Multiple Switching and Data Quality in the Multiple Price List

Y. Jane Zhang
Sharon Xuejing Zuo
Chi Wai Yu*

May, 2017

Abstract

Reliable and meaningful measurement of individual risk preferences is critical for understanding a wide range of economic decision-making. Researchers working with the widely used Multiple Price List (MPL) instrument have found that a substantial proportion of subjects switch back and forth between the safe and the risky choice columns in the instrument, which is behavior believed to indicate low quality decision-making (e.g., Charness et al. (2013)). In this study we develop a "nudge" protocol for the MPL that reduced multiple switching behavior (MSB) from 31% to 10% (p-value $<$ 0.001) without limiting the choice set. We further develop a conceptual framework to formally test three leading explanations for the nature of low quality decision-making in the MPL using the covariance of responses in the MPL with a second, simple risk instrument. Using a counter-balanced within and between-group experimental design, we find that low quality decision-making in the MPL is best explained by task-specific miscomprehension. Using our framework we propose a novel metric for data quality, which we use to make three observations: 1. the nudge treatment increased high quality responses by 143%. 2. MSB does not capture the full extent of low quality decision-making, as even non-multiple switchers generate relatively low data quality under the standard protocol. 3. While cognitive ability explains MSB, it does not predict higher data quality. To the extent that the nudge treatment elicited more cognitive effort, the findings suggest that cognitive effort is the limiting factor in achieving high data quality in the MPL.

*Zhang: Hong Kong University of Science and Technology. Email: janezhang@ust.hk. Zuo: University of Houston. Email: xzuo2@uh.edu. Yu: Hong Kong University of Science and Technology. Email: macwyu@ust.hk. We thank Catherine Eckel, Willa Friedman, Elaine Liu, Muriel Niederle, Angelino Viceisza, and Rick Wilson for helpful conversation and suggestions. We are indebted to Yang Haitao and Geirongduzhi for their field assistance. Keywords: risk preferences, multiple switching, data quality. JEL Codes: C91, D81, C81.

1 Introduction

Reliable and meaningful measurement of individual risk preferences is critical for understanding a wide range of economic decision-making. Experimental economics has contributed many tools to measure individual risk preferences using incentivized choice situations (see Harrison and Rutström (2008) for a survey), although different risk elicitation methods are not always correlated with one another (see, for example, Charness and Viceisza (2016) and the references in Niederle (2016)). The Multiple Price List (MPL) instrument, often called the Holt and Laury (2002) instrument, is one of the most widely used methods to elicit risk individual preferences. The MPL is also used in settings other than the measurement of risk preferences, such as pricing commodities (Kahneman et al., 1990; Cassar et al., 2016) and measuring discount rates (Harrison et al., 2002; Andersen et al., 2008). An attractive feature of the MPL is that it can be used to elicit arbitrarily precise intervals of risk aversion estimates (Charness et al., 2013; Tanaka et al., 2010; Jacobson and Petrie, 2009).

Despite its popularity, an empirical difficulty that researchers encounter when using the MPL is that a substantial proportion of subjects switch back and forth between the safe and the risky choice columns in the instrument (i.e., engage in multiple switching), which is behavior incompatible with standard assumptions on preferences (Charness et al., 2013). Such multiple switching behavior (MSB) is generally considered low quality decision-making, and the observed responses are treated as noise, although some studies argue that MSB may indicate indifference between a range of options (Andersen et al., 2006). MSB is especially pronounced in developing countries. Whereas typical studies in developed countries find multiple switching to affect approximately 10% of the subjects (e.g., Holt and Laury (2002) reports 13% multiple switchers; Dave et al. (2010) report 8.5% multiple switchers), in developing countries the multiple switching rate can be over 50% (Jacobson and Petrie, 2009; Charness and Viceisza, 2016).

A common experimental practice used to reduce MSB is to ask subjects to indicate the row in which they would like to switch from the risky option to the safe option (e.g., Andersen et al. (2006); Tanaka et al. (2010)). This eliminates MSB, but also reduces the choice set, so we do not know whether the subject would have engaged in multiple switching if they had been free to do so. In this study we develop a “nudge” protocol to increase cognitive effort without limiting the choice set.[1] After subjects complete the MPL task, we ask them if they are sure of their responses and give them the option of hearing the instructions one more time. We found a reduction in MSB from 31% using the standard protocol to 10% in the nudge protocol (p-value of difference < 0.001). This suggests that at least 2/3 of MSB can be categorized as mistakes which are corrected upon further reflection, which sets it apart from the deliberate randomization behavior described in Agranov and Ortoleva (2017), for example.

Although the literature generally views MSB as equivalent to low decision-making quality, an individual can make low quality decisions that do not result in MSB. The potential for non-multiple switchers to make low quality decisions is not well understood and may be an important source of noise in the data. We develop a conceptual framework, which formally defines decision-making quality independently of MSB, and, using the covariance between responses on the MPL task and responses on a simpler lottery selection task, allows us to test between three explanations for low-quality decision-making in the MPL suggested by the findings in the literature.

Of the three explanations, perhaps the most pessimistic is that bad decision-making in the MPL task is a stable attribute of the decision-maker. This explanation is supported by the evidence in Jacobson and Petrie (2009), which shows that multiple switchers in MPL instruments make sub-optimal decisions in other areas of their lives. Similarly, Choi et al. (2014) finds that people make bad decisions in many areas of their lives and individuals who make low-quality decisions in an experimental setting have less wealth, controlling for their current income and a slew of demographic and socioeconomic status variables.[2] Under

[1] A nudge, as defined by Thaler and Sunstein (2008) is “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives.”

[2] Low quality decisions in Choi et al. (2014) are defined as choices inconsistent with the generalized axiom

this view, decision-making quality is not readily improvable, and should not respond to an unintrusive stimulus such as the nudge treatment.

At first pass, this explanation seems immediately incompatible with our finding that MSB can be reduced by the nudge protocol. However, since we do not make the assumption that MSB captures the full extent of low quality decision-making, the finding that MSB can be reduced does not necessarily imply that decision-making quality can be improved. For example, individuals could have inferred that experimenters wanted less MSB and in their desire to be helpful make fewer multiple switches but continue to give noisy responses that do not reflect their true risk preferences. Charness et al. (2013) raises a similar concern in the context of treatments that eliminate MSB but do not induce higher quality decision-making, which would mask data quality issues.

A second explanation is that low quality decision-making in the MPL is incidental to the complexity of the MPL instrument. For example, Charness and Viceisza (2016) argue that a lack of comprehension is a serious concern with using the MPL in developing countries. Charness et al. (2013) uses the term "failure to understand". Andreoni and Sprenger (2011) directly tells subjects that "Most people begin by preferring Option A and then switch to Option B, so one way to view this task is to determine the best row to switch from Option A to Option B," in an effort to improve comprehension. Furthermore, Dave et al. (2010) demonstrates that the MPL produced noisier estimates of risk aversion than did a simple lottery selection instrument developed in Eckel and Grossman (2002), especially for low math ability individuals, which suggests that cognitive ability plays a role in comprehension of the MPL. Under this interpretation, short-term treatments such as the nudge protocol can improve decision-making quality in the MPL by improving comprehension.

A third explanation also views low quality decisions in the MPL as improvable in the short-term, but assumes that individuals who make low-quality decisions in the MPL also make low-quality decisions using other instruments because they are careless. This expla-

of revealed preferences in their experiment which asks subjects to choose bundles of goods under varying budget slopes.

nation was mentioned in Brick et al. (2012) but has not been given much attention in the literature. The main difference between carelessness and miscomprehension is that low quality decision-making due to carelessness is not unique to the MPL instrument. The implication is that the MPL, by virtue of allowing for MSB, which may be an indicator of low-quality decision making, may in fact be preferred to simpler instruments that do not allow for MSB and which obscures low quality decision-making.

To test the different explanations, we compare responses given in the MPL to responses given in a simple lottery selection task (henceforth, LS) in which MSB is not possible, in the control and nudge treatment groups.[3]

In the conceptual framework that we develop below, we show that the covariance of the responses given in the MPL and the LS tasks will be the same for the control and nudge groups under the stable attribute explanation, it will be larger for the nudged group than the control group under the miscomprehension explanation, and will be weakly larger for the control group than the nudge group under the carelessness explanation. To make inferences about the difference between two population covariances, we derived the variance of the difference between two sample covariances and propose an estimator which is consistent and asymptotically normal (see Appendix B.2).

Our test finds consistent evidence in support of the miscomprehension explanation. Furthermore, this finding in conjunction with our conceptual framework motivates a novel metric to quantify data quality using the correlation of the MPL task and the simpler LS task, which is independent of the multiple switch rate. Using this metric we find that the nudge treatment conservatively improves data quality by 143%. Data from both multiple switchers and non-multiple switchers under the standard protocol are characterized by low decision-making quality, implying that MSB does not capture the full extent of low quality

[3]The lottery selection task was independently developed in Binswanger (1980) and Eckel and Grossman (2002). To give a brief explanation of the LS task, it requires subjects to select one out of six different coin-flip lotteries. Lotteries with higher expected value also have higher variance. Selecting a lottery with higher expected value (and higher variance) is indicative of higher risk tolerance. In this task, all choices are compatible with standard assumptions on preferences.

decision-making, and that discarding multiple switchers does not ensure high data quality, as is often believed. Although cognitive ability predicts MSB, which is consistent with the previous literature, cognitive ability is not a significant determinant of data quality. To the extent that the nudge treatment was able to elicit more cognitive effort from respondents, the findings suggest that cognitive ability is not a limiting factor in achieving high data quality in the MPL, but cognitive effort is. Increased cognitive effort has also been proposed as an explanation for the reduction in MPL errors in choice situations where subjects can incur a loss (Von Gaudecker et al., 2011) and when stakes are high (Holt and Laury, 2002). This relates to the literature on bounded rationality, which finds cognitive effort plays an important role in decision-making independent of intelligence, and that humans are prone to cognitive laziness (Kahneman, 2003, 2011).

