1.

Figure 4: Violations of Prize Independence by Competence Criteria - Secondary Design



Each bar has a 95% confidence interval. These data are presented in tabular form in Table A4 in the Appendix. We detail the increasingly restrictive criteria, A to G, in Table 1. Subjects in Experiment 2 faced our Secondary Design and had Amazon MTurk's Master worker qualification.

check, of the data from Experiments 1 and 1_M. As we have noted, we intended the Secondary Design to suggest to subjects that their choices should conform with Prize Independence, and the subjects who faced the Secondary Design were exclusively Master Workers, so it seems natural to expect many fewer violations of Prize Independence; indeed, it would not seem shocking if there were *almost no* violations of Prize Independence.

However, the data, presented in Figure 4 are rather consistent with the data from the Primary Design. Although the rate of violations of Prize Independence is clearly lower – compare with Figure 3 – it is still the case that 25-30% make choices that violate Prize Independence.[20]

3.3 Patterns of Violation

Our intent is to test the empirical validity of Prize Independence, and thus of the main definitions of subjective probability in the literature. Hence, we focus on the proportion of subjects whose

[20]Five tests for a difference in proportions, bar by bar, across Experiments 1_M & 2 confirm the differences are significant save the last (Experiment $1_M\%$ − Experiment 2 % $= 16.3, 16.4, 20.2, 19.2, 12.5$; $p < 0.01$ for the first four, and $p = 0.55$ for the last; the analogous tests comparing Experiments 1 & 2 give the same conclusions).

choices violate Prize Independence. However, one might wonder whether any patterns can be detected in the choices which do not conform with Prize Independence. The remainder of this section addresses this issue. The purpose of our analysis is not to provide an explanation of the violations of Prize Independence but rather to detect any stylized facts.

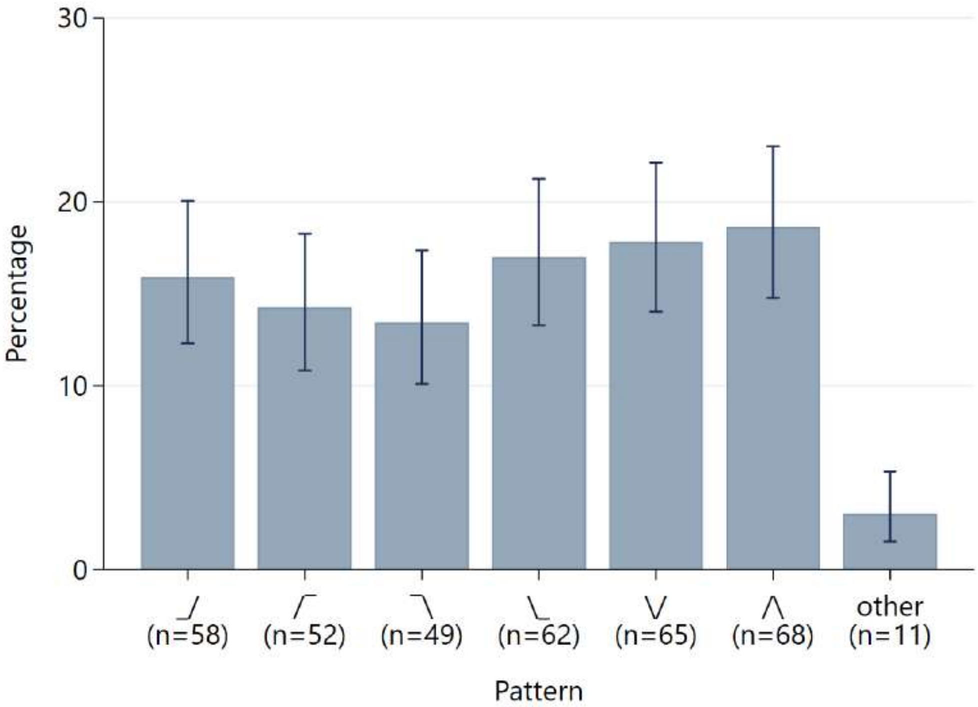
For each prize level, subjects face seven choices of which two had a dominant option (Figure 1). Ignoring those leaves us with six possible switch points, which we measure by the number of times a subject chose the half-open box: 1, 2, ..., or 6. In total, each subject's choices generate three switch points, one per prize level: \$1, \$5, \$20, which we denote by s_1, s_5, and s_{20}, respectively, with $s_1, s_5, s_{20} \in \{1, 2, \ldots, 6\}$.[21] If a subject's choices are consistent with Prize Independence, their switch points are identical, i.e., $s_1 = s_5 = s_{20}$.

We now examine how subjects' switch points vary as the prize level increases. We categorize subjects by graphing their switch points in the following way: assigning *prize level* to the x-axis and *switch point* to the y-axis, we plot a subject's data (three points), and draw straight lines to connect adjacent points. For example, consider a subject whose choices violate Prize Independence with switch points $s_1 = s_5 < s_{20}$. Their graph would appear flat and then upward sloping. We categorize such subjects by assigning them a symbol to represent the shape of the graph just described, "_/". The first bar of the histogram shown in Figure 5 reports the percent of subjects whose data fall into that category. The remaining bars of Figure 5 give the percentage of subjects in six similarly defined categories.

Figure 5 reveals a rich heterogeneity in the relationship between switch points and prize levels, and hence in what we term *the patterns of violation*. For example, there are a similar number of subjects with switch points that are weakly increasing (the first two bars), weakly decreasing (the second two), and non-monotonic (the third two) in prize level. Overall, there is no clear and common pattern to violations across subjects. This absence of a common pattern suggests the absence of a unifying theory. In the next section, we discuss several possible theories and their failure to explain our data.

[21]We continue to look at the data of those who faced our Primary Design (Experiments 1 and 1_M) and meet our basic criteria A-C, which leaves only subjects with a unique switch point per prize level. Data on the frequencies of each of the six switch points are presented in Table A5 in the Appendix.

Figure 5: Patterns of Violation



Subjects in Experiments 1 and 1_M whose choices satisfy A-C of Table 1 and violate Prize Independence (n=365). The seven pattern categories on the x-axis are defined as follows. For any subject, denote by $s_1, s_5, s_{20} \in \{1, 2, \ldots, 6\}$ their switch points when facing the \$1, \$5, \$20 prize levels, respectively (choices such that $s_1 = s_5 = s_{20}$ are consistent with Prize Independence). If their choices are such that $s_1 = s_5 < s_{20}$ we categorize their data by the left-most category shown. Similarly: if $s_1 < s_5 = s_{20}$, the second; $s_1 = s_5 > s_{20}$, the third; $s_1 > s_5 = s_{20}$, the fourth; $s_1 > s_5$ and $s_5 < s_{20}$, the fifth; $s_1 < s_5$ and $s_5 > s_{20}$, the sixth; $s_1 < s_5 < s_{20}$ and $s_1 > s_5 > s_{20}$, the last.

