DISCUSSION PAPER SERIES

Opting-In: Participation Biases in the Lab

Robert Slonim
Carmen Wang
Ellen Garbarino
Danielle Merrett

September 2012

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Opting-In: Participation Biases in the Lab

Robert Slonim
*University of Sydney
and IZA*

Carmen Wang
University of Sydney

Ellen Garbarino
University of Sydney

Danielle Merrett
University of Sydney

ABSTRACT

Opting-In: Participation Biases in the Lab[*]

Assuming individuals rationally decide whether to participate or not to participate in lab experiments, we hypothesize several non-representative biases in the characteristics of lab participants. We test the hypotheses by first collecting survey and experimental data on a typical recruitment population and then inviting them to participate in a lab experiment. The results indicate that lab participants are not representative of the target population on almost all the hypothesized characteristics, including having lower income, working fewer hours, volunteering more often, and exhibiting behaviors correlated with interest in experiments and economics. We discuss the implications and various methods for addressing non-representative biases.

JEL Classification: C9

Keywords: participation bias, laboratory experiments

Corresponding author:

Robert Slonim
School of Economics
370 Merewether Building
University of Sydney
Sydney 2006
Australia
E-mail: robert.slonim@sydney.edu.au

1. Introduction

One of the most profound changes in the understanding of economic behavior and theory in recent decades has been the wealth of insights and inferences arising from laboratory economics experiments.[1] The lab's influence on economists' understanding of preferences and theory has been especially pronounced over the past 20 years due to its ability to control contextual factors and precisely measure subjects' behavior. In response to the lab's growing importance, researchers have become interested in whether the conclusions drawn from lab experiment participants extend to non-participants. Studies have explored this question in various directions, most commonly by comparing results from lab participants recruited across different populations.[2] However, few studies have assessed whether the characteristics of the lab participants are the same as the characteristics of the target population. Moreover, the few studies that have directly compared the characteristics of the people who chose to participate with those who chose not to participate have primarily focused on just one or two preferences (Cleave, Nikiforakis and Slonim 2010; Falk, Meier and Zehnder 2011).

In this paper, we study the representativeness of lab participants. We assume that people rationally choose to participate; people will participate if and only if the expected utility of participation exceeds the expected utility of not participating. Central to the participation decision will be the common information people receive about economic lab experiments: monetary earnings, time commitment and (minimal) task information. Based on this common information and using a standard model of expected utility, we derive four core hypotheses regarding how the characteristics of lab participants might differ from non-participants.

Specifically, we hypothesize that rational decision-making will result in lab participants being less wealthy, having more leisure time, having interests related to the tasks in economic lab experiments, and being more pro-social in the dimension of volunteering time. The more general pro-social characteristics of lab participants were also conjectured by Levitt and List (2007a, b). To date, however, studies have found little evidence to support this conjecture, either by directly comparing pro-social measures of participants with those who chose not to participate (Cleave et al. 2010; Falk et al. 2011) or by indirectly comparing whether pro-social behavior differs across subject

---

[1] E.g., by 1998 over 200 experimental economics papers were published annually (Holt, 2006).
[2] Cleave, Nikiforakis and Slonim (2010) review this literature. Many studies compare results across populations (across occupation: Cooper et al. 1999; Hannah, Kagel and Moser 2002; Fehr and List 2004; Carpenter, Burks and Verhoogen 2004; Güth, Schmidt and Sutter 2007; Carpenter, Connolly and Meyers 2008; Andersen et al. 2009; Anderson et al. 2010; Hoffman and Morgan 2011; across age: Harbaugh et al. 2003, across nationality/culture: Roth et al. 1991; Henrich et al., 2001; Herrmann, Thöni and Gächter 2008; Cameron et al. 2009; Henrich, Heine and Norenzayan 2010, and between students and a representative sample of the population: Bellemare and Kröger 2007; Alatas et al. 2009; Anderson et al. 2010; Belot, Duch and Miller 2010; Remoundou et al. 2010; Brañas-Garza et al. 2011; and Falk et al. 2011). Other studies compare behavior across context (e.g., laboratory and field) using the same population (see Eckel and Grossman 2000; Güth, Schmidt and Sutter 2007; Benz and Meier 2008; Baran, Sapienza and Zingales 2010; and Glaeser et al. 2000). Other approaches explore whether participants select *out* of experiments after some experience (Casari, Ham and Kagel 2007) and whether recruitment procedures affect selection into experiments (which also follows from our hypotheses): Harrison, Lau and Rutström (2009) find that participants take greater risks when the advertised earnings have greater variation, and Krawczwyk (2011) finds that an invitation stressing monetary benefits attracted more students to register on ORSEE (Greiner 2004) than the invitation stressing non-pecuniary benefits.

populations (see a review by Bowles and Polonia-Reyes, 2011). However, these studies have focused on pro-social **monetary** decisions, whereas since (1) experimenters invite participants to *sacrifice time in exchange for receiving money* and (2) monetary donations may be positively correlated with wealth (which we predict will have a negative effect on participation), testing whether lab participants are more pro-social should also examine pro-social volunteer time in addition to pro-social monetary decisions. We also present additional hypotheses regarding risk attitudes and patience that could be important depending on the idiosyncratic features of the recruitment process (e.g., experiments that offer a large show up fee may minimize participation bias with respect to financial risk attitudes).

To examine the hypotheses, we collected individual level data from a large classroom population (n=892) that completed a 20-item survey and three incentivized experimental tasks (our classroom procedures closely follow Cleave et al. 2010, but we used entirely different measures). The classroom population was subsequently invited to attend a lab experiment. The invited population was informed that the experiment would take 60 to 90 minutes and they would receive at least $30 (roughly twice the minimum wage).[3] Our core analyses compare characteristics of the classroom population who chose to attend the lab experiment with those who chose not to.

Our results unambiguously document that the characteristics of participants attending the experiment differ from those who chose not to attend on all the core dimensions we hypothesize. In particular, the lab participants have significantly less income, more leisure time, are more interested in the economic lab tasks (e.g., more likely to major in economics) and are more pro-social on the dimension of volunteering. We also find that the characteristics of lab participants differ on other hypothesized characteristics that may be more idiosyncratic (e.g., patience). Moreover, the magnitudes of these effects suggest substantial non-representative biases among participants; even small differences in the participation decision across characteristics can lead to large non-representative biases because of the low overall participation rates in typical experiment recruitment.[4] To see this, suppose the population is *divided equally* between people with characteristics A and B who are respectively 4 and 8 percent likely to participate in an experiment. In this case, the *four percent absolute difference* (8%-4%) in the decision to participate results in a **2 to 1** (8%/4%) ratio of B to A types in the lab relative to the equal division in the population. Most of our results show greater over-representative ratios than in this example. Thus, the contribution of this paper goes well beyond being the first empirical evidence documenting specific non-representative biases; we provide clear evidence that in several cases the magnitudes of the non-representative biases are dramatic.

This paper is organized as follows. **Sections 2**, **3** and **4** present the hypotheses, design and results, respectively, on the non-representative characteristics of the population attending the lab. **Section 5** discusses several methods to deal with participation bias and **Section 6** concludes. Before turning to

---

[3] The time and payment offered was consistent with other economics experiments at the university and across the country.
[4] Few studies report participation rates. The few examples we found include Cleave et al. (2010) and Falk et al. (2011) indicating approximately 12% participation rate and Krawczyk (2011) reports less than 2%.

the details, we first discuss why inferences from economic experiments can be biased by subjects' participation decisions.

1.1 Why participation bias can affect inferences in lab experiments

Economists have generally recognized since Heckman (1979) that influences on the participation decision can critically affect inference from a study. For instance, in evaluating an education program (e.g., Head Start), if more educationally motivated families with higher resources choose to participate in the program, then ignoring the participation decision can lead to an over-estimation of the program's success. Since laboratory and (many) field experiments allow a participation decision,[5] participation bias is possible. An inference concerning a population's preferences (e.g., risk, social or discounting) based on behavior in the experiment will be biased if the recruitment procedures systemically attract subjects with certain characteristics (e.g., less risk-seeking, more pro-social, more patient).

Although random assignment of participants in laboratory experiments can reduce concerns of many biases, and recent studies have found behavior of laboratory subjects recruited from different populations do not in general differ, there are three important reservations on generalizing inferences from lab studies to broader populations:

(1) a cursory review of the experimental literature shows that many inferences are still made from *point estimates* in an experiment, including some very influential studies;

(2) inferences from *treatment effects in which subjects have been randomly assigned to treatment* are not immune to participation biases, as discussed by Al-Ubaydli and List (2012); and

(3) the existing evidence showing *few differences in behavior between subject populations* has its limitations if participation biases exist in recruitment from those populations, especially if these biases are similar we will observe smaller differences among participants.

We carefully examine each of these points in the following paragraphs and show that potential participation biases due to experiment recruitment remains a real and unavoidable concern that should always be considered when designing an experiment or drawing inferences from experimental results.

First, while many experimental studies draw inferences from randomization of subjects to treatment, not *all* experiments rely on within experiment randomization, and some seminal experiments' major contributions come from point estimates. We define point estimates to include not only estimates of the frequency of a specific choice (e.g., X% of offers less than half of the pie were rejected in an ultimatum game or Y% of the subjects transferred nothing in a trust game) but also estimates to characterize subject *types* that can be inferred from many decisions across treatments in an experiment (e.g., Z% of the subjects *exhibited preference reversals, fell prey to the winners curse, are hyperbolic, are other regarding, are inequity averse*, etc.). A review of the experimental literature

---

[5] One exception is the natural field experiment in which there is no participation decision (Harrison and List 2004)

suggests inferences from point estimates are not rare. Seminal papers in the following areas have all presented point or type estimates: social preferences, risk preferences, preference reversals and (bounded) rationality.[6] Based on the tens of thousands of citations, point estimates and inferences on subject "types" have had numerous and profound impacts on future studies and theoretical developments.

Second, randomly assigning subjects to treatment has no *a priori* reason to alleviate participation bias concerns (Al-Ubaydli and List 2012). Within experiment treatment comparisons provide an estimate on the effect of behavior b between treatments t_0 and t_1, $E[b(t_1 \mid p=1)] – E[b(t_0 \mid p=1)]$, where $p=1$ indicates the observed people in the population who chose to participate. The critical question for extending the inference beyond the immediate subject population is whether this behavior holds for people who chose not to participate, $p=0$, as well as for anyone else not invited whom we wish to generalize to, i.e., whether $E[b(t_1 \mid p=0)] – E[b(t_0 \mid p=0)] = E[b(t_1 \mid p=1)] – E[b(t_0 \mid p=1)]$. Within experiment randomization alone does not make inferences from one sample generalizable to other, different populations, e.g. if women are completely missing in an experiment, randomization alone would not satisfactorily make the treatment effects on male sample generalizable to female population. It follows that inference from a treatment effect in an experiment could under-estimate, over-estimate or even change the sign of the effect for the non-participants (see Al-Ubaydli and List (2012) for a formal analysis). For example, several studies reporting behavior changes with 'high stake' payoffs were conducted in developing countries to ensure experimental payoffs were high relative to income. This practice is predicated on the assumption that the wealth level of a sample affects treatment effects (measured by within experiment random assignment); i.e., this practice assumes that the effect of stakes would be different if the same absolute change in stakes in the lab was relatively small compared to the wealth of the participants.[7] Given that the current experiment finds low income people are much more likely to participate in an experiment, stakes in experiments are likely to bias results towards the behavior of the over-represented low income participants.

Finally, participation bias predicts that homogenization of subject populations will occur. Several recent studies reporting few differences in behavior observed between various subject populations suggest that the experimental results are generalizable to broader populations. This inference implicitly assumes no bias in the participation decision: with no participation bias, subjects would be

---

[6] **References:** <u>*Social preferences*</u>: Ultimatum game experiment: Güth, Schmittberger and Schwarze 1982, Google Citations (GS) 2,242; investment/trust game: Berg, Dickhaut and McCabe 1995, GS 2,132; models of social preferences relying on inferred social preference types from lab subjects: Bolton and Ockenfels 2000, GS 3,021, and Fehr and Schmidt 1999, GS 4,819. <u>*Risk attitudes*</u>: Holt and Laury 2002, GS 1,690; Machina 1982, GS 1,176 uses the lab results from risk studies to introduce new models of utility over risky gambles. <u>*Preference reversals*</u>: Grether and Plott 1979, GS 1,010, *Loss* <u>*aversion*</u>: Kahneman and Tversky 1979, GS 22,571; Tversky and Kahneman 1992 GS 4,520. <u>*Rationality*</u>: *(bounded) rationality* from the p-beauty contest game: Nagel 1995, GS 664 and the centipede game: McKelvey and Palfrey 1992, GS 551.

[7] One method to minimize this effect is to obtain measures of wealth w_0 among the lab subjects, estimate the (interaction) effect of wealth by each treatment, and then adjust for the distribution of wealth in the greater population. However, this approach relies on good measures of wealth in the subject population and appropriate weights in the population, and there appears to be little evidence that experimental studies are taking this approach with some notable exceptions (e.g., Andersen, Harrison, Lau and Rutström 2009; Harrison, Lau and Rutström 2009)

representative of the population they were recruited from, and thus to the extent that the 'representative' samples of subjects behave similarly in the lab, the populations they were recruited from would also behave similarly. However, as we demonstrate in this paper, these samples might not be representative of the populations from which they were recruited if subjects with certain characteristics are more likely to participate. Moreover, if the same participation biases exist in recruitment of both populations (for instance, towards lower income individuals and who volunteer more often), study participants will be more similar to each other than the populations they were recruited from, *ceteris paribus*. For instance, consider two populations H and L with mean incomes of $60,000 and $40,000, respectively, and identical variance in income σ^2. If people in both populations with income less than $30,000 choose to participate and everyone else chooses not to participate, the distribution of income among participants from both populations will be identical (and a higher percentage of the population in L than H will choose to participate).