This paper is organized as follows. In Section 2 we develop our conceptual framework. In Section 3 we describe our experimental design and our subjects. Section 4 presents the empirical analysis. Section 5 presents additional results and Section 6 concludes.

2 Conceptual Framework

In this section, we develop a simple conceptual framework (see Figure 1) to motivate our experimental design and analysis, presented in the next section. Let the signal in risk tolerance, or "true" risk preferences, be denoted S, with $\mu_s = E(S)$, and $\sigma_s^2 = Var(S) > 0$. η_{ij} and ν_{ij} are i.i.d. noise terms in measured risk aversion for the MPL and the LS tasks, respectively. $i \in \{1, 2\}$ denotes type, $j \in \{s, n\}$ denotes treatment status, where s represents standard MPL and n represents nudged MPL. Suppose there are two types of individuals - those who are confused by the MPL and those who are not. Type 1, occurring with probability p, are not confused by the MPL and give"high quality" responses which reflect both signal and noise. Without additional intervention, Type 2 individuals, occurring with probability $1 - p$, are confused by the MPL and give "low quality" responses that reflect only noise. Note that $1-p$ can be larger than the proportion of multiple switchers because Type 2 individuals could by chance avoid multiple switching. The maintained assumption is that $0 < p < 1$. The presence of the two types leads to a mixture distribution in measured risk aversion.

Scenario I: Low quality decision-making is a stable attribute of the decision-maker

For the control group using the MPL, the response from Type 1 is $X_{1s} = S + \eta_{1s}$, and the response from Type 2 is $X_{2s} = \eta_{2s}$. For the LS task, the response from Type 1 is $Y_{1s} = S + \nu_{1s}$. Because under this scenario confused individuals consistently make low quality decisions, we expect the response on the LS task for Type 2 to be $Y_{2s} = \nu_{2s}$.

Since the nudge treatment cannot induce confused individuals to make "high quality" decisions, the responses of the treatment group will be equal in distribution to those in the control group. That is to say, $X_{1n} \stackrel{d}{=} X_{1s}$; $X_{2n} \stackrel{d}{=} X_{2s}$; $Y_{1n} \stackrel{d}{=} Y_{1s}$, and $Y_{2n} \stackrel{d}{=} Y_{2s}$.

The response in the MPL is denoted by MPL_j, $j \in \{s, n\}$, whose density function is the mixture of the density functions of X_{1j} and X_{2j}, where p is the weight placed on the density function of X_{1j}. Similarly, the response in the LS task is denoted by LS_j, whose density function is the mixture of the density functions of Y_{1j} and Y_{2j}, where p is the weight placed on the density function of Y_{1j}.

Because of the equality in distribution of the component distribution functions, $MPL_s \stackrel{d}{=} MPL_n$, and $LS_s \stackrel{d}{=} LS_n$. This implies that $Cov(MPL_s, LS_s) = Cov(MPL_n, LS_n)$,. Similarly, $Var(MPL_s) = Var(MPL_n)$ and $Var(LS_s) = Var(LS_n)$, implying $Corr(MPL_s, LS_s) = Corr(MPL_n, LS_n)$.

Scenario II: Task specific miscomprehension

Under this scenario, the confusion is specific to the MPL task. For the control group using the MPL, the response is identical to Scenario I: The response from Type 1 is $X_{1s} = S + \eta_{1s}$, and the response from Type 2 is $X_{2s} = \eta_{2s}$.

For the control group using the LS task, because the confusion is specific to the MPL, the responses from Type 1 and Type 2 are $Y_{1s} = S + \nu_{1s}$ and $Y_{2s} = S + \nu_{2s}$, respectively, where $\nu_{1s} \overset{d}{=} \nu_{2s}$.

For the treatment group using the nudged MPL, the response from Type 1 is $X_{1n} = S + \eta_{1n}$, where $\eta_{1n} \overset{d}{=} \eta_{1s}$. If the treatment fully "unconfuses" Type 2s, then the response from Type 2 is $X_{2n} = S + \eta_{2n}$, where $\eta_{2n} \overset{d}{=} \eta_{1n}$.

Because the nudge treatment works only on the MPL, we expect no differences in the responses on the LS task by treatment status, for both types of individuals. That is to say, $Y_{1n} = S + \nu_{1n}$ and $Y_{2n} = S + \nu_{2n}$, where $\nu_{1n} \overset{d}{=} \nu_{1s}$ and $\nu_{2n} \overset{d}{=} \nu_{2s}$. Therefore, $LS_s \overset{d}{=} LS_n$.

By the properties of mixture distributions,

$$Cov(MPL_j, LS_j) = \sum_i [p_i Cov(X_{ij}, Y_{ij}) + p_i(\mu_{Xij} - \mu_{Xj})(\mu_{Yij} - \mu_{Yj})], \qquad (1)$$

where $\mu_{Xij} = E(X_{ij})$, $\mu_{Yij} = E(Y_{ij})$, $\mu_{Xj} = E(X_j)$, and $\mu_{Yj} = E(Y_j)$.

For the control group, $Cov(MPL_s, LS_s) = pCov(X_{1s}, Y_{1s}) + (1-p)Cov(X_{2s}, Y_{2s}) = p\sigma_s^2$, and for treatment group, $Cov(MPL_n, LS_n) = \sigma_s^2$. This yields the result that $Cov(MPL_s, LS_s) < Cov(MPL_n, LS_n)$.

More generally, we can assume that the nudge treatment "unconfuses" a proper subset of Type 2s. Appendix B.1 shows that in that case $Cov(MPL_s, LS_s) = p_1\sigma_s^2$ and $Cov(MPL_n, LS_n) = (p_1 + p_2)\sigma_s^2$ where p_1 is the proportion who are not confused in the control group and p_2 is the additional proportion who have become unconfused by the nudge treatment in the treatment group. We will have the same result that $Cov(MPL_s, LS_s) < Cov(MPL_n, LS_n)$ in this general case, if $p_2 > 0$. Scenario II also implies that for both the control and treatment groups, the lower the proportion of confused individuals, the higher will be the covariance between the responses on the MPL and the LS tasks.

Because this scenario does not produce clear predictions on the relative sizes of $Var(MPL_s)$ and $Var(MPL_n)$, it does not produce clear predictions on the relative sizes of $Corr(MPL_n, LS_n)$ and $Corr(MPL_s, LS_s)$.

Scenario III: Carelessness

This scenario assumes that the confusion of Type 2 individuals is due to non-task specific carelessness, affecting both the MPL and the LS tasks for the control group. The nudge treatment, which is only applied to the MPL, removes confusion only in the MPL task.

Identical to Scenario I and II, for the control group using the MPL, the response from Type 1 is $X_{1s} = S + \eta_{1s}$, and the response from Type 2 is $X_{2s} = \eta_{2s}$.

For the control group using the LS task, the response from Type 1 is $Y_{1s} = S + \nu_{1s}$, and the response from Type 2 is $Y_{2s} = \nu_{2s}$. Because carelessness is not task specific, the responses on the LS task for type 2 individuals also only capture noise.

For the treatment group using the nudged MPL, the response from Type 1 is $X_{1n} = S + \eta_{1n}$, where $\eta_{1n} \stackrel{d}{=} \eta_{1s}$. If the treatment fully "unconfuses" Type 2s, then the response from Type 2 is $X_{2n} = S + \eta_{2n}$, where $\eta_{1n} \stackrel{d}{=} \eta_{2n}$.

Because the nudge treatment works only on the MPL, we expect no differences in the responses on the LS task by treatment status. For the treatment group using the LS task, the response from Type 1 is $Y_{1n} = S + \nu_{1n}$ and the response from Type 2 is $Y_{2n} = \nu_{2n}$, where $\nu_{1s} \stackrel{d}{=} \nu_{1n}$ and $\nu_{2s} \stackrel{d}{=} \nu_{2n}$, so that $LS_s \stackrel{d}{=} LS_n$.

Using the fact that $\mu = p\mu_1 + (1-p)\mu_2$, Equation 1 simplifies to

$$\begin{aligned} Cov(MPL_j, LS_j) = pCov(X_{1j}, Y_{1j}) + (1-p)Cov(X_{2j}, Y_{2j}) \\ + p(1-p)(\mu_{X1j} - \mu_{X2j})(\mu_{Y1j} - \mu_{Y2j}). \end{aligned} \tag{2}$$

For the control group, $Cov(MPL_s, LS_s) = p\sigma_s^2 + p(1-p)(\mu_{X1s} - \mu_{X2s})(\mu_{Y1s} - \mu_{Y2s})$. For the treatment group, $Cov(MPL_n, LS_n) = p\sigma_s^2$. As long as the difference in the expected value of measured risk tolerance in the control group for Type 1 and Type 2 is not in the opposite direction for the two tasks, then $Cov(MPL_s, LS_s) \geq Cov(MPL_n, LS_n)$. We would violate this assumption if, for example, confused individuals without intervention are more risk averse in the MPL task, but less risk averse in the LS task. We do not know of any theory or evidence that predicts this pattern.

The intuition for this result is that unlike in scenario II, there are no gains in the covariances between the two tasks for Type 2 individuals under the nudge treatment because the carelessness of Type 2 individuals is not improved for the LS task. On the other the hand, the relative magnitudes of the expected values of risk tolerance for Type 1 and Type 2 individuals are allowed to have the same pattern for MPL and LS tasks in the control group, which adds to the overall covariance of the control group, but they do not have this "similarity" in the treatment group.

It can be demonstrated that under the more general assumption where the nudge treatment "unconfuses" a proper subset of Type 2s, we have the same result that $Cov(MPL_s, LS_s) \geq Cov(MPL_n, LS_n)$. See Appendix B.1 for the proof.