4 Discussion

The choices of a sizable fraction of subjects violate Prize Independence. Whatever the determinants of their behavior, it seems to us – and to much of the literature – that subjects whose choices violate Prize Independence *cannot* be said to have subjective probabilities. Although it is beyond the scope of this paper to explain *why* this is the case or what Decision Makers may do instead, we address below the extent to which various theories are able or unable to explain our data.

4.1 Mistakes, Inattention, Irrationality

The most obvious explanations for the observed violations of Prize Independence are that subjects make mistakes, are inattentive, or are somehow irrational. While this may explain *some* of the violations, and specifically those of subjects who do not satisfy competence criteria A-C, we believe that our experimental design and various checks argue against all of these possibilities for the majority of subjects whose choices violate Prize Independence.

In our design, subjects choose between boxes which offer different probabilities of drawing a Red ball. A subject interested in maximizing their payoff faces a simple task: in each choice, they should pick the box they believe has more Red balls in it, as we explicitly summarize for them on the same page that they make their choices. We do not require subjects to perform any calculations. We believe our design minimizes potential issues with the way in which a subject might internally represent their beliefs in numerical form – a much debated issue in cognitive psychology; see Manski (2004), for example. The straightforwardness of our design from the subject's perspective perhaps contrasts most starkly with the Becker et al. (1964) elicitation mechanism, which "is based on the intuition that if you want someone else to make a choice for you, then you should tell them your true preferences so that they can make the best choice [... which] is intuitive but complicated to implement" (Holt and Smith, 2009, pp.126-7).

Holt and Smith (2009) also point to a trade-off faced by experimental researchers on belief-updating between adopting detailed hypothetical scenarios versus more abstract designs ("book bags and poker chips") in which subjects can observe draws directly: "[hypothetical scenarios may] help subjects process complex information in logic-based inference tasks [... but they can affect] subjects' perceptions of the reliability of the test" (Holt and Smith, 2009, p.126).

Although we are not concerned with updating, and so do not require subjects to see any draws, we avoid this trade-off: we chose abstract but simple and unchanging choice objects that do not require or obviously prompt any complex mental work.

In both of our designs, after making their initial choices, subjects have the opportunity to check, compare, and amend or correct their choices. If a subject's choices were inconsistent across prize levels, and that inconsistency were the result of a mistake rather than a deliberate choice, they could correct the mistake and make their choices consistent. That so many violations persist would seem to indicate that they cannot be explained simply by mistakes. This is even more evident in the case of the Secondary Design, the structure of which is explicitly designed to make it *easier* for subjects to make choices consistent with Prize Independence.

The idea of analyzing choices in terms of switch points within lists of choices builds on the classic design of Holt and Laury (2002) whose aim was to elicit risk preferences. Charness et al. (2013) provide a review of this "multiple price list" method and highlight its potential weaknesses including subjects having multiple switching points or making "backwards" choices (in the context of our Figure 1, such a subject would start by choosing the open boxes and then switch to the half-open box, instead of the other way around). Subjects making either of those sorts of choices in our design violate standard definitions of subjective probability, but perhaps not in very interesting ways that may be more indicative of a lack of attention, understanding, or basic rationality. Our conclusion that almost half of subjects make choices that violate Prize Independence comes *after* excluding subjects whose choices suggest inattention, or fail basic criteria of rationality. Were we to include these subjects, we would conclude that around two-thirds of subjects make choices that violate Prize Independence.

Moreover, once we exclude subjects who are inattentive or whose choices violate basic rationality, the proportion of choices violating Prize Independence shows a remarkable invariance to our increasingly stringent criteria, and the requirement of subjects to have Master worker status.

4.2 Violations of Expected Utility

Despite the title of their paper, Anscombe and Aumann (1963) do not only provide a definition of subjective probability, but also provide axioms that imply a Decision Maker maximizes expected utility with respect to objective or subjective probability.[22] This may suggest an alternative explanation of our results because choices that violate Prize Independence must also violate one or more of the axioms in Anscombe and Aumann. In particular, subjects whose choices violate Prize Independence might not maximize expected utility – even for lotteries with objective probabilities. There is indeed substantial experimental evidence suggesting that many subjects do not maximize expected utility.

However, it seems to us that *any* sensible notion of subjective probability must satisfy Prize Independence. As an example, Machina and Schmeidler (1992) build on the classical framework and characterize "a probabilistically sophisticated individual, that is, one who possesses a subjective probability measure ... over events and evaluates acts on the basis of a non-expected preference function V over their implied probability distributions over outcomes" (Machina and Schmeidler, 1992, p.766). They drop the "Sure-Thing Principle" but retain Savage's "Eventwise Monotonicity" axiom P3 and strengthen P4 into a "Strong Comparative Probability" axiom. As a result, probabilistically sophisticated individuals hold subjective probabilities that are prize independent, even if they are *not* expected utility maximizers.

Similarly, consider two other prominent approaches to choice under uncertainty that depart from expected utility theory, namely the rank-dependent utility model (Quiggin, 1982) and cumulative prospect theory (Tversky and Kahneman, 1992). Both models reject the linearity of the value function in probabilities characteristic of expected utility theory and introduce a nonlinear probability transformation through a probability weighting function. In both approaches, the weighting function is assumed to capture the Decision Maker's intrinsic attitude towards probabilities, so that other aspects of the risky situation faced by the Decision Maker – and in particular, outcomes, which are evaluated via a utility or value function – should not affect this attitude. Hence, in these approaches too, Prize Independence holds.

Thus, it seems to us that choices that violate Prize Independence cannot be satisfactorily explained on the grounds that subjects do not maximize expected utility.

[22]Similarly, Savage's (1972 [1954]) axioms jointly characterize a utility function and a probability measure. However, objective probabilities play no role in Savage's theory and a Decision Maker maximizes expected utility with respect to subjective probability only.

4.3 State Dependent Preferences

Most of the literature on subjective probability assumes that the Decision Maker's utility function is state independent, but one can certainly imagine circumstances in which this would not be the case. For example, as Karni (2014, p.20) writes, "...loss of ability to work may affect a decision maker's willingness to take financial risks; a leg injury may reverse his/her preferences between going hiking and attending a concert." Similarly, the Decision Maker's marginal utility for the prize offered might be quite different in a state in which the Decision Maker is wealthy and in a state in which the Decision Maker is poor. As Karni and Safra (1995) point out, if the Decision Maker's utility function is state-dependent, then the Decision Maker's assessments of risk may also violate Prize Independence. Moreover, as Karni (2014, pp.18–19) writes, "...the utility function and probability measure that figure in the representations in Savage's theorem and in the theorem of Anscombe and Aumann are jointly unique (that is, the probability is unique given the utility and the utility is unique, up to a positive affine transformations, given the probability). Thus, the uniqueness of the probabilities in the aforementioned theorems depends crucially on assigning the consequences utility values that are independent of the underlying events."

While these considerations are relevant at a general, theoretical level, it is hard to imagine how they are relevant in our experiments; it seems implausible that a subject's wealth (aside from the prize) or their utility function differs according to the color of the ball drawn from a box.