The current paper emphasizes the importance of *experimental recruitment* procedures and how a rational (conscious) participation decision can lead to a non-representative sample. Our approach does not require recruitment from a general/representative population, rather we compare *participants* with the *target population* directly. We show that, for a population commonly used for lab experiments (university students), those who participate in a lab study are not representative of the target population due to the bias in the participation decision. Moreover, because a student population is more homogenous than a more general and representative sample, e.g., students have less variability in age, income, education, and possibly in political, moral and cultural values, there is already less opportunity for participation bias. To this extent, the substantial participation biases we present here very likely under-estimates participation bias that would occur with a more representative population.

2. Hypotheses

Voluntary participation is a core feature of almost all economic lab experiments. To advertise a lab experiment, researchers typically provide information on potential rewards (usually cash payments), participation time, and sometimes on the activities.[8] In this context, we assume individual i decides to attend ($a = 1$) or not attend ($a = 0$) a lab experiment k in order to maximize his expected utility,

(1) Individual i attends experiment k iff $E[U_i(a =1 \mid X_i, E_k)] \geq E[U_i(a =0 \mid X_i, E_k)]$.

X_i is a vector of individual characteristics and E_k is a vector of information on experiment k. We assume individuals are heterogeneous in X_i that includes income I_i, leisure time T_i, intellectual curiosity c_i, and social preferences s_i, but have the same beliefs regarding experiment k, E_k, which

---

[8] Offering monetary compensation not only increases the benefit to participation (and presumably participation rates), but is generally assumed to encourage participation among people who are motivated by the financial incentives that will be manipulated in the experimental design. Providing the participation time ensures that participants do not have other obligations that could result in them leaving before completing the experiment, and with information on money, allows participants to calculate per hour compensation. In environments where other researchers conduct experiments with the same population, indicating that the lab task is an economics experiment can signal a reputation that includes, among other features, no deception.

includes the task, potential earnings $r(a)$ and time $t(a)$.[9] The information E_k, serving several important practical purposes, may also attract a non-representative sample. Specifically, we assume individual i's expected marginal utility of participation is a function of the common information and heterogeneous characteristics:

(2) $U_i = U_i[W(I_i+r(a)), V(T_i-t(a)), M(c_i(a), s_i(a)) \mid E_k]$,

where W captures utility of wealth, a function of existing wealth I_i plus the expected lab earnings $r(a)$ if i participates; V captures the subjective utility of leisure time, a function of uncommitted time outside the lab T_i minus the lab time $t(a)$ if i participates; M captures the non-pecuniary benefits of participation that we separate into intellectual curiosity $c_i(a)$ and social preferences $s_i(a)$. If i does not participate in the experiment ($a = 0$), we assume $r(a) = 0$, $t(a) = 0$ and we normalize $M(c_i(0), s_i(0)) = 0$.

2.1 Main hypotheses

Using this model of utility and the common information given for a lab experiment, we first present our main hypotheses (H1-H4) that should be robust across most economic labs. We then consider three additional hypotheses (H5-H7) that rely on conditions that vary more widely across labs, and thus may be more idiosyncratic.

2.1.1 Wealth

We assume concave utility over wealth, $W' > 0$ and $W'' < 0$, thus higher wealth I_i will result in lower expected marginal utility from the identical belief in the lab payment $r(a)$. This implies, ceteris paribus, that the identical anticipated payment $r(a)$ will generate less expected marginal utility the greater an individual's wealth. Therefore, we anticipate non-random participation:

H1 (Wealth): Individuals with lower wealth, *ceteris paribus*, will be more likely to participate.

2.1.2 Leisure time

We also assume concave utility over leisure time, $V' > 0$ and $V'' < 0$, thus having more uncommitted (leisure) time T_i will result in a lower expected loss in marginal utility from the identical belief in the lab participation time $t(a)$. This implies, ceteris paribus, that the identical anticipated lab time $t(a)$ will generate a smaller loss in expected marginal utility the greater an individual's initial leisure time. Therefore, we anticipate non-random participation:

H2 (Time): Individuals with more leisure time, *ceteris paribus*, will be more likely to participate.

---

[9] We assume utility over time is only affected here by leisure time. We believe this is reasonable since most experiments give very short notice between the time of the invitation and the experiment, often one week or less, thus non-leisure time is more likely to be committed and not flexible (e.g., class and work scheduling). For experiments with longer planning time, the model can be made more flexible by allowing greater substitution of time in the lab with other activities, nonetheless, the hypotheses derived with respect to time would still hold, but we would expect the magnitude of the effect to be smaller the greater the flexibility a participant has with time commitments.

2.1.3 Intrinsic motivations

In addition to monetary benefits, lab experiments involve participating in tasks which individuals may anticipate deriving utility based on intrinsic interests $M(c_i(a), s_i(a))$. Thus, heterogeneous intrinsic interests may also lead to non-representative biases. Two intrinsic motivations that economists and psychologists have conjectured might affect participation are intellectual interests and pro-social preferences.

Intellectual curiosity - $c_i(a)$

A few older papers that examined participation in unpaid psychology and economics lab experiments noted that students who volunteered were more interested in the study. For instance, Rosenthal and Rosnow (1969, 1973) mentioned volunteers in social psychology experiments are 'scientific do-gooders that are interested in the research' (also discussed in Levitt and List, 2007b).[10] Running an unpaid economics experiment, Kagel, Battalio, and Walker (1979) found volunteers were more interested in the experiment than non-volunteers. People may thus attend experiments in part for the utility gained from the task, which could include interests in academic studies, intellectual activities or the specific lab tasks. Therefore, we anticipate non-random participation:

H3 (Intellectual Curiosity): Individuals with intellectual interests in the lab tasks, *ceteris paribus*, will be more likely to participate.

Additionally, in some economic labs participants are either explicitly informed or are implicitly aware that they are being invited to an *economics lab* experiment.[11] Thus, we anticipate that people who are more interested in economics or closely related areas of study, such as business, will have higher expected marginal utility from the lab activities. Therefore, we anticipate non-random participation:

H3a (Economics Interests): Individuals with interests in economic topics, *ceteris paribus*, will be more likely to participate in 'economics' experiments.

Pro-social preferences - $s_i(a)$

Levitt and List (2007a, b) conjecture that experimental subjects are more cooperative and pro-social based on evidence from previous psychology and economics experiments. In two early studies, Rosenthal and Rosnow (1969, 1973) noted that students who volunteered for social psychology experiments 'more readily cooperate with the experimenter and seek social approval' (see also Orne 1962; Rosen 1951). Pro-social preferences among participants may not be surprising for unpaid experiments (e.g., in psychology or health research) where participation often is framed around helping either the researchers or the greater community. These preferences could nonetheless also

---

[10] See also Dixon 1978, Jackson et al. 1989 and Jung 1969 on psychology experiment volunteers.
[11] Even if the advertisement does not explicitly indicate an economics experiment, potential subjects may infer an economics experiments either by the location of the experiment (e.g., in the economics building or where other economics experiments have been run) or if the population were initially signed up in a data base to be contacted for an economics experiment.

affect the participation decisions in incentivized economic lab experiments.[12] For instance, in a gift exchange experiment List (2006) observed that those who declined to participate in the lab experiment were less pro-social (less reciprocal) in a later field experiment than the lab participants.[13] Thus, participants who respond to researcher requests may obtain utility from participating in the same way as when they respond to other pro-social requests, e.g., a call for volunteers. In other words, students may be participating in an experiment to help out the researcher or general community. Therefore, we anticipate non-random participation:

H4 (Pro-social Preferences): Individuals with higher pro-social preferences, *ceteris paribus*, will be more likely to participate.

Two recent papers provide some initial evidence regarding this hypothesis. Falk et al. (2011) test whether students who donate money (when given the option while paying term fees) are more likely to participate in a later lab experiment, and Cleave et al. (2010) test whether students who reciprocate more (in a classroom trust game) are more likely to participate in a subsequent lab experiment. Neither study finds any relationship between the observed pro-social behavior and participating in the later lab experiment, suggesting no pro-social bias in participation. However, measuring pro-social behavior using monetary decisions may not be the most appropriate measure to understand whether pro-social preferences affect the participation decision. Wealthier people may be more likely to be donate money to charities given their larger budget, but less likely to participate as hypothesized in H1 above. Thus the correlation between participation and monetary donations may capture other countervailing factors than purely pro-social preferences.[14] More fundamentally, to participate in a lab experiment, participants sacrifice time in exchange for money, thus *they are deciding whether to give up time, not money*.[15] Camerer (2011) makes a similar point, "A likely reason why there are not systematic volunteering effects" (with respect to pro-social preferences) "is that subjects volunteer to earn money ..." Thus, the decision to participate in a lab study in which substantial monetary rewards are offered should motivate people to participate who are interested in earning money rather than donating money. On the other hand, to the extent that people perceive participation to assist university researchers, participation is similar to a decision to volunteer time. Our study adds to the literature on pro-social preferences and participation bias by recognizing that participation based on social preferences should reflect volunteering time:

---

[12] Extensive research examines the effects of monetary incentives on the supply of pro-social behavior. The results suggest that offering money can in some situations undermine (crowd out) intrinsic motivations, thus reducing the pro-social supply (e.g., see Bowles and Polinia-Reyes, 2011, for a review). To the extent that the monetary rewards crowd out the pro-social intrinsic benefits to participations, participation bias based on pro-social preferences will thus be less prevalent.

[13] List (2006) recruited by personally approaching potential subjects, which may contribute to screening more socially cooperative individuals into the lab. Selection effect was also not the main purpose of the study and the sample size for those who declined was small.

[14] In these studies, wealth was not directly controlled for. As we will show in the current work, controlling for wealth is critical for understanding the relationship between monetary donations and the participation decision.

[15] If volunteering time and donating money are perfectly correlated indicators of pro-social behavior, then there is no difference in measuring monetary donations or volunteering time. However, we are unaware of any study that has examined this correlation, and see no reason to assume *a priori* that they are perfectly correlated.

H4a (Volunteering Time): Individuals with higher pro-social preferences for volunteering time, *ceteris paribus*, will be more likely to participate.

2.2 Additional Hypotheses

The proceeding hypotheses (H1-H4) address features that are largely universal across lab studies, we now discuss three additional potential non-representative biases (H5-H7) arising from more idiosyncratic procedures which may vary substantially across labs.

2.2.1 Risk Attitudes

Assuming concave utility over wealth ($W' > 0$ and $W'' < 0$) and leisure time ($V' > 0$ and $V'' < 0$), higher uncertainty with respect to either the lab payment $r(a)$ or session time $t(a)$ will result in less participation among more risk averse individuals, ceteris paribus. Consistent with this hypothesis, Andersen et al. (2009) found that a decrease in the variance in the advertised lab payments led to greater risk aversion among lab participants.[16] Similar to the payment uncertainty, greater uncertainty regarding any other dimension of the experiment could bias the participation decision toward less risk-averse individuals on these dimensions.[17] For instance, there could be uncertainty with respect to whether someone will actually get to participate if he shows up, uncertainty regarding how long he might have to wait for a session to begin and uncertainty as to how much he will enjoy the tasks. Therefore, we anticipate non-random participation:

H5 (Risk Aversion): Individuals who are less risk averse with respect to either wealth, time or task activities, *ceteris paribus*, will be more likely to participate.

2.2.2 Patience

Because of the delay between individuals' decision to participate and the lab session time, the net benefit of the experiment will be discounted whereas additional costs will occur prior to the experiment, such as signing up (e.g., going online and filling out forms) and scheduling. Therefore, we anticipate non-random participation:

H6 (Patience): Individuals who are more patient, *ceteris paribus*, will be more likely to participate.

2.2.3 Recruitment conditions

In addition to individual characteristics, we also examine whether different recruitment procedures can increase participation rates (and consequently reduce participation bias). Economic lab experiments conventionally require students to make an appointment with a fixed starting time. The

---

[16] On the other hand, Cleave et al. (2010) find no overall evidence of more risk-averse individuals who participate than do not participate in a lab study. However, their disaggregated analysis finds that women who participated were less risk-averse than women who did not, which is consistent with the hypothesis, but they also find that men who participated were more risk-averse than men who did not while, which is inconsistent with our hypothesis. Thus, the existing evidence is in general consistent with our hypothesis with the exception of disaggregated behavior of men in Cleave et al. (2010).

[17] Roe et al. (2009) provides an example of risk attitudes affecting participation in research studies that involve greater risks. They find that more risk-averse subjects are less likely to participate in an fMRI study and a genetics study requiring a blood sample than the population they were recruited from.

fixed starting time solves the logistical need for having a fixed number of participants (e.g., for games and markets). However, compulsory appointments may affect participation because making an appointment incurs higher transaction costs than just showing up; for instance, subjects need to register in advance, thus adding a step to the recruitment process. Mandatory appointments can also reduce the likelihood of participation among people who receive negative utility from commitments.

An alternative recruitment method is to allow participants to show up (drop-in) at any time without making an appointment.[18] The ability to drop-in at any time should increase the likelihood of participation by offering greater flexibility and eliminating the costs associated with having to sign up in advance.

H7 (Recruitment conditions): A higher percentage of individuals will participate if they can drop-in rather than if they have to make an appointment.

One countervailing force when allowing potential participants to drop-in may occur among people who procrastinate since having to make an appointment can solve a time inconsistency problem (Strotz 1955). Further, if participants do not have an appointment they may have greater uncertainty regarding session availability and wait time, for instance whether the experimenters will be prepared and whether they will have to wait for others to show up. Thus, whether participation will increase depends on whether the added flexibility and lower transaction costs are greater than procrastination and uncertainty effects. While we anticipate that the hypothesized core participation biases (H1-H4) will be robust across recruitment conditions, other biases may differ. For instance, more risk-averse individuals may be more likely to participate when they have to make an appointment than when they must drop-in to the extent that fewer perceived uncertainties are associated with appointments.