3 Experimental Design

3.1 Experimental Setting

Subjects were recruited from a rural middle school (7th to 9th grade) in an ethnically diverse region of southwest China. The county in which this middle school is located has been on the register of nationally recognized "poor" counties since the criteria for the designation were established in 1986. According to the provincial statistical yearbook, in 2014, the county annual average GDP per capita was 11,345 RMB (1650 USD).

From the complete school rosters, we randomly drew students from all regular classes in the middle school.[4] Class size ranges from 51 to 72. Our final sample consists of 193 out of 212 students selected by us, for a response rate of 91%. Non-response is largely due to student absenteeism and the roster not being updated for students who had dropped out of school.

The experiments were conducted in the spring semester of 2015, mainly during the 4pm

[4]We omitted the honors class students at the request of school administrators.

to 7pm break time on campus. [5] Students completed the experiments one-on-one with our experimenters. At the beginning of the experiments, students were told that they would play two games and only one of them would be chosen randomly to realize their final payment. Students were also asked to fill out a short survey after all experiments are completed to capture basic demographic and socioeconomic status information. Subject payments were handed out after the surveys were completed. Average payout was 6.19 RMB, not including a pencil and eraser as a show up gift. Student test scores were separately obtained from the school administrators.

3.2 Balance tests

Subjects were randomly assigned to either the control or the treatment group.[6] Table 1 reports the balance of demographic and socioeconomic status variables between the treatment and the control groups. We found no statistically significant differences between the control and treatment groups in age, size of household, distance to school, mother's educational attainment, mother's occupation, monthly household income, monthly allowance, or students' test scores.

3.3 Experimental Design

Both the treatment group and control group are administered two tasks: the MPL task and the LS task. The control group is administered MPL using the standard protocol while the treatment group is administered the MPL with a nudge protocol, explained below. There is no difference in the administration of the LS task for the control and treatment groups. The order of the two tasks was randomized within each group. In Appendix Table A.1 we report balance by treatment status and the order in which the MPL and LS tasks were

[5] All students are in school through the evening self-study period (wanzixi) which ends at 10pm. This middle school is a boarding school. Over 85% of students live on campus and are only allowed to return home on weekends. The rest live within a 10 min walk to school.

[6] Because treatment status was assigned prior to the date of the experiment, the final number of subjects in each group was not identical.

administered. The results show that there are no statistically significant differences between the four groups defined by treatment status and task order.

3.4 MPL instrument and the Nudge Treatment

The MPL task follows the design in Dohmen et al. (2011).[7] Subjects are required to make six choices between a lottery and a certain payout (see Appendix C for the instrument). Option A is a coin flip lottery (risky choice) with 50% chance of paying 10 RMB and 50% chance of paying 0 RMB. Option A does not change across the six choices. Option B is the certain cash payout (safe choice), in increasing increments of 1 RMB, from 1 RMB to 6 RMB. One of the 6 pairs of options is randomly selected from the instrument for each subject after she makes her choices, and the option she chose from the selected pair will be implemented. The instrument is incentive compatible. For example, a participant who values the coin flip lottery (option A) at 3.5 RMB certain payout should choose the lottery for all values of the certain payout below 3.5 RMB and should choose the certain payout when it is above 3.5 RMB. For this subject we should observe three choices of the risky choice before switching to make three safe subjects. Under standard assumptions on preferences, subjects should make at most one switch from the risky choice to the safe choice, nevertheless, in practice we find that many subjects switch back to the risky choice after having made a safe choice.

In the nudge treatment, the subjects are first given the exact same instructions used in the control group to administer the MPL task. As each subject hands in his or her responses, we say the following: "Have you decided? You can think about your choices again carefully and can change your choices. If you would like, we can explain this game one more time." See Appendix C for the protocol. Those who indicate the need are given the instructions again. The treatment is designed to encourage the subjects to put more cognitive effort into the task, without taking away the ability of subjects to engage in MSB. Indeed, 10% of the

[7] While Dohmen et al. (2011) asks subjects after making the first switch from the risky choice to the safe choice whether they would also like higher amounts of the safe choice, we do not overtly discourage subjects from multiple switching in either the control or treatment groups.

treatment group exhibited MSB.

The choice results using the MPL instrument are shown in Table 2. We report the distribution of subjects choosing each possible number of lottery (risky) options (from 0 to 6) before making their first switch to the safe choice. We also report the range of the implied CRRA coefficient corresponding to each number of lottery options chosen, assuming no multiple switching. In the full sample, about 20% of subjects were multiple switchers, which falls in the range of many previous findings (Charness et al., 2013).

3.5 LS Instrument

The format of the LS instrument follows Barr and Genicot (2008). The appeal of this design is its simplicity. Subjects are only allowed to choose one coin flip lottery out of six, with the first lottery offering a certain amount and all other alternatives offering higher expected payoff along with higher variance (see Appendix C for the instrument). A more risk tolerant individual is more likely to choose lotteries with higher expected payoffs and higher variance. All choices are consistent with standard assumptions on preferences.

Table 3 reports the simple LS game results for the control and treatment group separately. We report the low and high payoffs for each lottery, the implied range of the CRRA coefficient corresponding to each choice and the percentage of subjects choosing each lottery in each group. The distribution of choices are similar in the two groups. In each group, the lottery chosen with the highest frequency is the third safest lottery and a Mann-Whitney test finds no significant distributional differences between the treatment and control groups in the lottery chosen (p-value = 0.33).

4 Empirical Analysis

The last row of Table 1 reports the share of subjects who are multiple switchers in the control and treatment groups. The nudge treatment reduces the share of multiple switchers from

31% in the control group to 10% in the treatment group. The p-value of the difference in the multiple switching rate is less than 0.001. The fact that the nudge treatment, which does not limit choice sets or overtly discourage MSB is able to eliminate 66% of MSB suggests that the majority of MSB are mistakes that are corrected upon further reflection rather than the result of deliberate choice.

As demonstrated in Section 2, the relative size of the covariances between the MPL and the LS task in the control and treatment groups will allow us to pin down the explanation for MSB most consistent with the data. Table 4 reports the covariance between choices made in the MPL task (number of risky choices) and the LS task (riskiness of lottery chosen), which correspond to the risk tolerance ranking of the implied CRRA coefficients of the choices made in each instrument. Because of the presence of multiple switching, we used three different methods to code the number of risky choices in the MPL task. Method A uses the total number of risky choices (this method was suggested by Holt and Laury (2002). Method B uses the number of risky choices made before the first point at which individuals switch from the risky choice to the safe choice, or the "first switch point" (this method was used in, for example, Harrison and Rutström (2008) and Meier and Sprenger (2013)). Method C uses the average of the decision number preceding the first switch from a risky choice to a safe choice and the decision number preceding the last switch from a risky choice to a safe choice, or the last decision number if the last decision was a risky choice (this method is inspired by the argument in the literature that multiple switching is due to indifference (Andersen et al., 2006)). The last column in Table 4 reports the p-value of the difference between the covariance in the control and treatment groups. Because we were unable to find an estimator in the literature for the difference between two population covariances, to do inference, we derived the variance of the difference between two sample covariances in Appendix B.2, and propose an estimator which is consistent and asymptotically normal. This method is also used to find the significance levels of the estimated covariances.

To simulate the magnitude of the covariance between the two tasks if subjects did not understand the MPL at all and made their choices randomly, we randomly generated 10,000 bootstrap samples of MPL choices from the empirical distribution of the number of risky choices made in the MPL, separately for the control and treatment groups.[8] We find the covariance of MPL responses in each bootstrap sample with the actual responses in the LS task and report the mean of sample covariance and the standard error of sample covariance in Table 4. The last column reports the average p-value of the difference between the covariance in the control and in the treatment groups over the 10,000 observations.

In the treatment group, methods A, B and C produce covariances between the MPL and LS tasks of 1.51, 1.67, and 1.58, respectively, all significantly different from 0 at the 1% level. However, in the control group, the covariance between the responses on these two tasks are small and only method B results in a marginally significant covariance of 0.71. The p-value of the difference in the covariances are 0.01, 0.08, and 0.03, respectively. The randomly generated MPL choices produce a mean covariance close to zero for both the control and treatment groups, which are not statistically different from each other. For the control group, the 95% confidence interval constructed from the bootstrap samples, using normal approximation, is $[-0.753, 0.747]$, which contains 0.300, 0.713, and 0.489, indicating that the covariance of control group responses (using any of the three coding methods) is not statistically significantly different from the mean covariance obtained from randomly generated MPL choices.[9]

Table 5 shows the Pearson correlation coefficients between responses in the MPL and LS tasks. Although the conceptual framework does not produce clear predictions on the relative sizes of the variances of the MPL and LS responses in the treatment and control groups, empirically we find that the variances of the MPL responses and the variances of the LS responses are not statistically different from each other in the treatment versus the control groups (see Table 6). Therefore, the correlation results should provide a similar pattern as the covariance results. In the treatment group, methods A, B and C produce

[8]The number of risky choices is coded using method B.

[9]See Appendix Figure A.1 for the empirical distribution of the bootstrap sample covariances.

correlation coefficients between the MPL and LS tasks of 0.424, 0.449, and 0.440, respectively, all significantly different from 0 at the 1% level. However, in the control group, the correlation coefficients between these two tasks are small and only method B results in a marginally significant correlation coefficient of 0.185. The p-value of the difference in the correlation coefficients are statistically significant at the 5% level for all three coding methods. Results from the 10,000 randomly generated bootstrap samples also show a similar pattern as the covariance results.