4.4 Ambiguity and Ambiguity Aversion

The number of Red balls in the half-open box is not known, and so is *ambiguous*. It is therefore legitimate to wonder whether individual attitudes towards ambiguity may explain our data. To understand the extent to which this might be the case, we consider the implications of the two leading models of ambiguity aversion.[23]

The first of these models is the α-MEU (Maxmin Expected Utility) model of Ghirardato et al. (2004), which is a generalization of the minimax model of Gilboa and Schmeidler (1989). To

[23]Because ambiguity aversion is frequently offered as an explanation of observed "paradoxical" behavior in the Ellsberg (1961) experiment, it seems worth noting that ambiguity plays a different role in our experiments than in the Ellsberg experiment. In our experiments, the lotteries are fixed but the prizes are changing; in the Ellsberg experiment, the prizes are fixed but the lotteries are changing.

apply the α-MEU model in our context, suppose the monetary prize is P and the subject's utility for monetary prizes is $u(\cdot)$; without loss of generality, normalize so that $u(0) = 0$.

First consider an open box with R Red balls and $Y = 20 - R$ Yellow balls, which yields the probability of success $R/20$. Now consider a half-open box; as in our experiments suppose that the composition of the 10 visible balls is 3 Red to 7 Yellow. A subject who assigns a subjective probability of π to drawing a Red ball from this box would presumably be indifferent between the half-open box and an open box with 20π Red balls and $20(1-\pi)$ Yellow balls. If such a subject maximizes expected utility, they will assign utility $U_h = \pi u(P)$ to the half-open box and utility $U_o = (20\pi/20)u(P)$ to the open box with 20π Red balls. Of course, $U_h = U_o$, independently of the prize P.

The α-MEU model of ambiguity aversion posits that a Decision Maker does not assign a single subjective probability to an ambiguous event, but an *interval* $[A, B]$ of possible probabilities. In our case, the true composition of the half-open box is somewhere between 3 Red to 17 Yellow and 13 Red to 7 Yellow, so presumably such a Decision Maker would assign some interval $[A, B]$ with $0.15 \leq A \leq B \leq 0.65$; i.e., they would believe the probability of drawing a Red ball from the half-open box would be somewhere between A and B. According to the α-MEU model, an ambiguity averse (or ambiguity loving) agent assigns utility to the half-open box of

$$\begin{aligned}
U_\alpha &= \alpha \min_{\pi \in [A,B]} [\pi u(P) + (1-\pi)u(0)] + (1-\alpha) \max_{\pi \in [A,B]} [\pi u(P) + (1-\pi)u(0)] \\
&= \alpha(Au(P) + (1-\alpha)Bu(P) \\
&= [\alpha A + (1-\alpha)B]u(P),
\end{aligned}$$

where α is the agent's degree of ambiguity aversion: $\alpha = 1$ corresponds to pure (Gilboa-Schmeidler) ambiguity aversion; $\alpha = 0$ to pure ambiguity loving. The Decision Maker would be indifferent between the half-open box and an open box with $20\,[\alpha A + (1-\alpha)B]$ Red balls and $20\,(1 - [\alpha A + (1-\alpha)B])$ Yellow balls. Thus, α-MEU ambiguity aversion would lead the Decision Maker to choose *as if* they were mixing the probabilities A, B in fixed proportion, *independently of the prize*. As such, α-MEU ambiguity aversion does *not* lead to different choices for different prizes and so does not explain our data.[24]

[24]We note that many other models of ambiguity aversion such as Schmeidler's (1989) Choquet expected utility and Maccheroni et al.'s (2006) variational preferences also imply Prize Independence and so do not explain our data.

It might be said that a Decision Maker who assigns an interval of probabilities, rather than a single point probability does not have subjective probabilities. As we have noted, however, the *choices* of such a Decision Maker would be precisely those of a Decision Maker who *does have* subjective probabilities – according to the definition given by Anscombe and Aumann (1963). Put differently: an α-MEU Decision Maker chooses *as if* they have subjective probabilities.

The second model we address is the smooth ambiguity model of Klibanoff et al. (2005). There, the Decision Maker assigns an interval of probabilities $[A, B]$, a probability distribution μ over that interval, and a function ϕ that represents the Decision Maker's attitude toward ambiguity (unknown probabilities), in the same way that the function u represents the Decision Maker's attitude toward risk (known probabilities). Given the function u, the interval of probabilities $[A, B]$, the beliefs μ, and the function ϕ, the Decision Maker's utility for the half-open box is

$$U = E_{\mu}\Big[\phi\Big(E_{\lambda}[u(P)]\Big)\Big] = \int_{A}^{B} \phi\Big(\lambda u(P)\Big)\, d\mu(\lambda).$$

This model has no predictive value in our context, and so cannot be said to provide a meaningful explanation of our data. To see why, note first, that the α-MEU model has only two parameters: the interval $[A, B]$ and the degree of ambiguity aversion, α, which is a real number in the interval $[0, 1]$. By contrast, the smooth ambiguity model has many parameters: the interval $[A, B]$, the belief μ, which can be an arbitrary probability distribution over $[A, B]$, and the ambiguity aversion function ϕ, which can be an arbitrary function. Hence, the smooth ambiguity model has infinite degrees of freedom. It might therefore not be a surprise that the smooth model makes no prediction whatsoever about Prize Independence. In this sense, the model is not inconsistent with the data, because (in our experiment) it would be possible to choose the belief μ and the ambiguity aversion function ϕ to rationalize *any* choices by a Decision Maker, including choices that display the *same* switch points for each of the three prize levels, *and* those that display different switch points for each of the three prize levels (in any order).

Consider again our earlier example and let $[A, B] = [0.15, 0.65]$. It is not difficult to show that even if we insist that the Decision Maker be risk neutral (that u be linear) and have the simplest, two-point beliefs (that the support of μ be the endpoints 0.15 and 0.65), it is possible to choose ϕ to derive patterns of behavior consistent with *every pattern* we see in our data. For example:[25]

[25]The calculations supporting the following claims are not particularly elightening and are omitted for space. They are available from the authors upon request.

(a) If ϕ is a CRRA function then choices will conform with Prize Independence.

(b) If ϕ is piecewise linear and *concave*, but has kinks at carefully chosen points, then choices will violate Prize Independence. Moreover, every monotonic pattern of violation (as shown in Figure 5) can be achieved by an appropriate choice of such a ϕ.[26]

What is more, it is straightforward to show that there exists no systematic connection between the Decision Maker's attitudes towards ambiguity (the function ϕ) and the relation between prizes and switch points. As the prize level increases, *both* ambiguity averse *and* ambiguity loving subjects can switch to open boxes either earlier, or later. By way of example, consider again the case with $[A, B] = [0.15, 0.65]$ and a risk-neutral Decision Maker with two-point beliefs. It it possible to show that:

(c) If ϕ is piecewise linear and *convex*, but has kinks at carefully chosen points, then choices will violate Prize Independence and every pattern of violation (as shown in Figure 5) can again be achieved by an appropriate choice of such a ϕ.

In sum, although the smooth ambiguity model has sufficient degrees of freedom to fit all patterns of behavior observed in our experiment, it cannot be said to *explain* our data. And, in any case, a Decision Maker who conforms to that model cannot be said to have subjective probabilities.