To further explore these potential effects, we also examine a third recruitment condition in which individuals have the option to either make an appointment or drop-in. We include this condition to test the effect of greater flexibility and importantly to examine the characteristics of participants who choose to make an appointment rather than to drop-in. We anticipate that more risk-averse participants will be more likely to make an appointment to reduce uncertainties associated with dropping in. Therefore, we anticipate non-random participation:

H7a (Recruitment conditions): More risk averse individuals will be more likely to participate when they have to make a mandatory appointment than when they have to drop-in, and will be more likely to make an appointment when given an option.

3. Study Design

We first discuss the population (**S3.1**), then the measures we collected for each hypothesis (**S3.2**) using survey questions and experimental data. We conclude this section by discussing the precise procedures and time flow (**S3.3**)[19] and recruitment conditions (**S3.4**).

---

[18] Allowing drop-ins may not be feasible for many kinds of experiments that need a fixed number of participants, such as market experiments and games.

[19] As mentioned in the introduction, the population and procedures follow closely those used by Cleave et al. (2010).

3.1 Population

The classroom data collection was conducted in the first year undergraduate Introductory Microeconomics class at the University of Sydney. This class provides a large heterogeneous population similar to typical target populations for economic lab experiments. Introductory Microeconomics is the first unit towards an economics or business degree. The course is required for an economics or business degree, but many other students take this course as an elective. The course is predominantly taken during students' first term at the university and our intervention occurred during the fourth week of the term when the course material would not have involved any topic related to our data collection.[20]

Our population data collection occurred during students' normal tutorials (typically small classes of at most 18 students). At the beginning of each tutorial, students were asked to participate in a 20 minute survey and experiment. Our target population (hereafter 'the population') consists of the 892 students who participated in the *classroom tasks*. Participation in the classroom task itself is unlikely to suffer from any *voluntary* participation bias for two reasons. First, participation in the tutorials (not a class requirement) is uncorrelated with our classroom tasks because the tasks were not announced in advance; thus students would not have chosen to attend the tutorial in order to participate in the classroom tasks.[21] Second participation in the classroom tasks is extremely high; over 96 percent of the students who attended a tutorial participated in our tasks. After the classroom data collection, the population was invited to participate in a lab experiment (see Procedures **S3.3** below). We are interested in whether the characteristics of the students who subsequently attended the lab experiment represent the characteristics of this population.

3.2 Classroom tasks

The classroom tasks contain two parts: three incentivized experiments and a 20 item survey. To address our hypotheses, the survey contained questions to measure income, leisure time, intellectual and academic interests, pro-social preferences (time and money based pro-social activities), patience and risk attitudes, and controls for demographics. To complement the survey, we measured individual's pro-social, risk and inter-temporal preferences using standard lab tasks over monetary outcomes. The experimental tasks were conducted first and were followed by the survey. All tasks were conducted with multiple choice responses using paper and pencil. Students worked alone at their own pace throughout these tasks. A copy of the instructions, experimental decisions and survey questions are included in supplemental materials (available from the authors). We discuss our measures in depth given the key role they play for examining non-representative characteristics.

---

[20] Topics covered: introduction, scarcity, choice; supply and demand; elasticity; consumer and firm behavior.
[21] Approximately 70% of the students enrolled in the class attended a tutorial during the week of the classroom tasks and thus tutorial attendees may not perfectly represent the whole class.

3.2.1 Measuring wealth (H1)

We measure both household and disposable income with the questions "What is your family's annual household income approximately?" and "How much money do you spend in a typical week?"[22] While the household income measure is common in surveys and reflects overall resources available to the family, it may not correlate well with the resources available to the individual. In contrast, measuring spending per week, used by Andreoni and Sprenger (2010) as a proxy for weekly consumption, may better identify the resources immediately available to the individual and thus be a better proxy for disposable income. The two measures thus capture different aspects of wealth. Household income is a more direct measure of wealth for a student who does not yet have her own stable income and is not affected by students' spending habits. However, weekly spending more accurately reflects the disposable income currently accessible to students, either from household income or their own earnings. Students may focus more on disposable income if they do not fully follow lifetime consumption smoothing models.

3.2.2 Measuring leisure time (H2)

We measure available leisure time as the inverse of weekly working hours, "How many hours per week do you currently work for pay?" We chose work hours because it reflects fixed commitments students would have difficulty altering in the short time between the invitation and participation and since most students have identical class commitment time given the rigid undergraduate structure. One potential drawback with work hours is that it may be correlated with higher income which we hypothesized will reduce the likelihood of participation. However, in our analyses we control for this potential correlation with our income measures.

3.2.3 Measuring intellectual curiosity (H3, H3a)

We measure two broad aspects of intellectual curiosity, interest in economics and interests related to general experimental economics lab tasks, using survey questions and consistency of behavior in one of the experimental tasks. The survey question "What is your major area of study?" reflects individual's interest in economics. Among the five responses (Economics; Business; Arts or Social Sciences; Science or Engineering; Other), majoring in economics or business suggests either innate interest, enjoyment or ability with the subject matter which could carry over to an economics lab experiment.[23]

We also used three additional measures to capture general interests in experimental lab activities. The first measure is students' academic performance based on their university admission ranking,

---

[22] We stressed that weekly spending only includes short term consumption items with examples for both short and long term items: "(This should be your daily expenses e.g. food, travel, mobile charges, excluding e.g. rent, tuition)."
[23] Had we instead advertised a "psychology experiment," we would have hypothesized that psychology rather than economics students would have been more likely to participate in the experiment.

"Please indicate your ATAR[24] (Australian Tertiary Admission Rank)," and, "If you don't have an ATAR, what was your equivalent ATAR (reflecting class standing) when applying to university?" Higher academic performance suggests, all else equal, greater academic curiosity, and thus may predict more interest in lab experiments to the extent that the experiments may involve academic subject matter.

The last two measures to capture general interest in experimental economic lab activities examine how much attention individuals give to tasks that require reflection. We assume that an individual who is more interested in a task, ceteris paribus, will put more effort and reflection into the task. The first of these measures is the three item Cognitive Reflection Test (CRT; Frederick 2005). The CRT attempts to assess how reflective people are on thinking about simple logical problems. Each CRT question has a quick/impulsive but incorrect response and a correct solution that requires a little reflection (Frederick 2005).[25] We anticipate that higher scores on the CRT, ceteris paribus, reflect individuals' interest in thinking about intellectual activities common in experimental economics lab activities. The second measure examines "consistent" choices over two sets of three inter-temporal saving decisions (discussed below in Measuring time preferences section **S3.2.6**) in which we asked subjects to allocate money between an early and later period. Within each set we only varied the interest rate. We define a set of decisions as *consistent* if an individual saves the same amount or more when facing higher interest rates.[26] Similar to the CRT measure, but with monetary incentives, we assume that people who make more consistent decisions are likely to have given more effort to the decisions, ceteris paribus, and hence enjoy thinking about the types of decisions in economic lab experiments.

3.2.4 Measuring pro-social preferences (H4, H4a)

Given the large body of experimental research on pro-social preferences, we included four survey questions and one experimental decision to measure the population's pro-social behavior. The survey questions measured the frequency and absolute amount of money (donation) and time (volunteering) related to charitable behavior over the past year. To measure monetary donation frequency, we asked: "Excluding today, how many times have you donated money to a charitable organization, such as an

---

[24] Australian Tertiary Admission Rank (ATAR) is a rank that allows the comparison of students who have completed different combinations of HSC (High School Certificate) courses. See Australian Universities Admissions centre website for details: http://www.uac.edu.au/undergraduate/atar/

[25] The three questions are: (1) "In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days" The potentially impulsive answer '24 days' is incorrect. People who reflect on the fact that if the lily patch cover the entire lake on day 48, on day 47 it must be half the size, will give the correct answer 47 days. (2) 'A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents' (impulsive answer is often reported to be 10 cents, correct answer is 5 cents), and (3) 'If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ minutes' (impulsive answer is 100 minutes, the correct answer is 5 minutes). In the survey, we provided four options in a multiple choice format and always included the impulsive response as one of the options. For control, we also asked "Have you seen any of the last three questions (Decision 9-11) before?" Anyone who had seen any of the three questions were treated as a missing observation in the data analysis for this question.

[26] Although the wealth effect could cause individuals to save less when interest rates increase, we consider it to be negligible in this case, due to the small overall earnings ($100), and relatively large ($20) increments in saving amount choices.

international aid organization, child agency, church and so forth, in the past year?" and for the total dollar amount: "Approximately how much money have you donated to charitable organizations in the past year?" To measure time-based volunteering frequency, we asked: "How many times have you volunteered some of your time to a charitable organization, such as a non-profit, university charity effort, church and so forth, in the past year?" and for the total hours of volunteering: "Approximately how many hours have you donated to charitable organizations in the past year?" The frequency measures capture the number of distinct times individuals did pro-social activities whereas the total dollar amount captures the overall financial contribution, ceteris paribus. Measuring time volunteering is central to our analyses since (1) participating in an experiment involves sacrificing time rather than income, (2) money donations are likely positively correlated with income, and (3) the Levitt and List (2007a,b) conjecture is based on past evidence of volunteering time (e.g., to help researchers) rather than sacrificing money to attend the lab.

The experimental donation decision is a modified dictator game initially used in the lab by Eckel and Grossman (1996) that pairs subjects with a well-known charity. In our classroom experiment, students received $100 to allocate between themselves and the Australian Red Cross (ARC). For each dollar donated, the ARC received two dollars, thus the opportunity cost, including tax considerations, to donate money through the experiment is lower than if the subject kept the money and donated outside of the experiment. We paid the classroom students for their experimental decisions after the lab experiment was completed (discussed in 3.3 Procedures below). To ensure that all classroom students would incur the identical transaction cost on the later date no matter how much they gave to the ARC, including giving everything to the ARC and keeping nothing for themselves, we included an additional $10 payment for the students to receive on the later date. Students were given the following six options:

Table 3.1: The donation decision

Donation Choice	To keep for myself	To donate to the Australian Red Cross
1	$10 + $100	$0
2	$10 + $80	$40
3	$10 + $60	$80
4	$10 + $40	$120
5	$10 + $20	$160
6	$10 + $0	$200

3.2.5 Measuring risk preferences (H5)

We included two measures of risk preferences, one to capture a broad risk assessment and the other to capture financial risk since, as discussed in the hypotheses, the risks to participate may be partially financial (uncertainty over money earnings) and partially other risks such as time involved and the enjoyment of the lab tasks. The survey question, "In general, how much of a risk taker are you compared to other students?" was used to measure self-perception of risks broadly. The financially based risk measure is a modified version of the ordered lottery sequence (OLS) initially used in the

lab by Eckel and Grossman (2008) and earlier by Binswanger (1980, 1981) in a field experiment. We use the same OLS values used in Garbarino, Slonim and Sydnor (2011). In this task, students chose one of the six lotteries shown in Table 3.2. Each lottery has two outcomes with a 50 percent chance of occurring. Lottery 1 has a sure payoff of $22, Lotteries 2-5 offer increasingly higher expected value with increasingly greater risk (measured by either variance or CRRA preferences), and Lottery 6 has the same expected value to, but higher risk than Lottery 5. Less risk averse individuals will choose higher numbered lotteries and a risk seeking individual will prefer Lottery 6. We chose a single OLS task over other common measures due to its simplicity to explain and administer given our limited time.

Table 3.2: The Lottery decision

Lottery Choice	Information for Subjects				Additional Information		
	Outcome 1		Outcome 2			Standard	CRRA*
	Payoff	Prob	Payoff	Prob	EV	Deviation	Range
1	$22	50%	$22	50%	22.00	0.00	r > 2.74
2	$30	50%	$18	50%	24.00	6.00	0.91 < r < 2.74
3	$38	50%	$14	50%	26.00	12.00	0.55 < r < 0.91
4	$46	50%	$10	50%	28.00	18.00	0.37 < r < 0.55
5	$54	50%	$6	50%	30.00	24.00	0 < r < 0.37
6	$60	50%	$0	50%	30.00	30.00	Risk Seeking

*CRRA: $U(x) = [x^{(1-r)}]/(1-r)$ for r>0 & r≠1; $U(x)= LN(x)$ for r=1.

3.2.6 Measuring time preferences (H6)

We included two measures of patience, one to capture a broad patience assessment and the other to capture financial patience since future participation may be partially discounted based on the future money and partially on other aspects of the future benefits and costs. The survey question, "In general, how patient are you compared to other students?" was used to measure self-perception of patience broadly.

The financially based measure is a modified version of the 'convex time budget' inter-temporal task developed by Andreoni and Sprenger (2010).[27] We gave each student the six multiple choice questions shown in Table 3.3. For each question, students allocated $100 between an early and later payoff date. The early payoff date was either two weeks (decisions 1-3) or six weeks (decisions (4-6) from the date of the experiment while the later payoff date was always four weeks after the early payoff date. Each decision offered one of three levels of simple interest for the later payment for each dollar saved: 5, 10 or 20 percent. Following Andreoni and Sprenger (2010), in all decisions we included an additional $10 payment to both the early and later payoff date to ensure participants would incur the identical transactions costs on the early and later payment dates. We also had the

---

[27] The novel feature of this method is that subjects are free to choose interior points under a convex time budget constraint rather than having to allocate all payments on a single date. This method allows jointly estimating a discount rate and the curvature of the utility function within a set of inter-temporal choices without needing to estimate risk preferences separately (Andreoni and Sprenger 2010). However, due to classroom time constraints, we only asked a few of these inter-temporal choices, thus we will only use a simple 'average saving' amount (and control for risk using our risk task) to measure possible representative biases based on time preferences.

identical procedures for receiving future payments regardless of the payment date in order to have the same transaction cost on each date. A student who allocates more to the later payoff date thus saves more and, controlling for risk, is more patient. We used the six decisions to let us test whether the relationship between savings and participation varies with (1) different frontend delays (decisions 1-3 vs. 4-6) or (2) different interest rates. (However, we find that once we control for an individual's average saving rate across the six decisions, variation in either frontend delay or interest rates offers no further explanatory power on the participation decision and is not discussed further.)