This set of results consistently support Scenario II - task specific miscomprehension of the MPL, which is the only explanation for low decision-making quality from the literature that predicts greater covariance in the treatment group. These results also rule out the interpretation that the lower rate of MSB in the treatment group is due to experimenter demand effects. If that were the case, choices in the nudged MPL would be just as noisy as the choices in the standard MPL, and there should be no difference in the covariance or correlation coefficient between the control and treatment groups.

Results using method C also speaks to the potential for indifference to account for MSB. If we assume MSB is a result of indifference between the lottery and a range of certain payout values defined by the first switch point and last switch point from the risky to the safe choice, then of the three methods, method C, which uses the midpoint of these certain payouts to value the lottery, should give the best approximation to true risk aversion. However, as Table 4 and 5 show, the covariance and the correlation coefficient between the two tasks are no larger using method C than the other two methods, for either the control or treatment group. Indifference in the true sense also should not respond to the nudge treatment, whereas we find a substantial reduction in MSB.

5 Additional Results

5.1 Data Quality Metric

Under the interpretation of Scenario II, a higher covariance between the two tasks is indicative of higher decision-making quality (that is to say, a higher proportion of individuals making high-quality decisions). The percent of maximum decision-making quality achieved is the ratio of the actual covariance to the maximum possible covariance: σ_s^2. To approximate σ_s^2 we have two candidates, $Var(MPL_j)$ and $Var(LS_j)$, which are equal to $\sigma_s^2 + \sigma_{\eta j}^2$ and $\sigma_s^2 + \sigma_{\nu j}^2$, respectively. The geometric mean of $Var(MPL_j)$ and $Var(LS_j)$ gives an upperbound estimate on σ_s^2, or maximum decision-making quality. Therefore, the correlation coefficient between the two tasks, $Cov(MPL_j, LS_j)/\sigma_{MPLj}\sigma_{LSj}$, gives a lowerbound estimate of percent of maximum decision-making quality achieved.[10] Interpreting the correlation coefficient as a data quality metric, we can further conclude from Table 5 that the nudge treatment increased data quality, more precisely, the proportion of high quality responses, by 143% to 371%. A lowerbound estimate of 42% to 45% of maximum decision-making quality is achieved using the nudge protocol.

5.2 Is MSB a Good Proxy for Data Quality?

Although the literature generally views MSB as indicative of low decision-making quality, potential low decision-making quality among non-multiple switchers is not well understood. Here we can separately identify decision-making quality and MSB. We explicitly test decision-making quality for multiple switchers and non-multiple switchers in Table 7. The results show that data quality is insignificantly different from 0 for both multiple switchers and non-multiple switchers in the control group, and is only significantly different from 0 for the non-multiple switchers in the treatment group. Data quality under the nudge treatment is

[10] An alternative method is to use the smaller of $Var(MPL_j)$ and $Var(LS_j)$ as an upperbound estimate of σ_s^2, which, according to Table 6, is $Var(LS_j)$. In this case, our data quality metric would be $Cov(MPL_j, LS_j)/Var(LS_j)$, or the slope coefficient in a regression of MPL_j on LS_j.

significantly higher than the data quality of both multiple and non-multiple switchers under the standard protocol (p-values = 0.016 and 0.019, respectively). This implies that lack of MSB does not necessarily indicate high quality decision-making and the common practice of restricting the data to non-multiple switchers does not necessarily resolve data quality issues (Charness et al., 2013).

5.3 Cognitive Ability, MSB, and Data Quality

The previous literature shows that MSB is related to cognitive ability, in particular math ability (Dave et al., 2010; Meier and Sprenger, 2013). In the following analysis we first check whether subjects' multiple switching behaviors are correlated with their school test scores and then check if a relationship exists between test scores and data quality. The tests are uniform across the middle schools in the county, and the test results are provided to us by the school administrators.

Table 8 reports the results from the linear regression of multiple switching on test scores for the control and treatment groups separately. Test scores are the average of standardized math and standardized verbal test scores, standardized within each grade. Column 1 shows that, in line with our expectations, individuals' cognitive ability is strongly correlated with MSB in the standard MPL protocol. Column 2 adds a set of control variables: gender, monthly household income, mother's educational attainment, mother's occupation, the number of children in the household, and grade fixed effects. The results are essentially unchanged. Column 2 shows that an increase of one standard deviation in test scores is associated with a 14.9 percentage point decrease in the likelihood of multiple switching. This corresponds to a 49% (0.112/0.307) reduction in the probability of multiple switching. Appendix Table A.2 shows that both math and low verbal scores are significant predictors of MSB. These findings suggest that one reason that multiple switchers using the standard MPL protocol make worse financial decisions (Jacobson and Petrie, 2009) could be their lower cognitive ability, which leads to both MSB and poor financial decision-making.

Columns 3 and 4 in Table 8 show that cognitive ability no longer predicts MSB when using the nudged MPL protocol. Appendix Table A.3 Table shows that this is also the case for both verbal and math test scores.

Because MSB is not a good proxy for data quality, to examine the relationship between decision-making quality and cognitive ability, Table 9 reports the correlation between the responses on the LS and MPL tasks by overall cognitive ability, for the control and treatment groups. The results show that data quality is low (insignificantly different from 0) for both high and low cognitive ability individuals using the standard MPL. For the treatment group, data quality is significantly different from 0 for both high and low cognitive ability individuals. The point estimates indicate that low cognitive ability individuals exhibited higher quality decision-making with the nudge protocol than high cognitive ability individuals did using the standard protocol. Table 9 also shows that cognitive ability is not a significant determinant of data quality, for either the control or treatment group.[11]

6 Conclusion

In this study we developed a conceptual framework defining decision-making quality in the MPL to test several prominent explanations of low decision-making quality suggested by the findings in the literature using a novel experimental design and treatment protocol. In a departure from previous literature, our study provides a direct test of and finds evidence in support of task specific miscomprehension as the explanation for low quality decision-making in the MPL.

Our framework further lead us to propose a novel metric to quantify data quality separately from MSB and we showed that MSB is not a good proxy for data quality, as is often believed. Using this metric we show that the nudge treatment conservatively increased high

[11] Using the regression coefficient in a regression of MPL responses on LS responses as an alternative data quality metric, we found that in a regression of MPL responses on LS responses interacted with test scores, the interaction term was insignificant for both the control and treatment groups (control group p-value = 0.789; treatment group p-value = 0.836), confirming the finding that cognitive ability is not a significant determinant of data quality.

quality responses by 143%. Data quality improvement is not limited to the multiple switchers. Data quality under the nudge protocol is significantly higher than data quality of both multiple and non-multiple switchers under the standard protocol.

We find that cognitive ability explains MSB in the standard protocol, but it is not a significant determinant of data quality. To the extent that the nudge treatment was able to elicit more cognitive effort from respondents, the findings suggest that cognitive ability is not a limiting factor in achieving high data quality, but cognitive effort is. This does not preclude cognitive ability from influencing decision-making quality in other areas of life, or in other instruments. Our findings speak narrowly to the MPL instrument, and imply that protocol design innovations in the MPL can reveal risk preferences that would otherwise be obscured due to poor decision-making quality.

These findings may be particularly relevant for researchers who would like to employ the MPL but are concerned that the subject pools they study will be prone to MSB. More broadly, our findings point to the importance of investing in efforts to increase subject comprehension and data quality using the MPL, which resonate with the conclusions in Dave et al. (2010) and Charness and Viceisza (2016). One strategy can be to use a nudge protocol similar to ours in conjunction with any existing MPL protocol. Other strategies include using the framing device in Andreoni and Sprenger (2011), discussed previously, reading instructions out loud in addition to providing written instructions (Bruner, 2011), and using a visual representation of the MPL (Bauermeister and Musshoff, 2016).

References

Agranov, Marina and Pietro Ortoleva, "Stochastic choice and preferences for randomization," *Journal of Political Economy*, 2017, *125* (1), 000–000.

Andersen, Steffen, Glenn W Harrison, Morten I Lau, and E Elisabet Rutström, "Eliciting risk and time preferences," *Econometrica*, 2008, *76* (3), 583–618.

_ , _ , Morten Igel Lau, and E Elisabet Rutström, "Elicitation using multiple price list formats," *Experimental Economics*, 2006, *9* (4), 383–405.

Andreoni, James and Charles Sprenger, "Uncertainty equivalents: Testing the limits of the independence axiom," Technical Report, National Bureau of Economic Research 2011.

Barr, Abigail and Garance Genicot, "Risk sharing, commitment, and information: an experimental analysis," *Journal of the European Economic Association*, 2008, *6* (6), 1151–1185.

Bauermeister, Golo and Oliver Musshoff, "Risk Aversion and Inconsistencies-Does the Choice of Risk Elicitation Method and Display Format Influence the Outcomes?," Technical Report, Agricultural and Applied Economics Association 2016.

Binswanger, Hans P, "Attitudes toward risk: Experimental measurement in rural India," *American journal of agricultural economics*, 1980, *62* (3), 395–407.

Brick, Kerri, Martine Visser, and Justine Burns, "Risk aversion: experimental evidence from South African fishing communities," *American Journal of Agricultural Economics*, 2012, *94* (1), 133–152.

Bruner, David, "Multiple switching behavior in multiple price lists," *Applied Economics Letters*, 2011.

Cassar, Alessandra, Feven Wordofa, and Y. Jane Zhang, "Competing for the benefit of offspring eliminates the gender gap in competitiveness," *Proceedings of the National Academy of Sciences*, 2016, p. 201520235.

Charness, Gary and Angelino Viceisza, "Three risk-elicitation methods in the field: Evidence from rural Senegal," *Review of Behavioral Economics*, 2016.

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

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

Dave, Chetan, Catherine Eckel, Cathleen Johnson, and Christian Rojas, "Eliciting risk preferences: When is simple better?," *Journal of Risk and Uncertainty*, December 2010, *41* (3), 219–243.