[26]With slightly more effort, the same patterns of violation can be achieved with a smooth ϕ. More exotic choices of ϕ would allow one to fit also the non-monotonic patterns of behavior in Figure 5.

References

Anscombe, Francis J and Robert J Aumann (1963), "A definition of subjective probability." *Annals of Mathematical Statistics*, 34, 199–205.

Arechar, Antonio A and David G Rand (2021), "Turking in the time of COVID." *Behavior Research Methods*, 53, 2591–2595.

Armantier, Olivier and Nicolas Treich (2013), "Eliciting beliefs: Proper scoring rules, incentives, stakes and hedging." *European Economic Review*, 62, 17–40.

Becker, Gordon M., Morris H. Degroot, and Jacob Marschak (1964), "Measuring utility by a single-response sequential method." *Behavioral Science*, 9, 226–232.

Chandler, Jesse, Pam Mueller, and Gabriele Paolacci (2014), "Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers." *Behavior Research Methods*, 46, 112–130.

Charness, Gary, Uri Gneezy, and Alex Imas (2013), "Experimental methods: Eliciting risk preferences." *Journal of Economic Behavior & Organization*, 87, 43–51.

Crump, Matthew JC, John V McDonnell, and Todd M Gureckis (2013), "Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research." *PloS One*, 8, e57410.

De Finetti, Bruno (1974), *Theory of Probability, Vol. 1*. Wiley, New York, USA.

DellaVigna, Stefano and Devin Pope (2017), "What motivates effort? Evidence and expert forecasts." *Review of Economic Studies*, 85, 1029–1069.

Ellsberg, Daniel (1961), "Risk, ambiguity, and the Savage axioms." *Quarterly Journal of Economics*, 643–669.

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci (2004), "Differentiating ambiguity and ambiguity attitude." *Journal of Economic Theory*, 118, 133–173.

Gilboa, Itzhak and David Schmeidler (1989), "Maxmin expected utility with non-unique prior." *Journal of Mathematical Economics*, 18, 141–153.

Goodman, Joseph K, Cynthia E 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.

Hao, Li and Daniel Houser (2012), “Belief elicitation in the presence of naïve respondents: An experimental study.” *Journal of Risk and Uncertainty*, 44, 161–180.

Holt, Charles A. and Susan K. Laury (2002), “Risk aversion and incentive effects.” *American Economic Review*, 92, 1644–1655.

Holt, Charles A. and Angela M. Smith (2009), “An update on Bayesian updating.” *Journal of Economic Behavior & Organization*, 69, 125–134.

Horton, John J, David G Rand, and Richard J Zeckhauser (2011), “The online laboratory: Conducting experiments in a real labor market.” *Experimental Economics*, 14, 399–425.

Kadane, Joseph B. and Robert L. Winkler (1988), “Separating probability elicitation from utilities.” *Journal of the American Statistical Association*, 83, 357–363.

Karni, Edi (2009), “A mechanism for eliciting probabilities.” *Econometrica*, 77, 603–606.

Karni, Edi (2014), “Axiomatic foundations of expected utility and subjective probability.” In *Handbook of the Economics of Risk and Uncertainty* (Mark J. Machina and W. Kip Viscusi, eds.), volume 1, 1–39, North Holland, Oxford, UK.

Karni, Edi, Fabio A. Maccheroni, and Massimo Marinacci (2015), “Ambiguity and nonexpected utility.” In *Handbook of Game Theory with Economic Applications* (H. Peyton Young and Shmuel Zamir, eds.), volume 4, 901–947, North Holland, Amsterdam.

Karni, Edi and Zvi Safra (1995), “The impossibility of experimental elicitation of subjective probabilities.” *Theory and Decision*, 38, 313–320.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji (2005), “A smooth model of decision making under ambiguity.” *Econometrica*, 73, 1849–1892.

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva (2015), “How elastic are preferences for redistribution? Evidence from randomized survey experiments.” *American Economic Review*, 105, 1478–1508.

Lian, Chen, Yueran Ma, and Carmen Wang (2019), "Low interest rates and risk-taking: Evidence from individual investment decisions." *Review of Financial Studies*, 32, 2107–2148.

Lipkus, Isaac M, Greg Samsa, and Barbara K Rimer (2001), "General performance on a numeracy scale among highly educated samples." *Medical Decision Making*, 21, 37–44.

Maccheroni, Fabio A., Massimo Marinacci, and Rustichini Aldo (2006), "Ambiguity aversion, robustness, and the variational representation of preferences." *Econometrica*, 74, 1447–1498.

Machina, Mark J. and David Schmeidler (1992), "A more robust definition of subjective probability." *Econometrica*, 60, 745–780.

Manski, Charles F. (2004), "Measuring expectations." *Econometrica*, 72, 1329–1376.

Paolacci, Gabriele and Jesse Chandler (2014), "Inside the Turk: Understanding Mechanical Turk as a participant pool." *Current Directions in Psychological Science*, 23, 184–188.

Paolacci, Gabriele, Jesse Chandler, and Panagiotis G Ipeirotis (2010), "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making*, 5, 411–419.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff (2003), "Common method biases in behavioral research: A critical review of the literature and recommended remedies." *Journal of Applied Psychology*, 88, 879–903.

Quiggin, John (1982), "A theory of anticipated utility." *Journal of Economic Behavior and Organization*, 3, 323–343.

Ronayne, David, Roberto Veneziani, and William R Zame (2020), "Subjective probability and prizes." *AEA RCT Registry*, URL `https://doi.org/10.1257/rct.6174-2.0`.

Savage, Leonard J. (1971), "Elicitation of personal probabilities and expectations." *Journal of the American Statistical Association*, 66, 783–801.

Savage, Leonard J (1972 [1954]), *The Foundations of Statistics, 2nd ed.* Dover Publications, New York, USA.

Schmeidler, David (1989), "Subjective probability and expected utility without additivity." *Econometrica*, 57, 571–587.

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

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski (2000), *The psychology of survey response*. Cambridge University Press, Cambridge, UK.

Tversky, Amos and Daniel Kahneman (1992), "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and Uncertainty*, 5, 297–323.