Table 3.3: The saving decision
Information for Subjects

Dec	Choice	Front End Delay	Duration	Early Payoff	Late Payoff	Saving	4-week Rates (simple)
1	1	2 weeks	4 weeks	$10 + $100	$10 + $0	$0	5%
	2	2 weeks	4 weeks	$10 + $80	$10 + $21	$20	
	3	2 weeks	4 weeks	$10 + $60	$10 + $42	$40	
	4	2 weeks	4 weeks	$10 + $40	$10 + $63	$60	
	5	2 weeks	4 weeks	$10 + $20	$10 + $84	$80	
	6	2 weeks	4 weeks	$10 + $0	$10 + $105	$100	
2	1	2 weeks	4 weeks	$10 + $100	$10 + $0	$0	10%
	2	2 weeks	4 weeks	$10 + $80	$10 + $22	$20	
	3	2 weeks	4 weeks	$10 + $60	$10 + $44	$40	
	4	2 weeks	4 weeks	$10 + $40	$10 + $66	$60	
	5	2 weeks	4 weeks	$10 + $20	$10 + $88	$80	
	6	2 weeks	4 weeks	$10 + $0	$10 + $110	$100	
3	1	2 weeks	4 weeks	$10 + $100	$10 + $0	$0	20%
	2	2 weeks	4 weeks	$10 + $80	$10 + $24	$20	
	3	2 weeks	4 weeks	$10 + $60	$10 + $48	$40	
	4	2 weeks	4 weeks	$10 + $40	$10 + $72	$60	
	5	2 weeks	4 weeks	$10 + $20	$10 + $96	$80	
	6	2 weeks	4 weeks	$10 + $0	$10 + $120	$100	
4	1	6 weeks	4 weeks	$10 + $100	$10 + $0	$0	5%
	2	6 weeks	4 weeks	$10 + $80	$10 + $21	$20	
	3	6 weeks	4 weeks	$10 + $60	$10 + $42	$40	
	4	6 weeks	4 weeks	$10 + $40	$10 + $63	$60	
	5	6 weeks	4 weeks	$10 + $20	$10 + $84	$80	
	6	6 weeks	4 weeks	$10 + $0	$10 + $105	$100	
5	1	6 weeks	4 weeks	$10 + $100	$10 + $0	$0	10%
	2	6 weeks	4 weeks	$10 + $80	$10 + $22	$20	
	3	6 weeks	4 weeks	$10 + $60	$10 + $44	$40	
	4	6 weeks	4 weeks	$10 + $40	$10 + $66	$60	
	5	6 weeks	4 weeks	$10 + $20	$10 + $88	$80	
	6	6 weeks	4 weeks	$10 + $0	$10 + $110	$100	
6	1	6 weeks	4 weeks	$10 + $100	$10 + $0	$0	20%
	2	6 weeks	4 weeks	$10 + $80	$10 + $24	$20	
	3	6 weeks	4 weeks	$10 + $60	$10 + $48	$40	
	4	6 weeks	4 weeks	$10 + $40	$10 + $72	$60	
	5	6 weeks	4 weeks	$10 + $20	$10 + $96	$80	
	6	6 weeks	4 weeks	$10 + $0	$10 + $120	$100	

3.2.7 Demographics

We also collected each individual's gender and ethnicity on the survey. For ethnicity, we asked "How would you describe your ethnicity (please pick the most applicable)?" to identify four main ethnic groups in Australia: Caucasian, East Asian, South Asian and Middle Eastern.

3.3 Procedures

We collected the population data during tutorials. Tutorials have 18 or fewer students, most with less than 15 (mean 11.6), so tutors could easily ensure no communication between students. We prepared detailed tutor instructions (Appendix 1) explaining the precise procedures to follow and scripts to read. All tutors attended a training session and received a take home packet for practice. Tutors were explicitly instructed not to encourage or discourage students to attend the subsequent lab experiment.[28] The classroom exercise was not announced prior to the tutorial to avoid biasing tutorial attendance based on our intervention.

On the day of the classroom intervention, each tutor announced the tasks at the beginning of class and handed out the material to each student. Each student was asked to read the cover page and then decide whether to complete the tasks. If a student decided *not* to participate, he could go over tutorial material or do anything else but was asked to remain quiet while other students completed the tasks. If a student decided to participate, the student would immediately begin working through the tasks at his own pace. After completing the tasks, the tutor collected all the material from the students. The tutor then made an announcement about an upcoming lab experiment and distributed the randomly assigned flyer invitation (Appendix 1) to the class.

In total, 96% of the students attending the tutorial agreed to participate in the tasks. However, tutors were instructed to begin tutorial topics after a maximum of 20 minutes, and since some students came late or worked slowly, not everyone completed all the tasks. For our analyses, we include everyone who began the tasks, and treat any incomplete responses as missing observations in the analyses. Nonetheless, virtually all students (97%) completed the experimental (first) part, and almost 2/3 provided a response to every survey question.

The classroom tasks and subsequent lab experiment spanned two weeks (see Time Flow below). The classroom tasks were run during the entire Week 4 to cover all tutorials in the class. The opportunity to participate in the lab experiment began exactly one week after the classroom intervention and remained open for exactly five weekdays. To have the identical time between the initial invitation and the time students could participate we staggered the invitations based on the day of the tutorials; for example, students in the *Tuesday* tutorials received invitations stating the lab sessions would be available from the following *Tuesday* to the Monday a week later. Exactly one week after the classroom tasks (and thus on the first day that a student was eligible to attend a lab session) tutors reminded students of the lab experiment by distributing the identical flyer invitation in the tutorials.[29] Finally, in the last two days of the second week of the lab experiment we emailed a third round of the identical flyer invitations.

---

[28] To compensate tutors for the time involved in preparing our classroom tasks, we held a lottery after we completed the laboratory experiment; we drew 5 of the 22 tutors to receive $100 gift certificates to a local major shopping mall.
[29] In principle, there was nothing preventing students from attending different tutorials from one week to the next. However, students who change tutorial sessions will switch in the first two or three weeks of the term. Thus, one of our motivations for

Time Flow: Classroom Tasks, Invitation to Participate and Lab Experiment
Day 0:   In class: intervention then lab experiment flyer invitation
Day 7:   In class: reminder lab experiment flyer invitation
Day 7 – 13:   Lab experiment open for participation
Day 11:   E-mail reminder sent with lab experiment flyer invitation
Day 13:   Last day for lab experiment
Day 14:   Students receive e-mail indicating whether chosen for pay for the class experiment

To pay classroom participants for the experimental tasks, we randomly chose 40 students who participated in the classroom tasks and paid each one for one of their decisions, also randomly chosen.[30] We informed students that they would learn whether they got paid two weeks after their classroom participation (coinciding with the first possible payment date for the saving decision). Importantly, since the classroom task only compensated a random sample of the classroom participants, we explained in the invitation flyer to the lab experiment that unlike the classroom exercise, everyone who participates in economics lab experiments gets paid.

3.4 Recruitment conditions

We varied the flyer invitation (Appendix 1) to examine three recruitment conditions: 'Appointment,' 'Drop-in' and 'Option.' The appointment condition required students to make an appointment using an online scheduling option on the course webpage.[31] The drop-in condition asked students to walk in anytime during lab hours and no appointment was mentioned. Students in the option condition were given the choice of making an appointment or walking in. The invitation provided a reason for the option by explaining, 'appointments are helpful since spaces may be limited.' Since the drop-in condition gave students the most flexibility, starting any time between 10:00 AM and 4:00 PM each day, in the appointment condition we allowed students to choose times to arrive in 20 minute intervals, thus on each day a student could sign up for an appointment for 19 different times, or 95 different times over the course of the week. Besides these differences, the three flyers were identical.

The recruitment conditions were balanced across both the tutorial day of the week and tutors. The class had 77 one-hour tutorials each week run by 22 tutors. Within each tutorial, the same flyer was handed out to all students in the tutorial to reduce the chance that students would have been aware of the different flyer conditions. Across tutorials, and since all tutors taught either three or six tutorials (except one who taught two tutorials), each tutor was assigned to hand out a different flyer in each tutorial if they taught three tutorials, or each flyer in two tutorials if they taught six tutorials. The tutors were given separate sealed envelopes with the flyers in them for each of their tutorials and were instructed to not open the envelopes until handing out the flyers. Thus, the tutors would not have been

---

running the classroom tasks in Week 4 was that we were unlikely to have students switch tutorials from Week 4 to in Week 5, thus students would always receive the same flyer invitation.

[30] We considered paying all students, however with the large sample the stakes would have to have been quite small and thus would have risked being salient sums to the students. We thus opted to pay a random sample so the decisions could be over much larger sums for the savings and donation decisions. We used the same scale for the risk decision as Garbarino, et al. (2011).

[31] Signing up for an appointment took about two minutes. To avoid any differences in the information participants had to provide across the conditions, we did not collect any information beyond their name when they made an appointment.

aware of the invitation condition until after the class tasks had been completed. The tutors were never told the purpose of our study or that there were different recruitment conditions, thus the study was conducted double anonymous.

4. Results

4.1 Non-representative biases

We examine whether the characteristics of lab participants are representative of the characteristics of the population from which they were recruited. We first examine the effect of the recruitment conditions on participation, followed by individual tests for the core hypothesized non-representative biases over income, leisure time, intellectual curiosity and social preferences, and then the more idiosyncratic hypotheses over risk and time preferences. We then test the robustness of the individual results estimating all biases simultaneously, and conclude by testing whether the biases differ across recruitment conditions.

4.1.1 Recruitment conditions

Figure 4.1 shows lab attendance by the three recruitment conditions: appointment, drop-in and option. The bars and left-hand y-axis indicate the participation rate for each condition; the line and the right-hand y-axis indicate the number of respondents for each condition (subsequent figures have the same format). Figure 4.1 shows that of the 306 students in the appointment condition, 26 percent participated in the lab experiment, while 23 percent of the 298 students in the drop-in condition participated and 21 percent of the 281 in the option condition participated. On average, 23 percent of the population attended the lab experiment.

Figure 4.1: Participation by recruitment condition
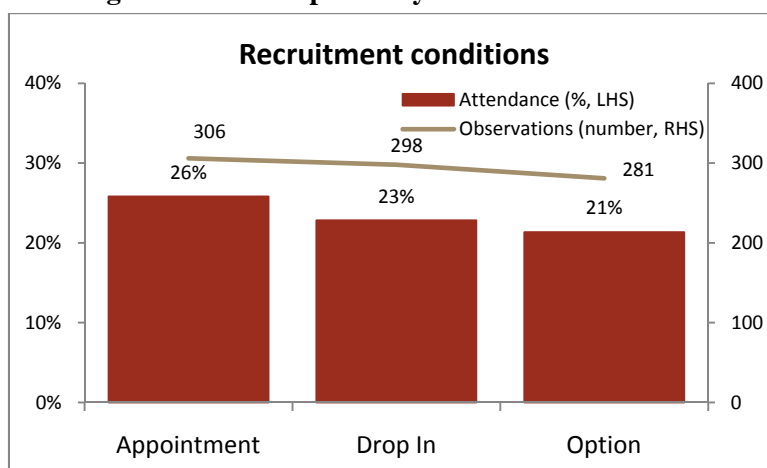
We estimate the following probit model to test whether the differences in attendance across conditions are statistically significant:

(3) $y_i = f(a + \beta_1 drop\text{-}in_i + \beta_2 option_i + \sum \delta_j X_{ji} + \varepsilon_{ti})$

where y_i equals 1 if student i in the population participated in the lab experiment, and 0 otherwise; f is the normal Probit function, *drop-in$_i$* and *option$_i$* are dummy variables indicating student i's recruitment condition. X_{ji} is a vector of dummy variables for the day of the week students attended their tutorials[32] and the 22 tutors who ran the classroom tasks. Although tutors were given identical training and instructed to follow the identical procedures, they may have inadvertently influenced students' participation decisions differently across their tutorials. Thus, in addition to controlling for tutor effects, we also estimate and report standard errors clustered at tutorial level to address possible differences across the 77 tutorials.

Table 4.1 presents the results. Column 1 shows the estimates without the controls for weekday and tutor (X_{ji}) and without clustering for the 77 tutorials, Column 2 includes the controls for weekday and tutor (X_{ji}) as well as clustering the errors at the tutorial level, and Column 3 reports the marginal effect based on the Column 2 estimates. In both regressions, the difference among recruitment conditions did not reach a conventional level of significance ($p > 0.10$), although directionally subjects in the drop-in and option conditions were 3.2 and 4.2 percentage points less likely to participate in the lab experiment, respectively, compared to those in the appointment condition.[33]

Result 0: *In contrast to hypothesis, H7, lab participation was not significantly different across the three recruitment conditions.*

Table 4.1: Participation by recruitment condition

	(1)	(2) †	Marginal Effect†
Drop-in	-0.0958	-0.109	-0.0322
	(0.112)	(0.127)	(0.0371)
Option	-0.145	-0.144	-0.0422
	(0.114)	(0.124)	(0.0355)
Constant	-0.649***	-0.942	
	(0.0774)	(0.579)	
Tutorial Effect	N	Y	Y
Observations	885[34]	885	885
Log Likelihood	-480.5	-463.9	-463.9

Probit regressions with robust standard errors in parentheses
†Regressions include dummy variables for tutors and tutorial days;
*** p<0.01, ** p<0.05, * p<0.1

---

[32] Students from Friday tutorials were more likely to attend the lab experiment ($p <0.01$) relative to every other day of the week, otherwise we found no differences across the days of the week. One concern with this Friday effect is that it might indicate that students on Friday, the last day of the classroom tasks, may have been more likely to have heard about the in class tasks, and thus Friday classroom attendees might reflect participation bias. However, we find no statistical difference in the percent of students attending Friday tutorials than any other day, thus attending Friday tutorials is unlikely to reflect participation bias in response the classroom tasks. A more likely explanation for the Friday effect is that students have fewer classes on Friday, thus the students who attend Friday classes have more spare time to participate in the lab experiment on the day they attend their tutorial.