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

Eckel, Catherine C and Philip J Grossman, "Sex differences and statistical stereotyping in attitudes toward financial risk," *Evolution and human behavior*, 2002, *23* (4), 281–295.

Gaudecker, Hans-Martin Von, Arthur Van Soest, and Erik Wengström, "Heterogeneity in risky choice behavior in a broad population," *The American Economic Review*, 2011, *101* (2), 664–694.

Harrison, Glenn and Elisabet Rutström, "Risk aversion in the laboratory," in "Risk aversion in experiments," Emerald Group Publishing Limited, 2008, pp. 41–196.

Harrison, Glenn W, Morten I Lau, and Melonie B Williams, "Estimating individual discount rates in Denmark: A field experiment," *The American Economic Review*, 2002, *92* (5), 1606–1617.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, December 2002, *92* (5), 1644–1655.

Jacobson, Sarah and Ragan Petrie, "Learning from mistakes: What do inconsistent choices over risk tell us?," *Journal of Risk and Uncertainty*, 2009, *38* (2), 143–158.

Kahneman, Daniel, "Maps of bounded rationality: Psychology for behavioral economics," *The American economic review*, 2003, *93* (5), 1449–1475.

_ , *Thinking, fast and slow*, Macmillan, 2011.

_ , Jack L Knetsch, and Richard H Thaler, "Experimental tests of the endowment effect and the Coase theorem," *Journal of political Economy*, 1990, pp. 1325–1348.

Meier, Stephan and Charles D Sprenger, "Discounting financial literacy: Time preferences and participation in financial education programs," *Journal of Economic Behavior & Organization*, 2013, *95*, 159–174.

Niederle, Muriel, "Gender," in "Handbook in Experimental Economics, second edition," Princeton University Press, 2016.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen, "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam," *American Economic Review*, March 2010, *100* (1), 557–71.

Thaler, Richard H. and Cass R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*, Yale University Press, 2008.

Figures and Tables

Figure 1: Conceptual Framework

Mixture Distribution Weight	I - Stable attribute		II - Task specific miscomprehension		III - Carelessness	
	MPL - standard	LS	MPL - standard	LS	MPL - standard	LS
p	$X_{1s} = S + \eta_{1s}$	$Y_{1s} = S + \nu_{1s}$	$X_{1s} = S + \eta_{1s}$	$Y_{1s} = S + \nu_{1s}$	$X_{1s} = S + \eta_{1s}$	$Y_{1s} = S + \nu_{1s}$
1-p	$X_{2s} = \eta_{2s}$	$Y_{2s} = \nu_{2s}$	$X_{2s} = \eta_{2s}$	$Y_{2s} = S + \nu_{2s}$	$X_{2s} = \eta_{2s}$	$Y_{2s} = \nu_{2s}$
	MPL - nudged	LS	MPL - nudged	LS	MPL - nudged	LS
p	$X_{1n} = S + \eta_{1n}$	$Y_{1n} = S + \nu_{1n}$	$X_{1n} = S + \eta_{1n}$	$Y_{1n} = S + \nu_{1n}$	$X_{1n} = S + \eta_{1n}$	$Y_{1n} = S + \nu_{1n}$
1-p	$X_{2n} = \eta_{2n}$	$Y_{2n} = \nu_{2n}$	$X_{2n} = S + \eta_{2n}$	$Y_{2n} = S + \nu_{2n}$	$X_{2n} = S + \eta_{2n}$	$Y_{2n} = \nu_{2n}$
Result	$Cov(MPL_s, LS_s) = Cov(MPL_n, LS_n)$; $Corr(MPL_s, LS_s) = Corr(MPL_n, LS_n)$		$Cov(MPL_s, LS_s) < Cov(MPL_n, LS_n)$		$Cov(MPL_s, LS_s) \geq Cov(MPL_n, LS_n)$	

Table 1: Balance Check and MSB

	Control	Treatment	P-values for
	(1)	(2)	H0: (1)=(2)
Female	.51	.61	0.192
	(.5)	(.49)	
Age	14.42	14.34	0.671
	(1.2)	(1)	
Number of children in the household	2.09	2.26	0.197
	(.87)	(.92)	
Number of family members in the household	5.34	5.72	0.290
	(2.05)	(2.8)	
Distance from home to school (=1 if less than or equal to 30min walk)	.38	.43	0.410
	(.49)	(.5)	
Mother's educational attainment (=1 if less than or equal to primary)	.68	.68	0.981
	(.47)	(.47)	
Mother's occupation (=1 if agricultural)	.71	.77	0.354
	(.45)	(.42)	
Monthly hh income (=1 if less than or equal to 750RMB)	.45	.45	0.999
	(.5)	(.5)	
Monthly allowance (=1 if less than or equal to 300RMB)	.82	.78	0.497
	(.38)	(.41)	
Test score	.08	-.09	0.177
	(.82)	(.92)	
Multiple switcher	.31	.1	0.000
	(.46)	(.3)	
Observations	101	92	

Notes: Means and standard deviations are presented. Standard deviations in parentheses. Exchange rate: 1RMB = 0.16 US dollars.

Table 2: Distribution of Responses in the Multiple Price List

Number of risky choices	Implied CRRA coefficient range	% of subjects
0	$r > 0.69$	17.10%
1	$0.57 < r < 0.69$	13.99%
2	$0.42 < r < 0.57$	18.13%
3	$0.24 < r < 0.42$	7.77%
4	$0 < r < 0.24$	8.81%
5	$-0.36 < r < 0$	7.25%
6	$r < -0.36$	26.94%

Notes: $N = 193$. The number of risky choices is coded using the number of risky choices made before the first point at which individuals switch from the risky choice to the safe choice. The implied CRRA coefficient range is calculated as the range of r in the function $u = x^{1-r}/(1-r)$ for which the subject makes the corresponding number of risky choices, assuming no multiple switching.

Table 3: Distribution of Responses in the Lottery Selection Task by Treatment

Lottery	Low payoff	High payoff	Implied CRRA coefficient range	% of subjects	
				Control	Treatment
1	3	3	$r > 4.17$	15.84%	13.04%
2	2.5	5	$0.99 < r < 4.17$	11.88%	9.78%
3	2	6	$0.81 < r < 0.99$	35.64%	29.35%
4	1.5	7.5	$0.32 < r < 0.81$	6.93%	17.39%
5	0.5	9	$0 < r < 0.32$	7.92%	8.70%
6	0	10	$r < 0$	21.74%	21.78%

Notes: $N = 101$ for the control group; $N = 92$ for the treatment group. The implied CRRA coefficient range is calculated as the range of r in the function $u = x^{1-r}/(1-r)$ for which the subject chooses each lottery.

Table 4: Covariance between Responses on the MPL and LS Tasks

	Control (1)	Treatment (2)	P-values of H0: (1) = (2)
Method A: Number of risky choices			
Covariance with LS task	0.300	1.514***	0.014
Method B: First switch point			
Covariance with LS task	0.713*	1.670***	0.076
Method C: Average switch point			
Covariance with LS task	0.489	1.580***	0.028
Randomly generated MPL choices			
Mean of Covariance with LS task	-0.0032	-0.0006	0.4941
Standard Error of Covariance with LS task	0.3825	0.3928	
N	101	92	

Notes: Method A defines MPL response as the total number of risky choices; Method B defines MPL response as the number of risky choices made before the "first switch point." Method C defines MPL response as the average switch point when the subject exhibits MSB. Randomly generated MPL choices uses 10,000 bootstrap samples of MPL choices from the empirical distribution of the number of risky choices made in the MPL (coded using method B), separately for the control and treatment groups. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 5: Correlation between Responses on the MPL and LS Tasks

	Control (1)	Treatment (2)	P-values of H0: (1) = (2)
Method A: Number of risky choices			
Correlation with LS task	0.090	0.424***	0.014
Method B: First switch point			
Correlation with LS task	0.185*	0.449***	0.043
Method C: Average switch point			
Correlation with LS task	0.135	0.440***	0.021
Randomly generated MPL choices			
Mean of Correlation with LS task	-0.0009	-0.0001	0.5000
Standard Error of Correlation with LS task	0.0995	0.1061	
N	101	92	

Notes: Method A defines MPL response as the total number of risky choices; Method B defines MPL response as the number of risky choices made before the "first switch point." Method C defines MPL response as the average switch point when the subject exhibits MSB. Randomly generated MPL choices uses 10,000 bootstrap samples of MPL choices from the empirical distribution of the number of risky choices made in the MPL (coded using method B), separately for the control and treatment groups. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6: Variance of Responses on the MPL and LS Tasks

	Control (1)	Treatment (2)	P-values of H0:(1) = (2)
Method A: Number of risky choices			
Variance of MPL	3.809	4.692	0.308
Method B: First switch point			
Variance of MPL	5.086	5.101	0.986
Method C: Average switch point			
Variance of MPL	3.974	4.753	0.382
Variance of LS	2.929	2.716	0.715
N	101	92	

Notes: Method A defines MPL response as the total number of risky choices; Method B defines MPL response as the number of risky choices made before the "first switch point." Method C defines MPL response as the average switch point when the subject exhibits MSB.