Appendix

Table A1: Data collection information and subject demographics

	main batch	batch 2	batch 3	total
Data collection dates (2020)	Aug 3-6	Sep 21	Dec 26-28	
Overall implied average hourly wage	$28.61	$34.23	$33.76	$29.41
Amazon MTurk's Master qualification required	no	yes	yes	
Gender				
Male	772 (64)	63 (53)	60 (50)	895 (62)
Female	425 (35)	56 (47)	61 (50)	542 (38)
Other / Prefer not to say	3 (0)	1 (0)	0 (0)	4 (0)
Age, mean years [sd]	36.7 [10.5]	41.3 [9.6]	44.6 [12.4]	37.8 [10.8]
Race				
White	794 (66)	99 (83)	92 (76)	985 (68)
Black or African American	288 (24)	9 (8)	10 (8)	307 (21)
Hispanic or Latino	48 (4)	6 (5)	3 (2)	57 (4)
American Indian or Alaska Native	8 (1)	0 (0)	1 (1)	9 (1)
Asian American	51 (4)	5 (4)	10 (8)	66 (5)
Native Hawaiian or Pacific Islander	0 (0)	0 (0)	0 (0)	0 (0)
Other / Prefer not to say	11 (1)	1 (1)	5 (4)	14 (1)
Mother tongue English	1191 (99)	120 (100)	121 (100)	1432 (99)
Income (household; pre-tax; ’000 USD)				
0 – 9.999	33 (3)	3 (3)	3 (2)	39 (3)
10 – 19.999	86 (7)	16 (13)	11 (9)	113 (8)
20 – 29.999	127 (11)	19 (16)	12 (10)	158 (11)
30 – 39.999	140 (12)	15 (13)	19 (16)	174 (12)
40 – 49.999	195 (16)	10 (8)	12 (10)	217 (15)
50 – 59.999	219 (18)	8 (7)	16 (13)	243 (17)
60 – 69.999	107 (9)	17 (14)	14 (12)	138 (10)
70 – 79.999	101 (8)	6 (5)	11 (9)	118 (8)
80 – 89.999	40 (3)	4 (3)	5 (4)	49 (3)
90 – 99.999	59 (5)	3 (3)	4 (3)	66 (5)
100 – 124.999	42 (4)	10 (8)	7 (6)	59 (4)
125 – 149.999	28 (2)	5 (4)	4 (3)	37 (3)
150+	23 (2)	4 (3)	3 (2)	30 (2)
Education				
Kindergarten and elementary (grades 1-8)	1 (0)	0 (0)	0 (0)	1 (0)
High school (grades 9-12, no degree)	13 (1)	4 (3)	3 (2)	20 (1)
High school graduate (or equivalent)	50 (4)	16 (13)	14 (12)	80 (6)
Some college (1-4 years, no degree)	167 (14)	35 (29)	32 (26)	234 (16)
Bachelor's degree (BA, BS, AB, etc)	671 (56)	55 (46)	56 (46)	782 (54)
Master's degree (MA, MS, MENG, MSW, etc)	285 (24)	9 (8)	15 (12)	309 (21)
Professional school degree (MD, DDC, JD, etc)	8 (1)	1 (1)	0 (0)	9 (1)
Doctorate degree (PhD, EdD, etc)	5 (0)	0 (0)	1 (1)	6 (0)
Political Affiliation[a] [sd]	59 [32]	35 [30]	33 [31]	55 [33]
MTurk join date before February 1, 2020	799 (67)	120 (100)	120 (99)	1039 (72)
N	1200	120	121	1441

Frequencies; (% within characteristic); [standard deviation]

[a] 0 = “Entirely Liberal”; 100 = “Entirely Conservative”

Table A2: Violations of Prize Independence by Competence Criteria - Experiment 1

Criteria	n	% Violations	95% CI
none	1200	67	[65, 70]
A-C	664	47	[43, 51]
A-D	641	47	[44, 51]
A-E	447	45	[40, 50]
A-F	356	49	[44, 54]
A-G	50	40	[27, 55]

The increasingly restrictive criteria, A through G, are detailed in Table 1

Table A3: Violations of Prize Independence by Competence Criteria - Experiment 1_M

Criteria	n	% Violations	95% CI
none	120	48	[39, 58]
A-C	116	47	[37, 56]
A-D	115	47	[38, 56]
A-E	96	46	[36, 56]
A-F	92	45	[34, 55]
A-G	8	38	[09, 76]

The increasingly restrictive criteria, A through G, are detailed in Table 1

Table A4: Violations of Prize Independence by Competence Criteria - Experiment 2

Criteria	n	% Violations	95% CI
none	121	36	[27, 45]
A-C	109	30	[22, 40]
A-D	108	31	[22, 40]
A-E	82	26	[17, 36]
A-F	79	25	[16, 36]
A-G	12	25	[05, 57]

The increasingly restrictive criteria, A through G, are detailed in Table 1

Table A5: Switch Point Frequencies

	Frequency (%) by prize level			
Switch point	\$1	\$5	\$20	Total
1	27	25	24	25
2	23	25	25	24
3	32	33	36	34
4	16	15	14	15
5	2	2	1	2
6	0	0	1	0

Switch point = # times the half-open box was chosen. For each prize level $n = 780$: subjects who face our Primary Design and whose choices met criteria A-C of Table 1. The column Total pools across prize levels, corresponding to $n = 780 \times 3 = 2340$ observations.

Probit Regressions

To complement the nested criteria approach presented in the main paper, we apply probit regressions to allow *ceteris paribus* interpretations and to include variables not clearly linked to competence. In this analysis, we pool subjects from Experiments 1 and 1_M, and only include those meeting the basic criteria A-C. We present estimates in Table A6.

We find that getting the comprehension questions correct (criterion E) is associated with more violations of Prize Independence (by approximately 10 percentage points on average). However, as we know from Figure 3 and Tables A2 and A3, even throwing away the data from subjects who failed any comprehension question leaves violation rates at about 45%.

We also find that getting the probability competence questions correct (criterion F) is associated with higher violation rates (by approximately 17 percentage points on average). The latter finding may suggest not only that violations do not stem from some basic failure to process probabilistic information, but that they may in fact be generated *more* often by those more able or more experienced in dealing with such information.

We find no effect of having a graduate degree (criterion G), being in Experiment 1_M (in which subjects were required to be Master workers) or having joined MTurk after January 31, 2020 (to pick up any differences in those joining after the COVID-19 pandemic - see the discussion of Section 2.7). The inclusion of demographic measures does not meaningfully alter the estimates.

Table A6: Violations of Prize Independence - Primary Design, by Regression

Average marginal effects	Dependent variable: violation of Prize Independence	
	(1)	(2)
Task comprehension	-0.098**	-0.099**
	(0.031)	(0.031)
Probability competence	0.168***a	0.161***a
	(0.022)	(0.031)
Graduate degree	-0.045	-0.028
	(0.062)	(0.056)
Experiment 1_M (Master workers only)	-0.030	-0.020
	(0.040)	(0.044)
Joined after Jan 2020	-0.015	-0.006
	(0.021)	(0.018)
Other demographics		X
Observations	776	776

$^{**}p < 0.01$, $^{***}p < 0.001$, ^{a}Bonferroni-adjusted $p < 0.001$. The five reported regressors are all indicator variables. Violation of Prize Independence is a binary variable = 1 if the subject's choices violate Prize Independence. Task comprehension = 1 if all correct on first attempt; probability competence = 1 if all correct. Standard errors clustered by date (defined by six data collection periods). Other demographics include gender, age, race, household income, and conservatism. Of the 780 subjects who face our Primary Design and whose choices meet criteria A-C (Table 1), 4 have not specified their gender, leaving 776 for analysis.

Primary Design's redacted transcript - intended for online publication only

[Bold text in square brackets was not seen by subjects.]

[Horizontal lines represent page-breaks.]

[HIT = Human Intelligence Task (which is the term used to refer to a job on MTurk)]

Participation Agreement

You have been invited to take part in a research study run by researchers at the University of Oxford. Please read the following statements carefully.