[33] We further find that the combined Drop-in plus Option conditions was also not jointly significantly different than the Appointment condition (p<0.10).

[34] We excluded seven classroom task participants in all our analyses because we were unable to match their student IDs from the self-reported classroom data with either university or course administrative records, and we had two lab participants who did not provide us with valid IDs, thus we could not determine the participation status of these seven students in the subsequent lab experiment.

Discussion: While we anticipated that the greater flexibility would lead to greater participation when the population could drop-in or could choose to drop-in or make an appointment, we find no evidence in this direction. Instead, the highest participation rate occurred directionally when an appointment was required. Thus, we find no evidence that two alternative recruitment methods examined here can improve the participation rate over the standard appointment system commonly used in experimental economics labs.

We now turn to our main hypotheses. In the subsequent analyses, we always control for recruitment conditions, fixed weekday and tutor effects (X_{ji}) and cluster standard errors at the tutorial level. For each subsection, we estimate variations of the following model:

(4) $\quad y_i = f(a + \sum \theta_j Z_{ji} + \beta_1 drop\text{-}in_i + \beta_2 option_i + \sum \delta_j X_{ji} + \varepsilon_{ti})$,

where Z_{ji} is a vector of the core characteristics j of interest (e.g., income and time) for individual i.

4.1.2 Wealth (H1)

Spending per week: Figure 4.2 and Table 4.2 Columns 1 and 4 examine lab participation based on students' weekly spending. Figure 4.2 shows a clear negative correlation of about a 10 percentage point decrease in participation for every $40 increase in weekly spending. The regressions in Table 4.2 Columns 1 and 4 show that students who spend more per week were less likely to attend the lab. The variable 'weekly spending' is highly significant ($p<0.001$) either on its own (Column 1) or with the household income and work hours variables in the regressions (Column 4).

It is worth discussing immediately the relationship between the magnitude of the participation bias and its (disproportionately large) effect on the representativeness of the lab participants relative to the population. To see this, suppose for simplicity there are an identical number of people in the population within each weekly spending category. The raw participation rates shown in Figure 4.2 suggest that lab participants, rather than being equally divided as assumed for the population, would instead include a ratio of 34 to 13 participants in the lowest and highest income categories, thus around 72% (34/47) of the lab participants would be members of the population from the lowest income group relative to the highest income group despite the population consisting of an equal number of people in these groups. We discuss the magnitudes of all the estimates after presenting the remainder of the results, but note here that even small differences in participation rates (e.g., 10 percentage points per $40 spending levels) can lead to large differences in the non-representativeness (e.g., about **2.6 to 1**) in the lab.
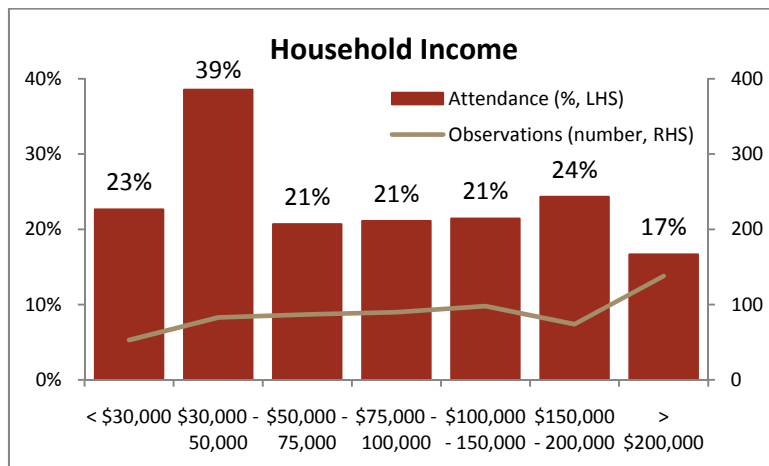
Figure 4.2: Participation by weekly spending
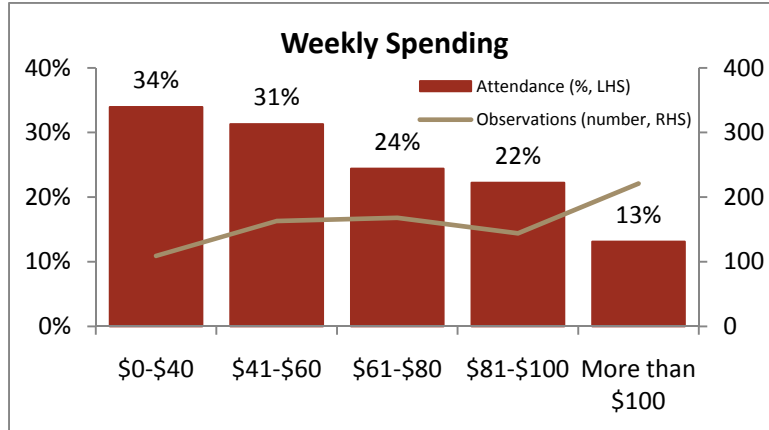
Household income: Figure 4.3 and Table 4.2 Columns 2 and 4 examine lab participation based on students' household income. Figure 4.3 places the population into seven categories from the lowest income (less than $30,000 per year) to the highest (over $200,000 per year). We observe similar participation rates of around 20-24 percent across the middle income range (from $50,000 to $200,000), with higher attendance, around 32 percent, averaging across the two lowest income groups (< $50,000) and lower participation, around 17 percent, for the top income group (> $200,000). The regression with household income alone, not controlling for spending per week, indicates that students with a higher household income were significantly less likely to attend the lab experiment (Column 2, $p<.05$). However, Column 4 shows that the effect of household income is insignificant when controlling for weekly spending and work hours.[35]

Result 1: *Students with less weekly spending or lower family income were more likely to attend the lab experiment compared to the population from which they were recruited.*

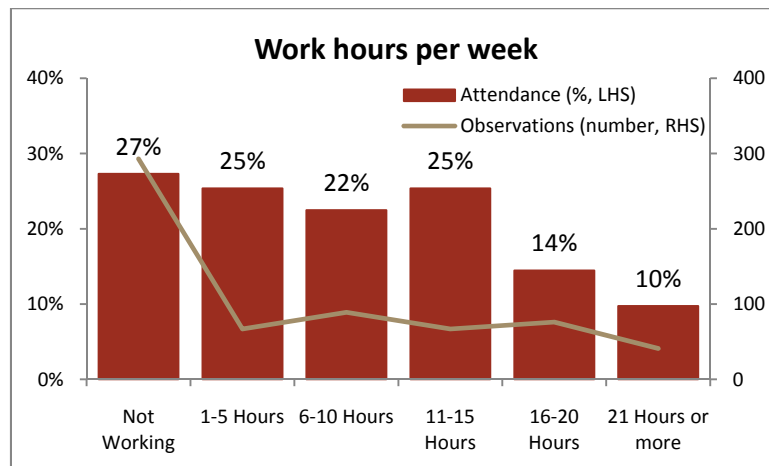**Figure 4.3: Participation by household income**

---

[35] The correlation between household income and weekly spending is only 0.126.

4.1.3 Leisure time (H2)

Work hours: Figure 4.4 and Table 4.2 Columns 3 and 4 examine lab participation based on hours worked per week. The work hours range from 0 (not working) to working 21 hours or more per week in 5-hour increments.[36] Figure 4.4 shows a downward trend in participation rates as hours worked increases. We observe higher participation rates for students who work 0-15 hours per week (22-27 percent) than for students who work more than 15 hours per week (10-14 percent). Regressions confirm that students who work more hours were less likely to come to the lab experiment. The variable work hours is significant both alone (Column 3, p<.01) and when controlling for the income variables (Column 4, p<.05).

Result 2: *Students who work fewer hours were more likely to attend the lab experiment compared to the population from which they were recruited.*

Figure 4.4: Participation by work hours

Table 4.2: Participation by income and work hours

	(1)	(2)	(3)	(4)	Marginal Effect
Weekly Spending	-0.171***			-0.157***	-0.0460***
	(0.0340)			(0.0352)	(0.0102)
Household Income		-0.0645**		-0.0448	-0.0131
		(0.0312)		(0.0315)	(0.00922)
Working Hours			-0.0991***	-0.0799**	-0.0234**
			(0.0344)	(0.0348)	(0.0102)
Constant	-0.407	-0.646	-0.632	0.0165	
	(0.561)	(0.596)	(0.585)	(0.586)	
Observations	885	885	885	885	885
Log Likelihood	-452.4	-461.5	-459.1	-447.0	-447.0

Probit regressions with robust standard errors at tutorial level in parentheses
*** p<0.01, ** p<0.05, * p<0.1;
Regressions include dummy variables for tutors, tutorial days and recruitment conditions

Discussion: Results 1 and 2 indicate that the participation decision is consistent with a rational response to monetary recruitment incentives and time requirements given individual's existing income

---

[36] The top three categories are 21-25, 26-30, and 31 hours or more are collapsed into the category 21 hours or more in Figure 4.4, due to small number of observations (less than 15 per cell). The regressions include all categories.

and time commitments. We interpret the negative correlation between income and participation as the result of lower marginal utility for the experimental payments among individuals with greater wealth (H1). When both weekly spending and household income are included in the model, the estimates suggest students are influenced more by disposable income than family wealth when deciding whether to participate. We included the variable 'work hours' to proxy for the opportunity cost of the time to participate (controlling for possible correlation with income); thus the negative correlation between work hours and participation confirms significantly lower participation among students with higher value for their leisure time (H2).

4.1.4 Intellectual curiosity (H3, H3a)

University majors: Figure 4.5 and Table 4.3 Columns 1 and 5 investigate the effect of students' major areas of study on participation. Figure 4.5 shows participation based on five areas of study: Economics, Business, Arts and social sciences, Science and engineering and Other majors. Approximately 26 percent of economics and business majors participated in the lab experiment compared to 20 percent among science students and 15 percent among arts and other majors. Regressions in Table 4.3 confirm that economics and business majors were significantly (p<.01) more likely to participate than the other majors when estimated alone (Column 1) or with the other intellectual curiosity measures (Column 5).

Saving consistency: Figure 4.5 and Table 4.3 Columns 2 and 5 investigate the consistency of the saving decisions on lab participation. Figure 4.5 categorizes students into three groups: 'consistent' in both savings decision sets, consistent in one set, or 'inconsistent' in both. Figure 4.5 shows that 26 percent of the population participated among those consistent in both decision sets whereas only 18 percent of those inconsistent in both decision sets participated. Coding the variable "consistent" equal to 0 if the student was inconsistent in both sets, equal to 1 if consistent in one of the sets, and equal to 2 if consistent for both sets, regressions in Columns 2 and 5 confirm that more consistent individuals were significantly more likely to participate in the lab experiment.

**Figure 4.5: Participation by major and consistency**
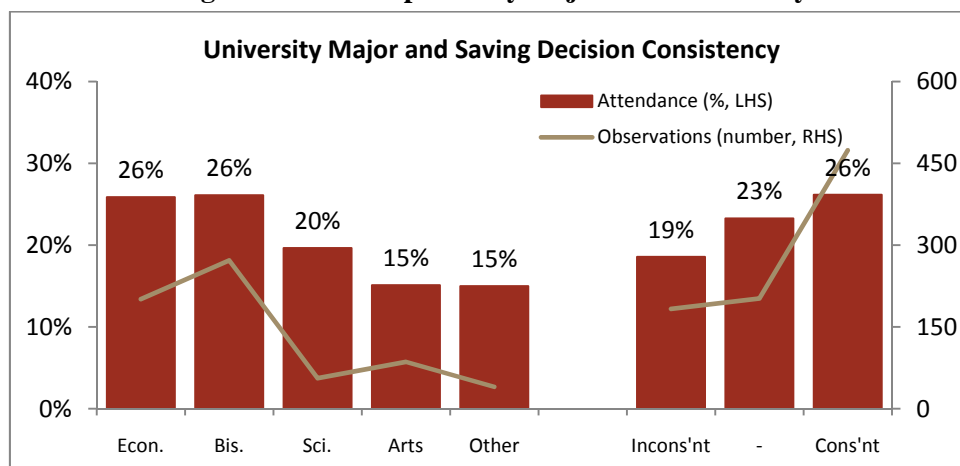
CRT score: Figure 4.6 and Table 4.3 Columns 3 and 5 investigate whether lab attendance differs based on students' Cognitive Reflection Test (CRT) score. Figure 4.6 shows that students who answered two or more CRT questions correctly are directionally more likely to participate than those who answered less than two correctly. Although the estimated CRT effect on participation does not reach a conventional level of significance, it is marginally significant in a one-tailed test in the direction hypothesized ($p < 0.10$) and in the full model (Table 4.7) controlling for other variables the CRT variable becomes significant ($p < 0.05$).

ATAR score: Figure 4.6 and Table 4.3 Columns 4 and 5 investigate whether lab attendance differs based on students' university admission ranking ATAR. Figure 4.6 shows no strong pattern of participation across the ATAR scores.[37] Estimates in Table 4.3 indicate academic performance measured by the ATAR score did not significantly affect lab participation when estimated alone (Column 4) or with the other intellectual curiosity measures (Column 5).