Table 7: MSB and Data Quality - Correlation between MPL and LS Responses

	Multiple Switchers (1)	Non-multiple Switchers (2)	P-values of H0: (1)=(2)
Control	-0.040	0.104	0.520
N	31	70	
Treatment	0.000	0.446***	
N	9	93	

Notes: MPL is coded using Method B, the number of risky choices made before the "first switch point." The p-value for the difference in correlations for the treatment group could not be calculated because the N for multiple switchers is too small. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 8: Cognitive Ability and MSB in the MPL

Dep.Var.= 1 if multiple switcher in MPL task

	Control		Treatment	
	Mean = 0.307		Mean = 0.098	
	(1)	(2)	(3)	(4)
Test scores	-0.136**	-0.149***	-0.037	-0.050
	(0.053)	(0.054)	(0.027)	(0.032)
Other controls	No	Yes	No	Yes
N	99	95	92	87

Notes: Students' test scores are the average of standardized math and standardized verbal tets scores within each grade. Robust standard errors are in parentheses. Other controls include: gender, monthly household income, mother's educational attainment, mother's occupation, number of children in the household, and grade fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 9: Cognitive Ability and Data Quality - Correlation between MPL and LS Responses

	Low	High	P-values of
	(1)	(2)	H0:(1)=(2)
Control	0.106	0.230	0.533
N	52	49	
Treatment	0.419***	0.477***	0.734
N	46	46	

Notes: MPL is coded using Method B, the number of risky choices made before the "first switch point." High cognitive ability is defined as above having median test scores (average of standardized verbal and math scores) in the control and treatment groups separately. * significant at 10%, ** significant at 5%, *** significant at 1%.

A Appendix Figures and Tables

Table A.1: Testing the Balance of Selected Observables by MPL II

	Control		Treatment		P-values for H0: (1)=(2)=(3)=(4)
	LS first	MPL first	LS first	MPL first	
	(1)	(2)	(3)	(4)	
Female	.49	.54	.65	.57	0.451
	(.51)	(.5)	(.48)	(.5)	
Age	14.23	14.55	14.3	14.38	0.568
	(1.3)	(1.11)	(.99)	(1.02)	
Number of children in the household	2	2.18	2.35	2.17	0.324
	(.76)	(.95)	(.92)	(.93)	
Number of family members in the household	5.17	5.5	5.33	6.09	0.291
	(2.11)	(2)	(2.13)	(3.29)	
Distance from home to school (=1 if less than or equal to 30min walk)	.43	.33	.41	.46	0.589
	(.5)	(.47)	(.5)	(.5)	
Mother's educational attainment (=1 if less than or equal to primary)	.65	.71	.61	.76	0.418
	(.48)	(.46)	(.49)	(.43)	
Mother's occupation (=1 if agricultural)	.73	.69	.78	.76	0.766
	(.45)	(.47)	(.42)	(.43)	
Monthly hh income (=1 if less than or equal to 750RMB)	.41	.48	.48	.41	0.821
	(.5)	(.5)	(.51)	(.5)	
Monthly allowance (=1 if less than or equal to 300RMB)	.8	.85	.85	.72	0.346
	(.41)	(.36)	(.36)	(.46)	
Test score	.06	.1	-.01	-.17	0.455
	(.82)	(.83)	(.86)	(.99)	
Observations	49	52	46	46	

Notes: Notes: Means and standard deviations are presented. Standard deviations in parentheses. Exchange rate: 1RMB = 0.16 US dollars.

Table A.2: Cognitive Ability and MSB in the MPL - Control group

Dep.Var.= 1 if multiple switcher in MPL task

	(1)	(2)	(3)	(4)
Math test score	-0.091**	-0.112**		
	(0.044)	(0.048)		
Verbal test score			-0.117**	-0.125**
			(0.049)	(0.056)
Other controls	No	Yes	No	Yes
N	99	95	99	95

Notes: Students' test scores are standardized within each grade. Robust standard errors are in parentheses. Other controls include: gender, monthly household income, mother's educational attainment, mother's occupation, number of children in the household, and grade fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table A.3: Cognitive Ability and MSB in the MPL - Treatment group

Dep.Var.= 1 if multiple switcher in MPL task

	(1)	(2)	(3)	(4)
Math test score	-0.024	-0.035		
	(0.031)	(0.033)		
Verbal test score			-0.033	-0.045
			(0.021)	(0.028)
Other controls	No	Yes	No	Yes
N	92	87	92	87

Notes: Students' test scores are standardized within each grade. Robust standard errors are in parentheses. Other controls include: gender, monthly household income, mother's educational attainment, mother's occupation, number of children in the household, and grade fixed effects. * significant at 10%, ** significant at 5%, *** significant at 1%.

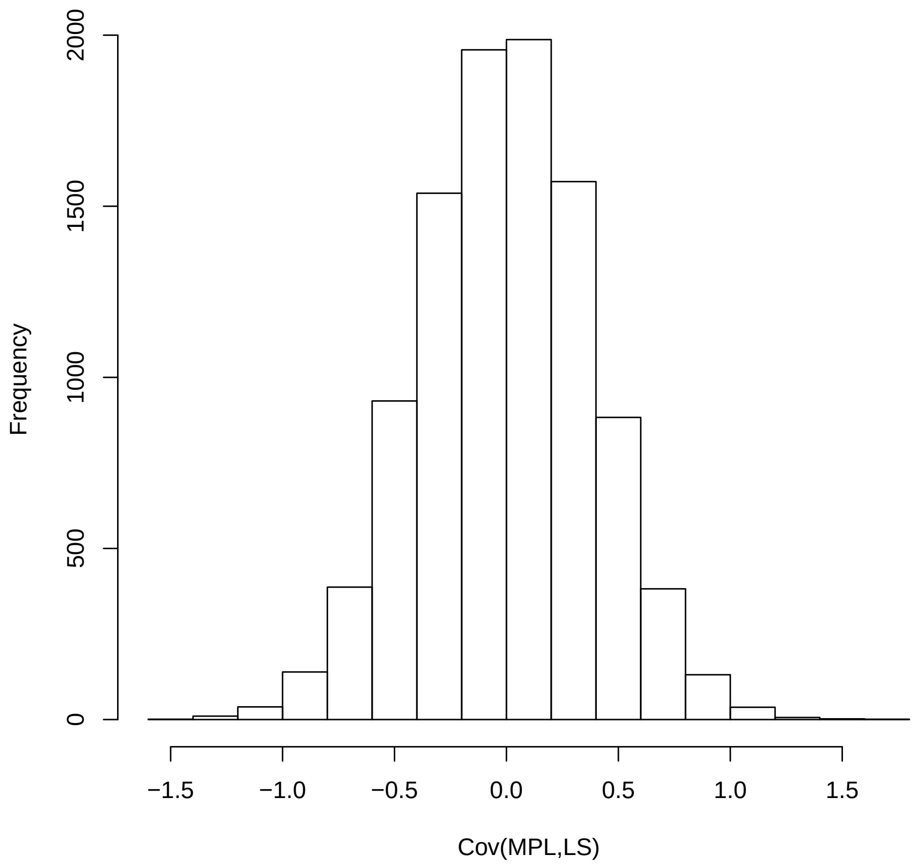


Figure A.1: Histogram of the covariance of the responses on the MPL and LS Tasks using bootstrap MPL samples.

B Proofs

B.1 Conceptual Framework

Scenario II In the general case under Scenario II, the nudge treatment unconfuses a proper subset of individuals. To describe this scenario we introduce a third type. Type 3 individuals, occuring with probability p_3, are confused using the MPL even when nudged. Type 1 and Type 2 occur with probability p_1 and p_2, respectively, with $p_1 + p_2 + p_3 = 1$.

Under this general Scenario II, for the control group using the MPL, the response of Type 1 individuals is $X_{1s} = S + \eta_{1s}$; the response of Type 2 individuals is $X_{2s} = \eta_{2s}$; and the response of Type 3 individuals is $X_{3s} = \eta_{3s}$, $\eta_{2s} \overset{d}{=} \eta_{3s}$. For the control group using the LS task, because the confusion is specific to the MPL, the response of Type 1, Type 2 and Type 3 individuals are $Y_{1s} = S + \nu_{1s}$, $Y_{2s} = S + \nu_{2s}$, and $Y_{3s} = S + \nu_{3s}$, respectively, where $\nu_{1s} \overset{d}{=} \nu_{2s} \overset{d}{=} \nu_{3s}$.

For the treatment group using the MPL, the response of Type 1 individuals is $X_{1n} = S + \eta_{1n}$; the response of Type 2 individuals is $X_{2n} = S + \eta_{2n}$; and the response of Type 3 individuals is $X_{3n} = \eta_{3n}$. $\eta_{1n} \overset{d}{=} \eta_{2n} \overset{d}{=} \eta_{1s}$; $\eta_{3n} \overset{d}{=} \eta_{2s} \overset{d}{=} \eta_{3s}$.

Because the nudge treatment works only on the MPL, we expect no differences in the responses on the LS task by treatment status, for all three types of individuals. That is to say, $Y_{1n} = S + \nu_{1n}$, $Y_{2n} = \nu_{2n}$, and $Y_{3n} = \nu_{3n}$, where $\nu_{1n} \overset{d}{=} \nu_{1s}$, $\nu_{2n} \overset{d}{=} \nu_{2s}$, and $\nu_{3n} \overset{d}{=} \nu_{3s}$.

Then, for the control group, $Cov(MPL_s, LS_s) = p_1 Cov(X_{1s}, Y_{1s}) + p_2 Cov(X_{2s}, Y_{2s}) + p_3 Cov(X_{3s}, Y_{3s}) = p_1 \sigma_s^2$, and for treatment group, $Cov(MPL_n, LS_n) = (p_1 + p_2)\sigma_s^2$. This yields the result that $Cov(MPL_n, LS_n) > Cov(MPL_s, LS_s)$ if $p_2 > 0$.

Scenario III In the general case under Scenario III, the nudge treatment unconfuses a proper subset of individuals. As in the section above, we introduce a third type. Type 3 individuals, occuring with probability p_3, are confused using the MPL even when nudged. Type 1 and Type 2 occur with probability p_1 and p_2, respectively, with $p_1 + p_2 + p_3 = 1$.

Under this general Scenario III, for the control group using the MPL, the response of Type 1 individuals is $X_{1s} = S + \eta_{1s}$; the response of Type 2 individuals is $X_{2s} = \eta_{2s}$; and the response of Type 3 individuals is $X_{3s} = \eta_{3s}$. $\eta_{2s} \overset{d}{=} \eta_{3s}$.