Our Commitments and Privacy Policy

- We never deceive participants.
- We keep our promises made to participants. For example, if we promise a certain payment, participants will indeed receive it.
- In the event that we are responsible for a mistake that is to the disadvantage of participants, we will inform and compensate the respective participants.
- We design, conduct and report our research in accordance with recognized scientific standards and ethical principles.

We adhere to the terms of our privacy policy as stated below.

- The data in the participants' database will only be used for the purpose of the study.
- There is no link between the personal data in the participants' database and the data collected during a study.
- The generated anonymous data will be used for analysis. The end product will be publicly available.
- Your participation in this study is purely voluntary, and you may withdraw your participation or your data at any time without any penalty to you.
- Please be aware that Amazon user information, connected to MTurk worker IDs, can be visible to the public, depending on the privacy settings of your Amazon.com account. See also https://www.mturk.com/mturk/privacynotice for further information on Amazon.com's privacy policies.

There are no known risks associated with your participation in this research beyond those of everyday life. If there is anything about the study or your participation that is unclear or that you do not understand, if you have questions or wish to report a research-related problem, you may contact the Requester via MTurk. For questions about your rights as a research participant, you may contact the Department of Economics, Manor Road Building, Manor Road, Oxford, OX1 3UQ, United Kingdom, at ethics@economics.ox.ac.uk.

○ I agree

Welcome

Please read the following instructions carefully, as they are followed by some comprehension questions.

If you get those questions wrong you will have to read the instructions again and re-take the questions.

○ I understand

Instructions - page 1 of 5

We are conducting an academic study to examine choices under uncertainty.

You will be offered 21 choices between pairs of boxes.

Each time, choose the box you prefer because one of your choices will determine your **bonus payment**.

Instructions - page 2 of 5

Each box has 20 balls inside it. Balls can either be red or yellow.

When a box is selected, one ball is drawn at random from it. If the ball drawn is **red, you win** a prize in dollars. If the ball drawn is yellow, you win nothing.

So, **to maximize your chance of winning a prize:** in each choice, pick the box that you think has the highest number of red balls inside it.

There are three sets of seven choices. The only thing that is different across sets is the prize. One set has a prize of **$1**, another set has **$5** and another has **$20**, but you may not see them in this order.

But there is a difficulty: boxes can be open or half-open. Every choice you make will be between an open box and one fixed half-open box.

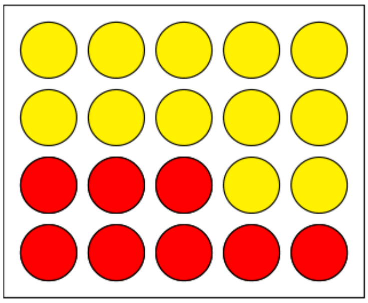
Instructions - page 3 of 5

There are seven **open boxes**. They appear in each set of choices and you can see all the balls inside them.

One is shown below as an example; it has 8 red balls.

Because you can see this, you know your chance of getting a winning (red) ball from this box is **8 in 20**.

Example of an open box

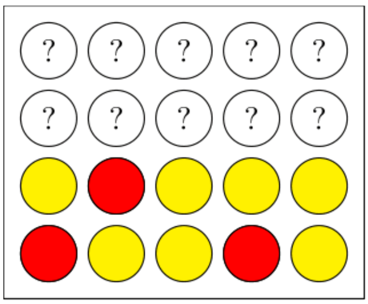


Instructions - page 4 of 5

There is **one half-open box** (shown below), which **NEVER changes across all of the 21 choices you will make**.

Every choice you make will be between this half-open box and an open box.

The half-open box



You can see there are 3 red balls for sure, but you do not know how many of the remaining 10 balls marked "?" are red, so you do not know the total number of red balls in the half-open box.

The true color of each ball marked "?" has been set (randomly) once and for all, and so **is the same throughout the whole HIT**.

Instructions - page 5 of 5

One of the 21 boxes you choose will be selected at random and one ball will be drawn randomly from it.

If you win (draw a red ball), the relevant bonus payment will then be made to your Mturk account. If you do not win, you will not be notified.

Basically, **to maximize your chance of winning a bonus payment:** in each choice, pick the box that you think has the highest number of red balls inside it.

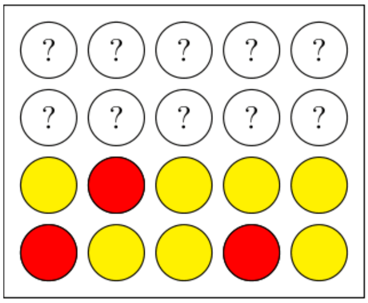
Comprehension Questions

1. Which color of ball corresponds to a win?

- ◯ Red
- ◯ Yellow

2. Consider the half-open box (shown again below).

The half-open box

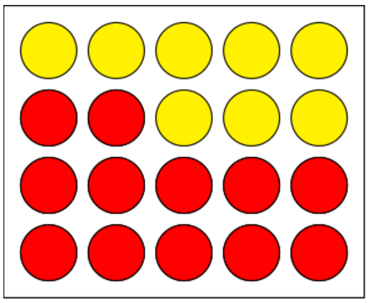


Does the number of red balls in the half-open box change during the HIT?

- ◯ Yes
- ◯ No

3. Suppose the box shown below was selected in order to determine your bonus. What is the chance of having a red ball drawn from it?

An open box



- ◯ 0 in 20
- ◯ 6 in 20
- ◯ 8 in 20
- ◯ 12 in 20

[IF subject got exactly one comprehension question wrong]

Unfortunately, you did not get everything correct.

One page of the instructions is repeated on the next page.

You will then be asked to re-try the comprehension questions.

[Subject is then taken back to the page of instructions containing the information needed to correctly answer the question they got wrong before re-visiting the comprehension questions.]

[IF subject got more than one comprehension question wrong]

Unfortunately, you did not get everything correct.

The instructions will now be presented again.

You will then be asked to re-try the comprehension questions.

[Subject is then required to go through the instructions again before re-visiting the comprehension questions.]

Thank you. The task begins immediately on the next page.