Result 3: *Students who major in economics and business, made more consistent decisions and had a higher CRT score were more likely to attend a laboratory economics experiment. However, relative academic performance upon entering the university does not predict participation.*

Figure 4.6: Participation by CRT and academic performance

---
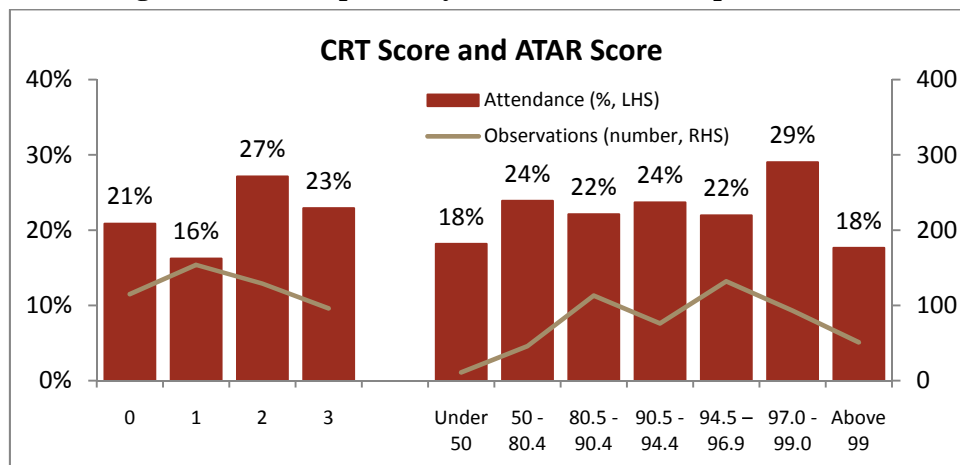
[37] Under 50 refers to below average in academic performance and likewise above 99 refers to above the top one percentile in academic performance.

25

Table 4.3: Participation by intellectual curiosity

	(1)	(2)	(3)	(4)	(5)	Marginal Effect
Economics/Business	0.352***				0.387***	0.112***
	(0.118)				(0.123)	(0.0345)
Consistent Saving		0.138**			0.136**	0.0398**
		(0.0536)			(0.0556)	(0.0167)
CRT Score			0.102^		0.100^	0.0295^
			(0.0658)		(0.0701)	(0.0204)
ATAR Score				0.0153	0.00331	0.000973
				(0.0444)	(0.0471)	(0.0138)
Constant	-1.239**	-1.069*	-1.102*	-1.011*	-1.581***	
	(0.612)	(0.564)	(0.577)	(0.596)	(0.601)	
Obs.	885	885	885	885	885	885
Log Likelihood	-460.1	-458.6	-461.9	-463.8	-452.0	-452.0

Probit regressions with robust standard errors at tutorial level in parentheses
*** p<0.01, ** p<0.05, * p<0.1; ^ p<0.1 (one tailed test);
Regressions include dummy variables for tutors, tutorial days & recruitment conditions

Discussion: Result 3 indicates a non-representative bias based on interests in economics or business, consistent with *Hypothesis 3a* that people were more likely to participate in experimental studies related to their major field of interest. In our case, the 'economics decision-making' task advertised in the recruitment invitations likely attracted more economics and business students than had we advertised another discipline-specific type of experiment, or no specific field.[38] The results on broader intellectual curiosity generally support our hypothesis (H3). Although we find that academic performance measured by ATAR scores does not predict lab attendance, evidence from the more direct measures of decision consistency and accuracy on the CRT suggests that students with higher interests or ability in intellectually challenging activities are over-represented among the lab attendees. Additionally, savings consistency may also imply lab participants are more responsive to incentives (in this case, the interest rate variations). Since participation is a decision with monetary incentives (experiment pay), the correlation between participation and saving consistency may indicate a non-representative bias based jointly on responsiveness to monetary incentives and interest in the experimental tasks.

4.1.3 Pro-social preferences (H4, H4a)

Volunteering (frequency and total hours): Figure 4.7 and Table 4.4 Columns 1, 2, 6 and 7 examine participation based on the population's volunteering behavior. Figure 4.7 presents the participation rates by the number of times students volunteered in the past year in the left panel and by the total

---

[38] Psychology studies show that the description and title of an experiment could impact on subjects' self-selection into the experiment (Jackson et al. 1989, Senn and Desmarais 2001, Saunders et al. 1985, Silverman and Margulis 1973). An alternative explanation for the stronger participation effect among economics and business majors could be that the lab's location was in the economics and business building, thus potentially making the location more convenient for economics and business students. However, since virtually all (73 of the 77) tutorials were held in the same location in the economics and business building, the students in the population would be in the lab building at least twice during the time of the experiment (lecture and tutorial), thus the location is not likely to have played a major role.

hours volunteered in the right panel. Students who volunteered more than 10 times in the past year were 10 to 12 percentage points more likely to participate in the lab experiment than students who volunteered fewer than 10 times. Regressions show that students who volunteered more frequently were significantly more likely to participate when estimated alone (Column 1) or with controls for other charitable behaviors (Column 6), consistent with H4a. Students volunteering more hours were also directionally more likely to attend the lab; for instance, 27 percent of students who volunteered six or more hours participated in the lab experiment whereas only 22 percent of students who volunteered less than 6 hours participated. However, this does not reach significance alone (Column 2, $t=1.26$), or controlling for other pro-social behaviors (Column 6).

Figure 4.7: Participation by volunteering
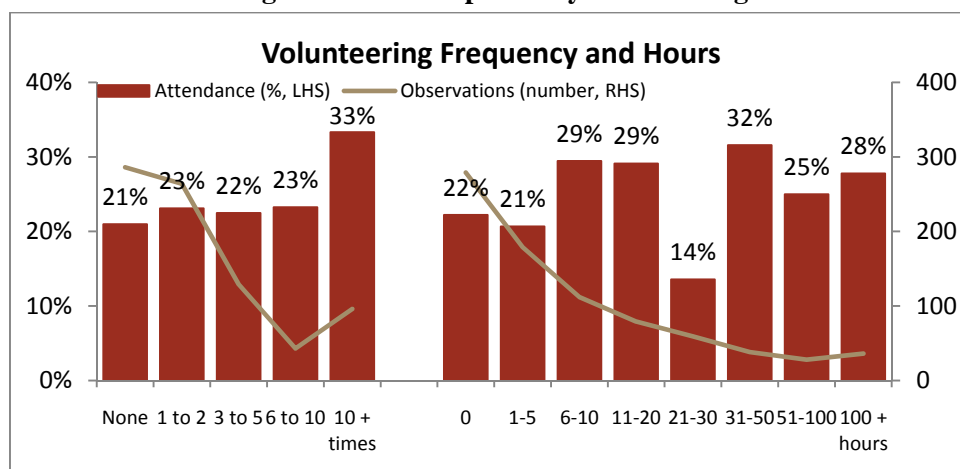
Donation (frequency and total dollars): Figure 4.8 and Table 4.4 Columns 3, 4, 6 and 7 examine participation based on the monetary donation frequency and total dollars donated. The left and right panels of Figure 4.8 present participation by the frequency of monetary donations in the past year and by the total dollars donated, respectively. We observe no strong pattern with the number of donations or with total dollars donated, but Figure 4.8 suggests directionally that students who donated money more frequently or gave fewer total dollars were more likely to participate. For instance, 19% of students who never donated participated whereas 24 percent and 26 percent of students who donated 1 to 5 times and more than 5 times, respectively, participated. On the other hand, 25 percent of students who donated less than $100 participated, while only 17 percent of students donating more than $100 participated.

Regressions with either monetary donation frequency or total dollars donated alone (Columns 3 and 4) indicate that neither is significant. However, Column 6 shows that when controlling for the other pro-social measures, donation frequency becomes marginally significant indicating that students who donate money more frequently are more likely to participate (consistent with H4) while students who donate *less* money are significantly more likely to participate. While the donation frequency may reflect how often students think about acting pro-socially, consistent conceptually with the number of times they volunteer, donation dollars as discussed earlier may be positively correlated with wealth

which we hypothesized and showed negatively affects lab attendance (Result 1). Adding in controls for the wealth measures, Column 7 shows indeed that donation dollars is no longer significant. Finally, the full model (Table 4.7) shows that when controlling for all of the measures, neither the frequency nor the total amount of monetary donations have a significant effect on participation.

Figure 4.8: Participation by monetary donations
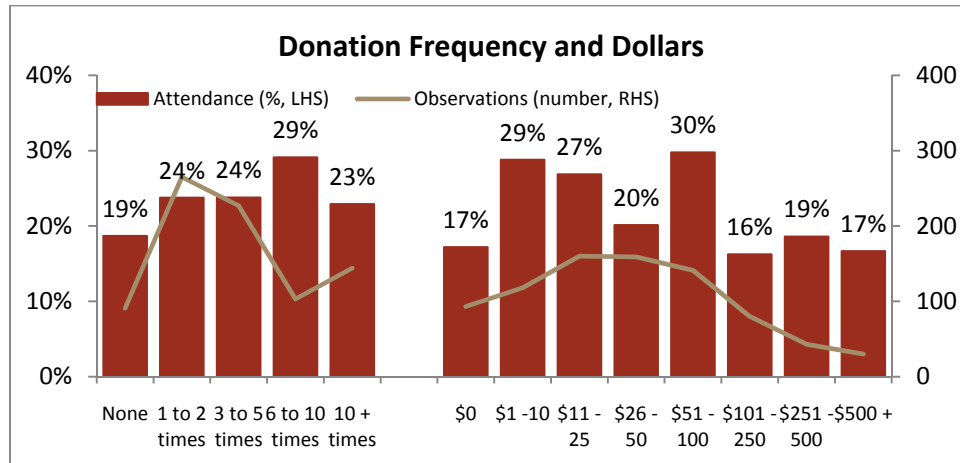
Dictator game decision: Figure 4.9 shows participation rates based on the dictator game decision between each student and the Australian Red Cross (ARC). Participation was highest, over 28 percent, for those who donated $160 or more, around 23 percent for those who donated between $40 and $120 and 17 percent for those who donated $0. Regressions in Table 4.4 show that students who gave more money were significantly more likely to participate when estimated alone (Column 5) or with controls for the other pro-social measures (Column 6) and controls for wealth (Column 8, p<.06). Unlike the total monetary donation (outside the classroom experiment) effect that depends on controlling for wealth, the dictator game donation effect is unaffected when we controlled for wealth, consistent with Andersen et al.'s (2009) finding that subjects only partially integrate wealth when making lab decisions.

Result 4: *Lab participants were not representative of the target population based on pro-social preferences. The lab participants more frequently volunteered their time, but not money, and also donated more in the experimental dictator game than the population from which they were recruited.*

Figure 4.9: Participation by dictator game decision
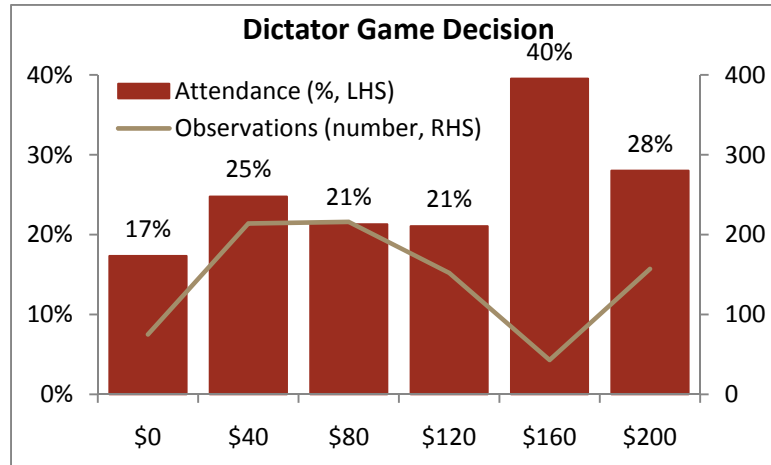
Table 4.4 Participation by pro-social behaviors

	(1)	(2)	(3)	(4)	(5)	(6)	Marginal Effect	(7)	(8)
Volunteering Frequency	0.095** (0.037)					0.159** (0.076)	0.045** (0.0215)	0.148* (0.077)	
Volunteering Hours		0.029 (0.023)				-0.037 (0.046)	-0.011 (0.013)	-0.037 (0.047)	
Donation Frequency			0.030 (0.040)			0.053 (0.045)	0.015 (0.013)	0.068 (0.049)	
Donation Dollars				-0.022 (0.028)		-0.074** (0.033)	-0.021** (0.0093)	-0.057 (0.035)	
Dictator Game (x100)					0.143** (0.067)	0.146** (0.072)	0.0414** (0.0203)		0.142* (0.075)
Weekly Spending								-0.145*** (0.036)	-0.152*** (0.036)
Working Hours								-0.087** (0.036)	-0.079** (0.0356)
Household Income								-0.044 (0.032)	-0.052 (0.032)
Constant	-1.115* (0.574)	-1.000* (0.579)	-1.079* (0.599)	-0.893 (0.585)	-1.076* (0.580)	-1.307** (0.608)		-0.328 (0.607)	-0.0979 (0.579)
Obs.	885	885	885	885	885	885	885	885	885
LL	-460.7	-462.6	-462.9	-463.5	-458.9	-449.3	-449.3	-438.1	-442.5

Probit regressions with robust standard errors at tutorial level in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$; ^ $p<0.1$ (one tailed test); Regressions include dummy variables for tutors, tutorial days and recruitment conditions

Discussion: We find a significant positive correlation between participation and time-based pro-social behavior but no significant relationship between participation and monetary-based pro-social behavior outside the lab. Money- and time-based pro-social behaviors are not perfectly correlated and in fact can be fairly distinct. For example, the highest correlation between the two volunteering and two donation behavior measures is only 0.37 between volunteer hours and donation dollar amount. We anticipated that time-based pro-social preferences would be critical for the *participation decision* since participation does not explicitly involve donating money but instead involves earning money in exchange for giving up time. Therefore people who are more willing to volunteer their time would be

more likely to participate in the lab experiment while people who donate more money are not equally generous with their time. This distinction may explain the insignificant results trying to link money-based charitable behavior with lab participation in previous studies (Cleave et al. 2010; Falk et al. 2011) [39].