For the control group using the LS task, the response of Type 1 individuals is $Y_{1s} = S + \nu_{1s}$; the response of Type 2 individuals is $Y_{2s} = \nu_{2s}$; and the response of Type 3 individuals is $Y_{3s} = \nu_{3s}$. $\nu_{2s} \overset{d}{=} \nu_{3s}$.

For the treatment group using the MPL, the response of Type 1 individuals is $X_{1n} = S + \eta_{1n}$; the response of Type 2 individuals is $X_{2n} = S + \eta_{2n}$; and the response of Type 3 individuals is $X_{3n} = \eta_{3n}$. $\eta_{1n} \overset{d}{=} \eta_{2n} \overset{d}{=} \eta_{1s}$; $\eta_{3n} \overset{d}{=} \eta_{2s} \overset{d}{=} \eta_{3s}$.

Because the nudge treatment works only on the MPL, we expect no differences in the responses on the LS task by treatment status. For the treatment group using the LS task, the response of Type 1 individuals is $Y_{1n} = S + \nu_{1n}$; the response of Type 2 individuals is $Y_{2n} = \nu_{2n}$; and the response of Type 3 individuals is $Y_{3n} = \nu_{3n}$. $\nu_{1n} \overset{d}{=} \nu_{1s}$, $\nu_{2n} \overset{d}{=} \nu_{2s}$, and $\nu_{3n} \overset{d}{=} \nu_{3s}$.

For the control group, we have $Cov(MPL_s, LS_s) = \sum_{j=1}^{3} p_i Cov(X_{is}, Y_{is}) + p_1 \mu_{X1s}\mu_{Y1s} + p_2 \mu_{X2s}\mu_{Y2s} + p_3 \mu_{X3s}\mu_{Y3s} - (\mu_{Xs})(\mu_{Ys})$, while for the treatment group, we have $Cov(MPL_n, LS_n) = \sum_{j=1}^{3} p_i Cov(X_{in}, Y_{in}) + p_1 \mu_{X1n}\mu_{Y1n} + p_2 \mu_{X2n}\mu_{Y2n} + p_3 \mu_{X3n}\mu_{Y3n} - (\mu_{Xn}\mu_{Yn})$.

Note that $\sum_{i=1}^{3} p_i Cov(X_{is}, Y_{is}) = \sum_{i=1}^{3} p_i Cov(X_{in}, Y_{in}) = p_1 \sigma_s^2$, so we can rearrange to obtain $Cov(MPL_s, LS_s) = Cov(MPL_n, LS_n) + p_2 \mu_{X2s}\mu_{Y2s} - p_2 \mu_{X2n}\mu_{Y2n} + \mu_{Xn}\mu_{Yn} - \mu_{Xs}\mu_{Ys} = Cov(MPL_n, LS_n) + p_2 \mu_{Y2s}(\mu_{X2s} - \mu_{X2n}) - \mu_{Ys}(\mu_{Xs} - \mu_{Xn}) = Cov(MPL_n, LS_n) + p_2 \mu_{Y2s}(\mu_{X2s} - \mu_{X2n}) - p_2 \mu_{Ys}(\mu_{X2s} - \mu_{X2n}) = Cov(MPL_n, LS_n) + p_2(\mu_{X2s} - \mu_{X2n})(\mu_{Y2s} - \mu_{Ys})$. Thus, as long as the difference in the expected value of measured risk tolerance in the control group for Type 1 and Type 2 is not in the opposite direction for the two tasks, i.e., either $E(X_{1s}) - E(X_{2s})$ and $E(Y_{1s}) - E(Y_{2s})$ have the same sign, or one or both of the differences is 0, then $(\mu_{X2s} - \mu_{X2n})(\mu_{Y2n} - \mu_{Yn}) \geq 0$, and we have $Cov(MPL_s, LS_s) \geq Cov(MPL_n, LS_n)$.

B.2 Test Statistic

To test the explanations in our conceptual framework, we need to compare the covariances between responses on the MPL and LS tasks for the control group which uses the standard protocol and the treatment group which uses the nudge protocol. To the best of our knowledge, covariance comparison tests have not been previously considered. Tests for correlations cannot be used because the variances in our case are indeterminate.

For notational convenience, let $M_s = MPL_s$ and $L_s = LS_s$ for the control group, and let $M_n = MPL_n$ and $L_n = LS_n$ for the treatment group.

Denote the sample covariance between M_s and L_s by

$$S_{M_sL_s} = \frac{1}{n_s - 1}\sum_{i=1}^{n_s}\Big[(M_{s_i} - \overline{M}_s)(L_{s_i} - \overline{L}_s)\Big], \tag{3}$$

where $\overline{M}_s$ is the average of the data $\{M_{s_i}, i = 1, \ldots, n_s\}$ for MPL, and $\overline{L}_s$ is defined for LS in a similar way.

The sample covariance (3) is used because it is unbiased, i.e. $E(S_{M_sL_s}) = Cov(M_s, L_s) = \sigma_{M_sL_s}$. It is easy to verify, so we omit the proof here. In what follows, we derive the variance of $S_{M_sL_s}$ to construct our test.

First, consider

$$\begin{aligned}
E\Big[(n_s-1)^2 S^2_{M_sL_s}\Big] &= E\Big[\sum_{i,j=1}^{n_s}\big[(M_{s_i} - \overline{M}_s)(L_{s_i} - \overline{L}_s)(M_{s_j} - \overline{M}_s)(L_{s_j} - \overline{L}_s)\big]\Big] \\
&= E\Big[\sum_{i,j=1}^{n_s}\big[(M_{s_i} - \mu_{M_s})(L_{s_i} - \mu_{L_s})(M_{s_j} - \mu_{M_s})(L_{s_j} - \mu_{L_s})\big]\Big] \\
&\quad - 2n_s E\Big[(\overline{M}_s - \mu_{M_s})(\overline{L}_s - \mu_{L_s})\sum_{j=1}^{n_s}(M_{s_j} - \mu_{M_s})(L_{s_j} - \mu_{L_s})\Big] \\
&\quad + n_s^2 E\Big[(\overline{M}_s - \mu_{M_s})^2(\overline{L}_s - \mu_{L_s})^2\Big],
\end{aligned} \tag{4}$$

where $\mu_{M_s} = E(M_s)$ and $\mu_{L_s} = E(L_s)$.

For the last term on the right hand side in (4), we have

$$\begin{aligned}
n_s^2 E\Big[(\overline{M}_s - \mu_{M_s})^2(\overline{L}_s - \mu_{L_s})^2\Big] &= \frac{1}{n_s^2}E\Big[\sum_{i,j,k,l=1}^{n_s}\big[(M_{s_i} - \mu_{M_s})(M_{s_j} - \mu_{M_s})(L_{s_l} - \mu_{L_s})(L_{s_k} - \mu_{L_s})\big]\Big] \\
&= \frac{1}{n_s^2}\sum_{i=1}^{n_s}E\Big[(M_{s_i} - \mu_{M_s})^2(L_{s_i} - \mu_{L_s})^2\Big] + \frac{n_s - 1}{n_s}\sigma^2_{M_s}\sigma^2_{L_s} + \frac{2(n_s-1)}{n_s}\sigma^2_{M_sL_s},
\end{aligned} \tag{5}$$

where $\sigma^2_{M_s} = Var(M_s)$ and $\sigma^2_{L_s} = Var(L_s)$.

Similarly, for the first and second terms on the right in (4), we have

$$E\Big[\sum_{i,j=1}^{n_s}\big[(M_{s_i} - \mu_{M_s})(L_{s_i} - \mu_{L_s})(M_{s_j} - \mu_{M_s})(L_{s_j} - \mu_{L_s})\big]\Big] = \sum_{i=1}^{n_s}E\Big[(M_{s_i} - \mu_{M_s})^2(L_{s_i} - \mu_{L_s})^2\Big] + n_s(n_s-1)\sigma^2_{M_sL_s}, \tag{6}$$

and

$$
\begin{aligned}
& -2n_s E\Big[(\overline{M}_s - \mu_{M_s})(\overline{L}_s - \mu_{L_s}) \sum_{j=1}^{n_s} (M_{s_j} - \mu_{M_s})(L_{s_j} - \mu_{L_s})\Big] \\
=& \frac{-2}{n_s} \sum_{i=1}^{n_s} E\Big[(M_{s_i} - \mu_{M_s})^2 (L_{s_i} - \mu_{L_s})^2\Big] - 2(n_s - 1)\sigma^2_{M_s L_s}. \qquad (7)
\end{aligned}
$$

Substituting (5)-(7) into (4) and then dividing both sides by $(n_s - 1)^2$ yields

$$
E\Big[S^2_{M_s L_s}\Big] = \frac{1}{n_s} E\big[(M_s - \mu_{M_s})^2 (L_s - \mu_{L_s})^2\big] + \frac{1}{n_s(n_s - 1)} \sigma^2_{M_s} \sigma^2_{L_s} + \frac{(n_s - 1)^2 + 1}{n_s(n_s - 1)} \sigma^2_{M_s L_s}, \qquad (8)
$$

where $E\big[(M_s - \mu_{M_s})^2 (L_s - \mu_{L_s})^2\big]$ is the common value of $E\Big[(M_{s_i} - \mu_{M_s})^2 (L_{s_i} - \mu_{L_s})^2\Big]$ for $i = 1, \ldots, n_s$.

Finally, the variance of $S_{M_s L_s}$ is given by

$$
Var(S_{M_s L_s}) = \frac{1}{n_s} Var(W_s) + \frac{\sigma^2_{M_s} \sigma^2_{L_s}}{n_s(n_s - 1)} (\rho^2_{M_s L_s} + 1), \qquad (9)
$$

where $W_s = (M_s - \mu_{M_s})(L_s - \mu_{L_s})$ and $\rho_{M_s L_s}$ is the correlation between M_s and L_s.