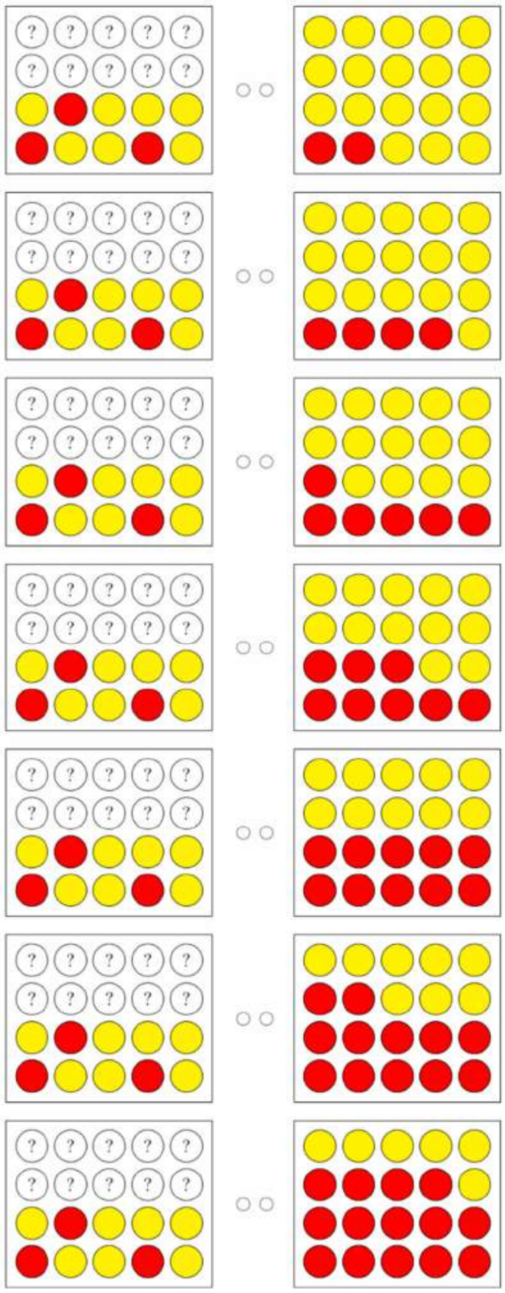
[The three sets were shown all on the same webpage, in a random order. Choice order *within* a set was as shown. While on this webpage, subjects could freely change any of their answers.]

Your Choices

To maximize your chance of winning a bonus: in each choice, choose the box you think has the highest number of red balls.

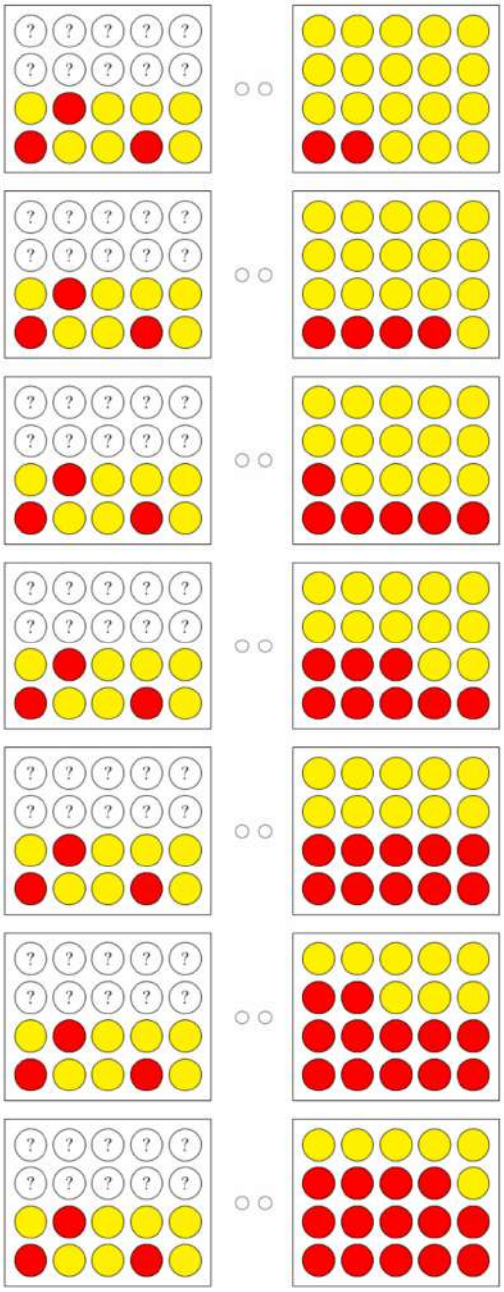
(Remember that the total number of red balls in the half-open box is fixed, and does not change throughout the whole HIT.)

$1 set
If a red ball is drawn from the box you choose: **$1**. (Yellow: $0.)
Choose: the half-open box (left) or an open box (right).



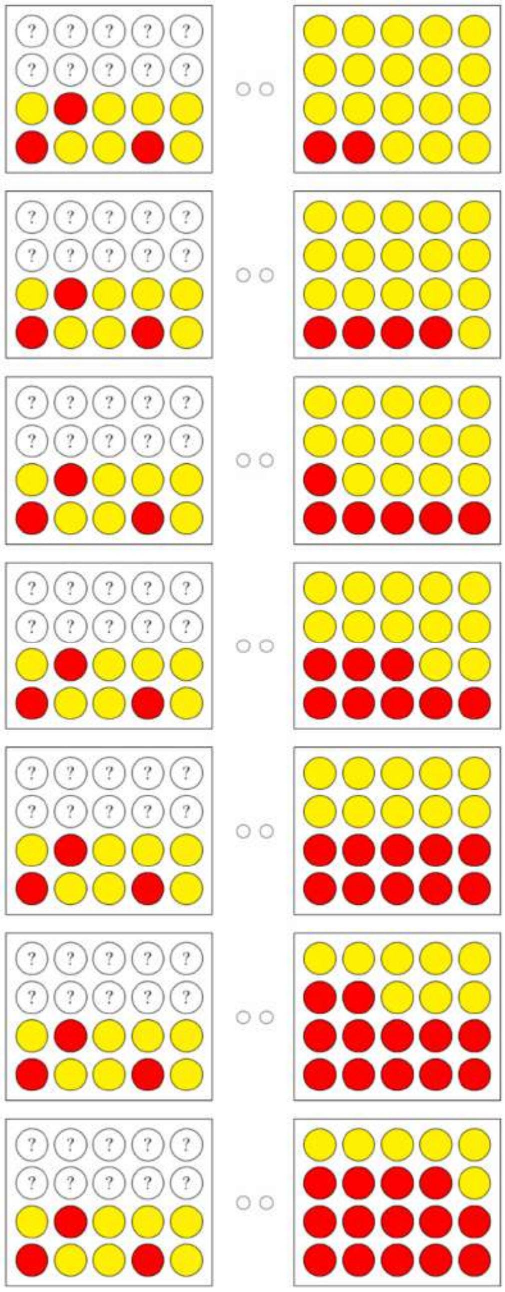
$5 set

If a red ball is drawn from the box you choose: **$5**. (Yellow: $0.)
Choose: the half-open box (left) or an open box (right).



$20 set

If a red ball is drawn from the box you choose: **$20**. (Yellow: $0.)
Choose: the half-open box (left) or an open box (right).



Final Questions - page 1 of 2 **[Probability Competence Questions of Lipkus et al. (2001)]**

These questions test your familiarity and understanding of probabilities and risk.

1. Which of the following numbers represents the biggest risk of getting a disease?

- ◯ 1 in 100
- ◯ 1 in 1000
- ◯ 1 in 10

2. Which of the following numbers represents the biggest risk of getting a disease?

- ◯ 1%
- ◯ 10%
- ◯ 5%

3. If person A's risk of getting a disease is 1% in ten years, and person B's risk is double that of A's, what is B's risk?

- ◯ 0.5%
- ◯ 1%
- ◯ 2%
- ◯ 10%

4. If person A's chance of getting a disease is 1 in 100 in ten years, and person B's risk is double that of A's, what is B's risk?