The result from the dictator game donation decision indicates that lab participants over-represent individuals who are more generous with money received in an experiment. The contrast between the positive effect between the dictator game and participation, and the insignificant (and directionally negative) donation dollars and participation is interesting. One possibility is that the dictator game may reflect the charitable feeling of the population at the closest time to the participation decision, and is thus capturing the population's temporal pro-social preferences and therefore manifest themselves on the participation decision. Further, by measuring the dictator donation decision over the same options and same charitable entity (the Australian Red Cross) at the same time, it might provide a more parsimonious measure of charitable behavior than the other two donation measures. Further, since the dictator game decision was over "house money," the population may not treat the decision in the same manner as donating money they have earned outside the lab; the correlation between the dictator game decision and the two monetary donation dollars was just 0.12. For instance, Andersen et al. (2009) find that lab subjects only partially integrate outside wealth with experimental decisions. Thus, it is possible that a combination of the dictator game's temporal proximity to the participation decision, parsimony across the charitable decision and the abstraction from outside wealth may result in a measure that captures pro-social behavior related to the pro-social aspect of the participation decision.[40]

4.1.5 Time and Risk preferences (H5, H6)

Risk perception: Figure 4.10 shows participation rates by risk perception in the right panel and by each of the six lottery choices in the left panel. Both panels suggest directionally that *more* risk-averse individuals are more likely to participate. For instance, excluding the small sample of the least and most risk taking individuals, 32, 24 and 21 percent of the below average, average and above average risk takers, respectively, participated. Likewise, 19 percent of the population that chose the two riskiest lotteries participated, whereas 24 percent of that chose the four least risky lotteries participated. Regressions indicate that both risk attitudes and lottery choice are directionally negative (Table 4.5), with risk perceptions marginally significant. However, controlling for all the variables in

---

[39] Several studies have also compared students with non-students and found students are no more pro-social than non-students. (Fehr and List (2004), List (2004), Cardenas (2005), Carpenter et al. (2004), Bellemare et al. (2008), Carpenter et al. (2008), Burks, Carpenter, and Goette (2009), Baran, et al. (2010), Cardenas (2011), Falk, et al. (2011). However, all of these studies reported monetary-based pro-social preferences. It would be interesting to see whether those results change under time-based pro-social preference measures.

[40] In both Cleave et al. (2010) the pro-social monetary decision and the participation decision were several months apart. In Falk et al. (2011) the pro-social decision involved sacrificing their own money.

the full model (Table 4.7), neither risk perception nor lottery choice reach even marginal significance (p>.20).

Figure 4.10 Participation by risk perception and lottery choice
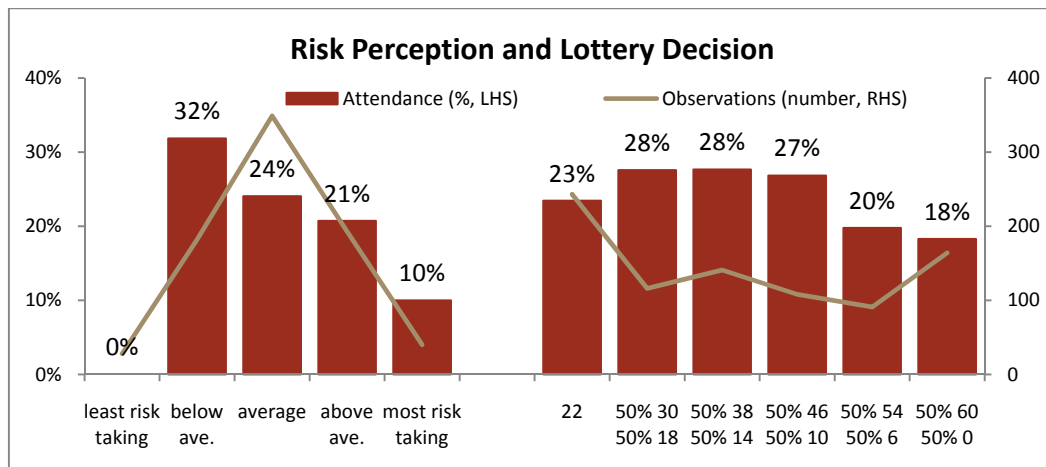
Table 4.5 Participation by risk perception and lottery choice

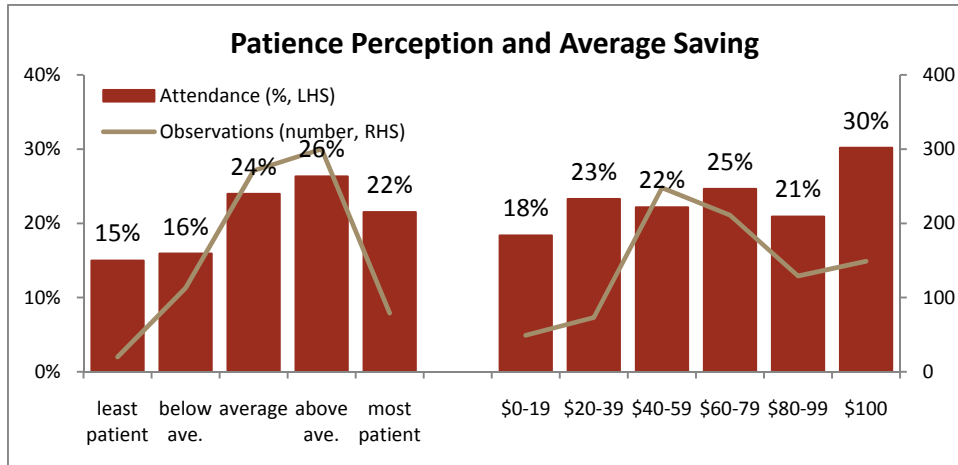|	(1)	(2)	(3)	Marginal Effect
Risk Attitude	-0.0947*		-0.0828	-0.0246
	(0.0542)		(0.0543)	(0.0162)
Lottery Decision		-0.0311	-0.0237	-0.00705
		(0.0239)	(0.0238)	(0.00709)
Constant	-0.640	-0.847	-0.601	
	(0.601)	(0.574)	(0.594)	
Observations	885	885	885	885
Log Likelihood	-462.3	-461.2	-460.2	-460.2

Probit regressions with robust standard errors at tutorial level in parentheses
*** p<0.01, ** p<0.05, * p<0.1; ^ p<0.1 (one tailed test)
Regressions include dummy variables for tutors, tutorial days & recruitment conditions

Patience: Figure 4.11 shows participation rates based on patience perceptions in the left panel and for the average savings amount across the six savings decisions in the right panel. Both panels show directionally that more patient individuals were more likely to participate. For instance, 16 percent of students who regarded themselves as less patient than their peers participated, whereas 24 percent of those assessing their patience as average or above average participated. Similarly, 18, 23 and 30 percent of those who chose to save less than $20, between $20 and $99, and $100, respectively, participated. The regressions in Table 4.6 show that greater perceived patience marginally (p<.10) increased participation (Column 1), while higher saving in the experimental task significantly (p<.05) increased participation (Column 2), and the significance level for both variables is the same when estimated together (Column 3).

Result 5: *Lab participants saved significantly more in the saving decisions and view themselves to be more patient than the population they were recruited from.*

Figure 4.11 Participation by patience perception and average savings

Table 4.6 Participation by patience perception and savings choices

	(1)	(2)	(3)	Marginal Effect
Patience	0.0882^		0.0799^	0.0237
	(0.0572)		(0.0582)	(0.0171)
Saving Decision		0.383**	0.381**	0.113**
(x100)		(0.184)	(0.182)	(0.0545)
Constant	-1.274**	-1.139**	-1.436**	
	(0.632)	(0.563)	(0.618)	
Observations	885	885	885	885
Log Likelihood	-462.6	-459.0	-457.7	-457.7

Probit regressions with robust standard errors at tutorial level in parentheses
*** p<0.01, ** p<0.05, * p<0.1; ^ p<0.1 (one tailed test)
Regressions include dummy variables for tutors, tutorial days and recruitment conditions

Discussion: We anticipated that lab participants would be less risk averse (H5) and patient (H6) than the population they were recruited from. Our evidence confirms the patience hypothesis, but rejects the risk hypothesis. It is possible that the lab experiment we invited our population to attend was perceived to have minimum risk since we advertised earnings "on average of at least $30," that is nearly twice the minimum wage and perhaps since the lab runs effectively with turn-away and delayed starting times almost never occurring, there was little perceived risk to participation.

4.1.6 Full model: implications for representative biases

Table 4.7 shows the estimates of the model including all the measures we collected (Column 1) and their marginal effects (Column 2). The estimates indicate that the core non-representative biases are robust to the inclusion of all the measures with most of the significant results reported above remaining significant when controlling for all other variables.

We thus focus here on the magnitude of non-representative biases in terms of the marginal effects (Column 3) and the implied large ratio of disproportionate representation of the characteristics of the lab participants relative to the population (Column 4). Consider the variable 'spending per week;' for every $20 more a student spent per week (one level shift in the survey), students were 4.0 percentage points less likely to participate in the lab experiment, which amounts to a 16 percentage point difference in participation between students with the highest and lowest weekly spending (H1 income).

Column 3 shows similar large differences in participation for all the core hypotheses: a 9.1 percentage point overall difference in participation for 'hours worked' (H2 leisure time), 11.6 percentage point for 'economics major' (H3 intellectual interests) and 17.9 percentage point for 'volunteering frequency' (H4 pro-social preferences).

The disproportionate representation of lab participants in Column 4 reports the ratio of the over-represented to under-represented group, assuming for simplicity an equal number of people in each response category in the population. Consider again weekly spending; the raw participation rate for each weekly spending response (Figure 4.2) show a ratio of 34 to 13 for lab participants in the lowest to highest income categories, which is approximately a **2.6 to 1** over-representation of the lowest to highest income group instead of a 1 to 1 ratio in the population. Column 4 shows a larger ratios of disproportionate representation for all core non-representative biases: **2.6 to 1** for students not working to working 16-20 hours a week, **1.5 to 1** for students majoring in economics and business to other majors, and **1.6 to 1** for students who volunteer 10 or more times a year compared to students who did not volunteer. We also include the disproportionate ratios for intellectual curiosity overall of **2.9 to 1** for students majoring in business who were both consistent in the savings decision and correctly answered all of the CRT questions correctly compared to non business and economics majors who were inconsistent in the savings decisions and answered all of the CRT questions incorrectly; and for more pro-social preferences **2.8 to 1** for students who volunteered more than 10 times and gave $200 in the dictator game compared to those who never volunteered and gave nothing in the dictator game.

Discussion: To test whether any of the variables led to significant differences in participation across the recruitment conditions, we re-estimated the full model in Table 4.7 with interaction terms for each variable by each recruitment condition. The model thus produces 51 pair-wise tests consisting of the 17 variables with three comparisons each: (1) Appointment vs. Drop-in; (2) Appointment vs. Option; and (3) Drop-in vs. Option. The results indicate that none of the 51 comparisons are significant at the $p = .05$ level, and only one of the 51 tests is significant at the $p=.10$ level.[41] Therefore, we cannot reject the null hypothesis that non-representative biases are the same across conditions. This result suggests that switching from the standard appointment recruiting procedures to either a drop-in or option procedure would not alleviate the participation biases.

---

[41] Students who volunteered more hours were less likely to participate in the appointment than drop-in condition (p<.10).

Table 4.7 Participation by all characteristics

	(1) Probit Est.	(2) Marginal Effects	(3) percentage point difference Top v. Bottom Category		(4) Non-repres. bias
Income H1					
Spending per week	**-0.147*****	**-0.0399*****	**-16.0%**	$0-40 v. > $100	2.6 to 1
	(0.0380)	(0.0101)			
Household Income (per year)	-0.0397	-0.0108	-6.5%	< $30,000 v. > $200,000	
	(0.0365)	(0.00992)			
Leisure Time H2					
Work Hour (per week)	**-0.0839****	**-0.0227****	**-9.1%**	no work v. 16-20 hours	2.6 to 1
	(0.0392)	(0.0106)			
Intellectual Interest H3					All three: 2.9 to 1
Econ/Business	**0.434*****	**0.116*****	**11.6%**	Econ/bus v. Other majors	1.5 to 1
	(0.130)	(0.0327)			
Saving Consistency	**0.136****	**0.0368****	**7.4%**	inconsistent v. consistent	1.4 to 1
	(0.0638)	(0.0174)			
CRT Score	**0.137***	**0.0372***	**11.2%**	0 v. 3 Correct	1.4 to 1
	(0.0740)	(0.0197)			
ATAR Score	-0.0285	-0.00774	-4.6%	Under 50 v. Above 99	
	(0.0462)	(0.0125)			
Pro-social Pref. H4					Both: 2.8 to 1
Volunteering Freq.	**0.165****	**0.0447****	**17.9%**	Never v. > 10 times	1.6 to 1
	(0.0738)	(0.0200)			
Volunteering Hours	-0.0427	-0.0116	-9.3%	0 v. > 100 hours	
	(0.0446)	(0.0122)			
Donation Freq.	0.0500	0.0135	5.4%	0 v. > 10 times	
	(0.0485)	(0.0132)			
Donation Dollars	-0.0559	-0.0152	-10.6%	0 v. > $500	
	(0.0357)	(0.00957)			
Dictator Game (x100)	**0.145***	**0.0394***	**7.8%**	0 v. $200	1.8 to 1
	(0.0753)	(0.0205)			
Risk Attitude	-0.0250	-0.00677	-2.7%	least v. most risk taking	
	(0.0558)	(0.0152)			
Lottery Decision	-0.0250	-0.00677	-3.4%	certainty v. risk seeking	
	(0.0257)	(0.00699)			
Patience	0.0592	0.0160	6.4%	least v. most patient	
	(0.0579)	(0.0155)			
Saving Decision (x100)	**0.317***	**0.086***	**8.5%**	0 v. $100	
	(0.184)	(0.051)			
Female	-0.155	-0.0407	-4.1%	female v. male	
	(0.137)	(0.0345)			
Caucasian	-0.0270	-0.00730	-0.7%	Caucasian v. Asian	
	(0.129)	(0.0349)			
Constant	-1.200*				
	(0.710)				
Observations	881	881			
Log Likelihood	-416.3	-416.3			

Probit regressions with robust standard errors at tutorial level in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Regressions include dummy variables for tutors, tutorial days and recruitment conditions

4.2 Appointment vs. Drop-in choices

This section briefly investigates students' decision to make an appointment or drop-in if given the option. Of the 60 students in the option condition who participated in the lab experiment, 43 made an appointment and 17 dropped in. We can thus easily reject that students who participated in the lab were equally likely to make an appointment or drop-in at the *p = 0.001* level. While this behavior signifies that the majority of students prefer to make an appointment when given an option, understanding what individual characteristics explain this preference is unfortunately difficult to infer since only 17 students chose to drop-in. To understand this choice given the small sample, we estimated probit regressions on the students who participated where the dependent variable equals 1 if they made an appointment and 0 if they dropped in, including only one independent variable at a time from the classroom tasks.