By the central limit theorem and the independence of standard and nudge groups, we have the result that

$$
\frac{(S_{M_s L_s} - S_{M_n L_n}) - (\sigma_{M_s L_s} - \sigma_{M_n L_n})}{\sqrt{Var(S_{M_s L_s}) + Var(S_{M_n L_n})}} \qquad (10)
$$

will have a limiting standard normal distribution.

Note that $Var(S_{M_s L_s})$ and $Var(S_{M_n L_n})$ are unknown, and need to be estimated in order to give a p-value for our test. We use

$$
\frac{1}{n_s} \widehat{Var}(W_s) + \frac{\sigma^2_{M_s} \sigma^2_{L_s}}{n_s(n_s - 1)} (\hat{\rho}^2_{M_s L_s} + 1) \qquad (11)
$$

to estimate $Var(S_{M_s L_s})$, where

$$
\widehat{Var}(W_s) = \frac{1}{n_s} \sum_{i=1}^{n_s} (W_{s_i} - S_{M_s L_s})^2 - (\overline{W}_s - S_{M_s L_s})^2,
$$

$W_{s_i} = (M_{s_i} - \overline{M}_s)(L_{s_i} - \overline{L}_s)$ for $i = 1, \ldots, n_s$, $\overline{W}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} W_{s_i} = \frac{n-1}{n} S_{M_s L_s}$, and $\hat{\rho}_{M_s L_s}$ is the sample Pearson's correlation. Similar estimates are used for $Var(S_{M_n L_n})$.

C Experimental Protocol

Verbal Instructions (Translation) – Control Group

Thank you for participating in our game today!

First, we want you to understand that this game has no influence on your academic performance, and there is no right or wrong answer. Your choice is only related to your personal preference. How much cash you can earn in the end essentially depends on your choices and luck.

Do you have any questions?

Now, we are going to explain these two games. You are going to play both of them. After that, we randomly choose one to realize your payment. (We are going to toss a coin, if the heads shows up, you will get your payment based on the results of game one; if the tail shows up, and you will get your payment based on the results of game two.)

Since your ID number is odd, you will play game one first. (If your id is even, you will play game two first). Please read the instructions carefully, and write your choice on the paper. In this game, there are six lotteries; you are allowed to choose one of them. For each lottery, there is 50/50 chance to win each payoff. For example, if you choose lottery 2 and this game is chosen for payout at the end, and the ping pang ball you draw is yellow, you will get 5 yuan.

Now, we are going to play game two and I will explain the instructions now. In this game, there are six groups and you will need to make a choice in each group. There are two choices – A and B—in each group, your task is to pick either A or B. If you pick A, it means you would like to enter into a lottery, there is a 50% probability you will win 10 Yuan and 50% probability you will get nothing. If you choose B, it means you don’t want to enter the lottery and would like to get the cash directly. If this game is chosen for payout at the end, we will randomly choose one of these six groups and realize your payment based on your choice (A or B). Now, you can read these six groups one by one, if you would like to choose A, put a mark in the blank; if you would like to choose B, put a mark on the other blank.

Verbal Instructions (Translation) – Treatment Group

Thank you for participating in our game today!

First, we want you to understand that this game has no influence on your academic performance, and there is no right or wrong answer. Your choice is only related to your personal preference. How much cash you can earn in the end essentially depends on your choices and luck.

Do you have any questions?

Now, we are going to explain these two games. You are going to play both of them. After that, we randomly choose one to realize your payment. (We are going to toss a coin, if the heads shows up, you will get your payment based on the results of game one; if the tail shows up, and you will get your payment based on the results of game two.)

Since your ID number is odd, you will play game one first. (If your id is even, you will play game two first). Please read the instructions carefully, and write your choice on the paper. In this game, there are six lotteries; you are allowed to choose one of them. For each lottery, there is 50/50 chance to win each payoff. For example, if you choose lottery 2 and this game is chosen for payout at the end, and the ping pang ball you draw is yellow, you will get 5 yuan.

Now, we are going to play game two and I will explain the instructions now. In this game, there are six groups and you will need to make a choice in each group. There are two choices – A and B—in each group, your task is to pick either A or B. If you pick A, it means you would like to enter into a lottery, there is a 50% probability you will win 10 Yuan and 50% probability you will get nothing. If you choose B, it means you don't want to enter the lottery and would like to get the cash directly. If this game is chosen for payout at the end, we will randomly choose one of these six groups and realize your payment based on your choice (A or B). Now, you can read these six groups one by one, if you would like to choose A, put a mark in the blank; if you would like to choose B, put a mark on the other blank.

Have you decided? You can think about your choices again carefully and can change your choices. If you would like, we can explain this game one more time.

ID: _______________

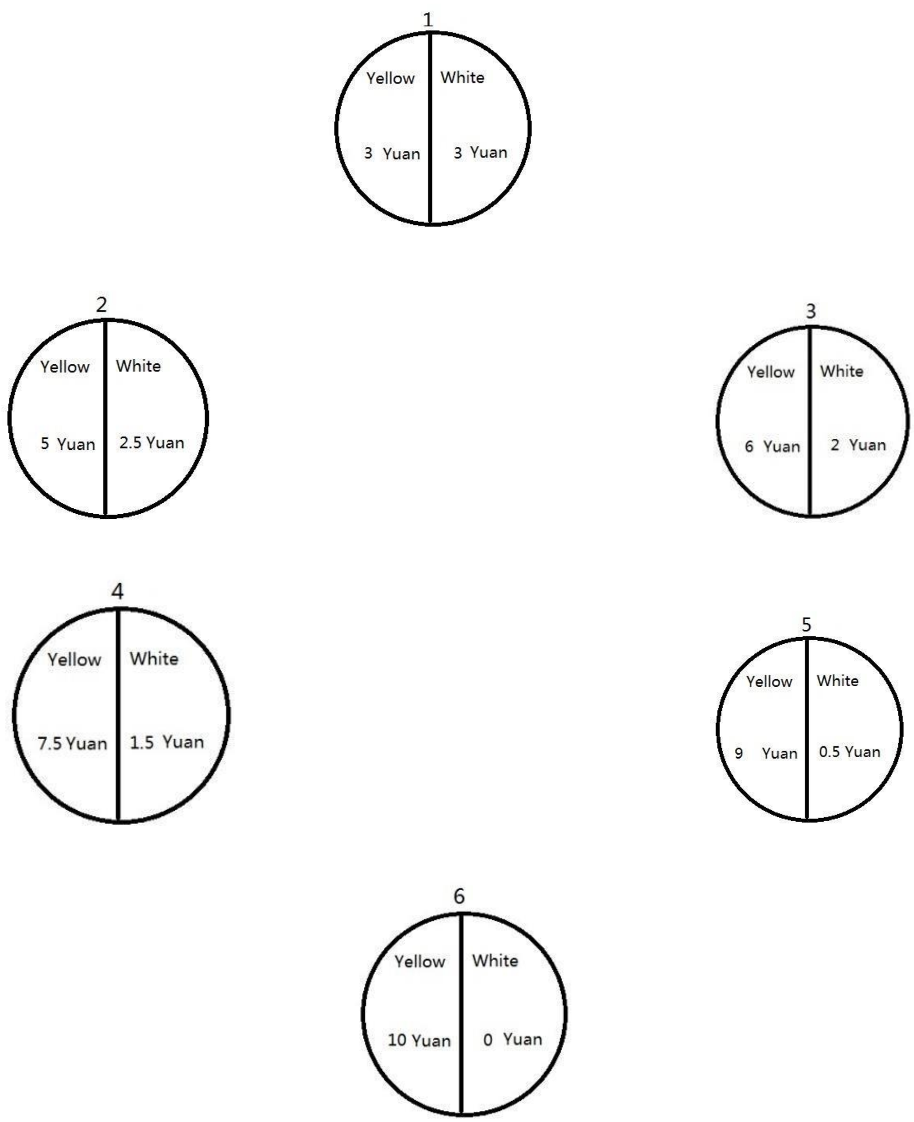
Experimental Instructions (Translation)

Hello everyone, welcome to today's game. Today's game has no impact whatsoever on your academic performance. Depending on your choices in the game, you will have a chance to win a cash prize, so please make your decisions carefully. This game is simple, but please listen carefully to our instructions, and if you have any questions, please raise your hand.

In this game, there are six groups of prizes, each group has two payouts. You can choose any one of the groups. After you've made your choice, how much money you can win depends on the color of the ping pong ball you draw. There are 2 ping pong balls in this black plastic bag, one is yellow and the other is white.

Please write down the number of the group you choose in the blank space below.

I would like to choose number_______.



Experimental Instructions (Translation)

Hello everyone, welcome to today's game. Today's game has no impact whatsoever on your academic performance. Depending on your choices in the game, you will have a chance to win a cash prize, so please make your decisions carefully. This game is simple, but please listen carefully to our instructions, and if you have any questions, please raise your hand.

In each of the 6 questions that we ask you, you must choose: Option A – lottery, or Option B – fixed payment. If you choose A, each outcome of Option A has 50% chance of occurring.

We will record your six choices, and then randomly choose a number between 1-6 to determine which choice will decide your payment.

ID:__________

1	☐ **Option A: Lottery**				☐ **Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 1 RMB
2	☐ **Option A: Lottery**				☐ **Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 2 RMB
3	☐ **Option A: Lottery**				☐ **Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 3 RMB
4	☐ **Option A: Lottery**				☐ **Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 4 RMB
5	☐ **Option A: Lottery**				☐ **Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 5 RMB
6	☐ **Option A: Lottery**				☐ **Option B: Fixed payment**
	 yellow = 10 RMB	**or**	 white = 0 RMB		 6 RMB