- ◯ 1 in 100
- ◯ 2 in 100
- ◯ 5 in 100
- ◯ 10 in 100

5. If the chance of getting a disease is 10%, how many people would be expected to get the disease out of 100?

- ◯ 1
- ◯ 2
- ◯ 5
- ◯ 10

Final Questions - page 2 of 2

What is your sex?

- ◯ Male
- ◯ Female
- ◯ Other
- ◯ Prefer not to say

What is your age?

What is your race?

- ◯ White
- ◯ Black or African American
- ◯ Hispanic or Latino
- ◯ American Indian or Alaska Native
- ◯ Asian American
- ◯ Native Hawaiian or Pacific Islander
- ◯ Other
- ◯ Prefer not to say

Which language(s) do you consider to be your mother tongue?

- ☐ English
- ☐ Spanish
- ☐ Chinese (any variety inc. Cantonese and Mandarin)
- ☐ Other

What is your household's annual income? (US dollars, before tax)

- ◯ 0-9,999
- ◯ 10,000 - 19,999
- ◯ 20,000 - 29,999
- ◯ 30,000 - 39,999

- 40,000 - 49,999
- 50,000 - 59,999
- 60,000 - 69,999
- 70,000 - 79,99
- 80,000 - 89,999
- 90,000 - 99,999
- 100,000 - 124,999
- 125,000 - 149,999
- 150,000 +

What is the highest grade of school you have completed, or the highest degree you have received?

- No schooling (or less than 1 year)
- Nursery, kindergarten, and elementary (grades 1-8)
- High school (grades 9-12, no degree)
- High school graduate (or equivalent)
- Some college (1-4 years, no degree)
- Bachelor's degree (BA, BS, AB, etc)
- Master's degree (MA, MS, MENG, MSW, etc)
- Professional school degree (MD, DDC, JD, etc)
- Doctorate degree (PhD, EdD, etc)

Generally speaking, which point on this scale best describes your political views?
[A slider was presented with range [0,100] with "Liberal" over 0 and "Conservative" over 100.]

I joined Mturk as a worker:

- before February 1, 2020
- after February 1, 2020

What is your Mturk ID? (Please copy and paste it to avoid typos)

Secondary Design's redacted transcript - intended for online publication only

[Bold text in square brackets was not seen by subjects.]

[Only the main choice pages are shown below. The rest of the secondary design was identical to the primary design (save a few trivial wording changes were necessary to reflect the different presentation of the choices), so we refer the reader to the prior transcript for the rest.]

[Each of the three choices for each pair of boxes was revealed one at a time, below the preceding one. After all three choices for a given pair were made, the next pair and the first ($1) question for it would appear, below the last question answered, on the same webpage.]

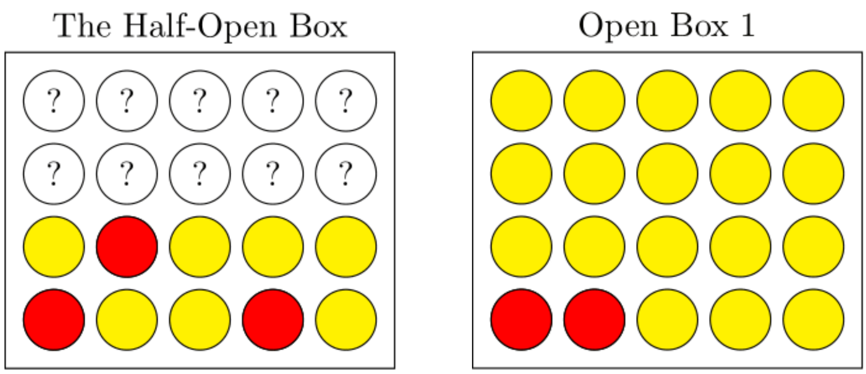
[The order of the choices was as shown below.]

[While on this webpage, subjects could freely change any of their answers.]

Your Choices

To maximize your chance of winning a bonus: in each choice, choose the box you think has the highest number of red balls.

(Remember that the total number of red balls in the half-open box is fixed, and does not change throughout the whole HIT.)



For a prize of **$1**, which box do you choose?

◯ The Half-Open Box ◯ Open Box 1

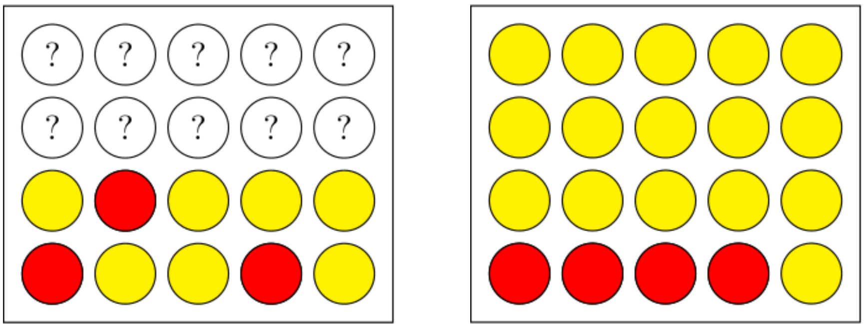
For a prize of **$5**, which box do you choose?

◯ The Half-Open Box ◯ Open Box 1

For a prize of **$20**, which box do you choose?

◯ The Half-Open Box ◯ Open Box 1

The Half-Open Box　　　　Open Box 2



For a prize of **$1**, which box do you choose?

◯ The Half-Open Box　　◯ Open Box 2

For a prize of **$5**, which box do you choose?

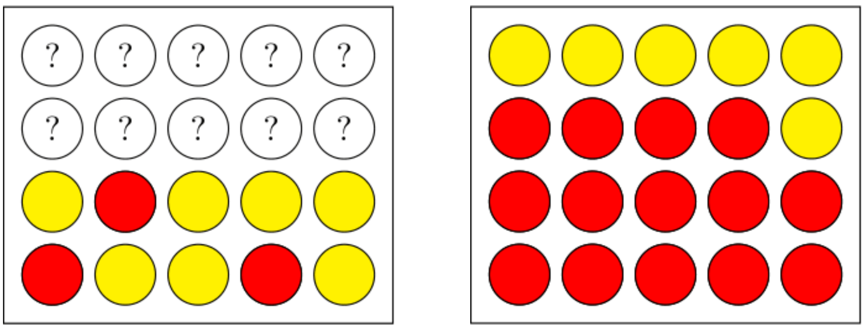
◯ The Half-Open Box　　◯ Open Box 2

For a prize of **$20**, which box do you choose?

◯ The Half-Open Box　　◯ Open Box 2

[…Choices continued with the only difference being the Open Boxes' number of red balls, which increased in increments of 2, until the final choice pair, shown below.]

The Half-Open Box　　　　Open Box 7



For a prize of **$1**, which box do you choose?

◯ The Half-Open Box　　◯ Open Box 7

For a prize of **$5**, which box do you choose?

◯ The Half-Open Box　　◯ Open Box 7

For a prize of **$20**, which box do you choose?

◯ The Half-Open Box　　◯ Open Box 7

School of Economics and Finance

This working paper has been produced by the School of Economics and Finance at Queen Mary University of London

School of Economics and Finance Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/