The estimates from each regression indicate that only one measure was significant, and no other measure was even marginally significant at the p = .10 level. Consistent with our hypothesis (H7a), we found that more risk-averse participants on the lottery decision were more likely to make an appointment when given the option ($p < 0.06$). Table 4.8 shows that students were 6.9 percentage points more likely to make an appointment for each additional less risky choice, thus a student who chose the sure $22 payoff was almost 35 percentage points more likely to make an appointment than someone who chose the riskiest option.

Result 7: *Among students given the option to make an appointment or drop-in, participants preferred making an appointment, and more risk-averse participants were more likely to make an appointment.*

Table 4.8 Chose to make an appointment given option condition and participated

	(1)	Marginal Effect
Lottery Decision	-0.205*	-0.0685**
	(0.105)	(0.0347)
Constant	1.220***	
	(0.388)	
Observations	59†	59†
Log Likelihood	-33.44	-33.44

Probit regressions with robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
†One lab participant in the option condition did not make a choice for classroom lottery decision.

Discussion: Result 7 may in part be due to the sentence we included in the option recruitment flyer in which we justified the option by writing, "appointments are helpful since spaces may be limited," which may suggest risks (e.g., being turned away or longer delays) with dropping in than having an appointment. Moreover, we found that subjects' lottery choice on participation was not significantly different across the recruitment conditions suggesting that while risk-aversion might affect how subjects choose to participate (appointment or drop-in), it does not affect whether they participate.

5. Addressing non-representative biases

The results indicate that the characteristics of lab participants are not representative of the characteristics of the population they were recruited from. We hypothesized and found several over

represented characteristics among lab participants including having less income, more leisure time, greater interest in economics lab activities and more pro-social over volunteering time. Researchers can address the potential concerns in both recruitment procedures and econometric analyses, for example if quantitative results of point estimates are important, or if a treatment (e.g. stakes) effect is likely correlated with certain characteristics (e.g. income) that also affects experiment participation. Indeed, several actions with respect to recruitment procedures are already in practice by some researchers.

5.1 Recruitment procedures

5.1.1 Reward based procedures

For biases caused by monetary rewards, **increasing experimental payments** could alleviate over-representation of *low income* groups in the lab. Although expensive, researchers have an option to offer higher payments if this bias would substantially affect the lab results. Our model and empirical finding suggest that higher anticipated earning will increase the likelihood of participation for everyone, thus potentially reducing the *low income* participation bias.[42] For biases caused by monetary rewards or the opportunity cost of time, experimenters may **reduce the lab time**, thus raising the hourly earnings rate $r(a)/t(a)$ to the same effect as increasing the monetary rewards. **Reducing the lab time** also allows more individuals with *higher value of time* to attend the lab experiment. Several additional procedures are possible to increase participation by **improving flexibility and convenience**, which reduces the *total* time and increases the ease for participants. For instance, the experimenter may 1) run experiments in more convenient locations to reduce the transit time to the lab, 2) offer more convenient or more flexible lab times when most of the population would be free, or 3) run experiments on-line if the study design permits, which allows people to choose the time and location. The use of the Internet to run 'traditional' lab studies has been growing increasingly common (e.g., Bellemare, Kröger and van Soest 2008, 2011; Slonim and Garbarino 2006; Garbarino and Slonim 2009; Horton, Rand and Zeckhauser 2011).

Another reward, potentially as powerful as monetary payments, to attract students to participate is **course credits**. Many business school and psychology departments create 'subject pools' and offer modest course credits for participation. To avoid coercion, students are typically given multiple options for earning the extra course credit besides participating in a study, and students are also free to choose none of the options. This approach often generates very high levels of participation, for instance the course credit offer by the Marketing Discipline at the University of Sydney typically has over 90 percent of students choosing to participate in a study during the term. Importantly, studies

---

[42] While the experimental economics literature has looked extensively at the effects of varying the stakes on behavior within a study, to our knowledge no study has directly varied the advertised earnings in the recruitment procedures in order to test non-representative biases. Some early psychology studies suggest that paying volunteering subjects might increase their representativeness (Rush et al. 1978, Wagner and Schubert 1976).

offering course credit can still be incentivized for the decisions within the study identically to current economic laboratory studies.[43]

5.1.2 Information based procedures

For biases caused by intrinsic motivations over the nature of experiments, carefully minimizing relevant signals in the recruitment information (by withholding or framing of the task information) can help recruiting participants with more neutral interests. Specific suggestions follow.

Not mentioning economics in the recruitment process can address over-represented *interest in economic lab tasks* among economics or business students. Our explicit indication in the recruitment flyer of an 'economics decision-making experiment' could be partly responsible for this 'economics' student bias we identified. If it is not possible to conceal a connection with economics, for example due to either ethics considerations or reputation, the experiments can be held in more convenient times and locations for non-economics and business students, for instance in locations near engineering, science and humanities buildings, or using a mobile lab (the mobile 'lab' can typically be set up in any large room).[44] Another approach is to run a lab in which a broader set of researchers beyond economists (e.g., psychologists) run experiments such that the recruited population would not know in advance what kind of lab activities might occur. Feasibility of this approach may depend on whether the standards for recruitment can be unified across the research communities (e.g., over compensation and deception).

Similarly, **not suggesting social value** of the lab experiment, e.g. never mention helping researchers or the broader community, can reduce participation bias based on *Pro-social preferences*. However, since most experiments are specifically run to enhance knowledge, experimenters can only go so far to minimize the communication of the value of the research without outright deception. Moreover, in our experiment the recruitment material never mentions helping researchers or the broader community, yet we still found significant participation bias on pro-social volunteering time.

5.2 Econometric analyses

Experimenters can also take several approaches to econometrically address non-representative biases. First, researchers can measure participants' characteristics on which there is anticipated bias (e.g., income, work hours, major areas of study, pro-social volunteering behavior) and then control for their influence in the analyses. If the researchers are making point estimate inferences, then weighted analyses can be reported for average population inferences. If researchers are making within experiment treatment inferences, they can control for bias by controlling for the interaction of the within experiment manipulations by the participant's characteristics and/or can report weighted

---

[43] It is possible though that offering course credits may cause a different set of participation biases such as attracting students who are more concerned about course grades.

[44] The 'economics' bias could also be due to potential participants believing knowledge of economics is important. To address this bias, recruitment materials can explicitly state that no knowledge of economics is necessary.

analyses for population inferences. However, this approach assumes the experimenters are able to find reliable population weights.

Second, experimenters can collect information on the population they recruit from. In many labs, this means collecting information on the population they invite into their recruitment databases, not just into the lab; Cleave et al. (2010) report that only 25 percent of the population invited joined the subject pool database, and Krawczwyk (2011) reports that less than 2.5% of his population joined the subject pool, suggesting that participation bias could mainly occur before an invitation to a specific experiment occurs. If information on participants who do and do not participate is available or can be collected, then standard approaches to selection (e.g., IV strategies) are possible (Heckman 1979; Heckman et al. 1998).

Third, experimenters can manipulate recruitment procedures to econometrically estimate the nature and extent of potential participation bias. Andersen et al. (2009) to our knowledge is one of the rare papers to have taken this approach. Their study held the expected payoffs constant and manipulated the variability of earnings in their recruitment advertisement. As hypothesized, they found greater risk aversion among the participants recruited with the lower variance. In this direction, experimenters can vary the expected earnings, for instance recruiting subjects for three earnings conditions: $r(a)$ and $M_1r(a)$ and $M_2r(a)$, where $M_2 > M_1 > 1$. Experimenters can then compare the behavior of the three distinctly recruited groups to test whether their behavior systematically differs, and use any differences to extrapolate for greater external validity.[45] Likewise, experimenters could vary the advertised time to be $t(a)$, $N_1t(a)$ and $N_2t(a)$, with $0 < N_1 < N_2 < 1$, and then in all conditions run an experiment that only lasts for $N_1t(a)$. In this setup, the experimenter can anticipate higher turnout the shorter the advertised experimental time, and can compare behavior to test whether the participants in the higher turnout condition (presumably with the shorter advertised lab time) behave differently than those in the longer advertised lab time conditions. If any differences occur, the experimenter can than extrapolate to improve external validity.

Fourth, a common technique used in survey work to address response bias involves examining subjects' behavior based on the temporal order of participation. The temporal order can be measured by either the sequential time in which subjects sign up or show up, or by the researchers sending advertisements multiple times to the same individuals (removing those who have participated from the subsequent advertisements). The implicit assumption with this approach is that those who sign up to participate sooner or with fewer announcements are more likely to exhibit the participation biases than those who sign up later, and if so, researchers can then calibrate the extent of the bias among those who participate. There are also several additional methods that experimentalists can consider

---

[45] One difficult challenge with this approach is that in the lab, whatever stakes actually get used, r(a), M1*r(a) or M2*r(a), in the lab only one amount will reflect the lab stakes, and thus this amount will be a surprise in the other conditions, and so any difference in behavior in the lab across the groups could be due to the surprise or the variance in the advertised payments.

adopting that are common to address participation bias in survey work (Miller and Smith 1983; Groves 2006).

5.3 Disclosing recruitment procedures

Experimenters can adopt a convention to report recruitment materials and procedures either in manuscripts or in on-line supplemental material like the current convention to provide experimental instructions. We believe this will not only enable readers to better assess the potential concern for participation bias, but will also help disseminate state of the art recruitment techniques and further standardize and facilitate comparison of results across labs.

6. Conclusion

Lab experiments are an increasingly common source of empirical data for economic research. We modeled the decision to participate in a lab experiment as a rational decision over costs and benefits, and derived several hypotheses regarding the characteristics of the people who would be more likely to participate. Consistent with our hypotheses, we found that recruited lab participants were not representative of the target population they were recruited from; participants had lower income, more leisure time, more interests in economics and lab activities, and were more pro-social on the dimension of volunteering time. Our estimates indicate that the rational participation decision led to an over-representation of most of the hypothesized characteristics by a ratio of more than 2 to 1 among the participants compared to the target population.

As mentioned in the introduction, any behavior measured in a lab study that is correlated with any non-representative characteristic may result in biased estimates. In such cases, lab researchers could follow two simple steps to address participation bias. First, researchers can adopt a convention to report recruiting procedures, both into subject pools as well as into specific studies, to allow readers to understand and compare results across labs for potential participation biases. It is also important to report the percent of individuals in the population invited who attend the lab study to allow readers to assess the potential for participation bias. Second, researchers can collect individual characteristics within their lab studies to control for potentially biased characteristics that would be correlated with outcome behaviors; indeed many researchers already routinely collect and control for socio-economic and demographic data. A third step, but more involved, would be to collect socio-economic and demographic data on individuals in the population before recruitment, either using existing administrative sources and or by collecting it themselves as we did in the current study. Researchers can then use standard econometric procedures to estimate and control for participation bias.

The most common approach on the more general question of generalization of lab studies has been to compare the behavior of lab subjects recruited from one population with lab subjects recruited

from another population.[46] This approach allows researchers to directly test the robustness of results across different subject populations. However, the recruitment procedures, which are typically the same across the subject populations, likely "homogenize" the participants between the populations so that their characteristics are more similar than the populations they were recruited from, and hence comparing participants could under-estimate differences in the populations they were recruited from. The potential for this homogenization will likely increase in studies with lower participation; in the current study, 23% of the population contacted ultimately participated in our lab study. Cleave et al. (2010) and Falk et al. (2011) report approximately 12% participation in their lab studies, and Krawczyk (2011) reports that only 2% joined the subject pool among those who were sent a mass email invitation.[47]

This paper focuses on the recruitment procedures and presents the first systematic comprehensive study motivated by theory to examine the voluntary participation decision. Our study used university students as the target population, which likely under-estimated the magnitude of participation biases since the characteristics of this population (e.g., age, income, education) are more homogeneous than a more general population. Beyond students, we hope to raise a general concern that experiment samples from a well-chosen population could potentially be biased due to recruitment. Nonetheless, given that many lab studies still primarily recruit from student populations, the results presented here are immediately relevant for a large body of research.

Lab research has made tremendous contributions to the economics literature. Its advantages for empirical study, including control, measurement, and replication, are well known. Perhaps one of the current remaining challenges for further influence involves addressing questions regarding robustness and generalizability. The current work suggests that the characteristics of the participants in lab studies are not representative of the population they were recruited from.[48] Well-known techniques are available to address bias, and this paper takes the first step to identify the sources and magnitudes of potential biases.

---

[46] Cleave et al. (2010) review this literature: example include comparing results across occupation (Cooper et al. 1999; Hannah et al. 2002; Fehr and List 2004; Carpenter et al. 2004; Güth et al. 2007; Carpenter et al. 2008; Andersen et al. 2009; Anderson et al. 2010), across age (Harbaugh et al. 2003), across nationality/culture (Roth et al. 1991; Henrich et al. 2001; Herrmann et al. 2008; Cameron et al. 2009; Henrich et al. 2010) and between student and a representative sample of the population (Bellemare and Kröger 2007, Anderson et al. 2010 and Falk et al. 2011).

[47] The current experiment was run one week after the initial recruitment, whereas Cleave et al. (2010) and Falk et al. (2011) ran their studies several weeks to months after the initial recruitment, which could explain the difference in lab participation rates.

[48] This conclusion should not seem shocking; we similarly do not assume professional athletes, entrepreneurs, truck drivers, data entry workers, bicycle messengers, sports card dealers, academics, blue collar workers, teachers or almost any other specific population we study would be perfectly representative of the broader population they come from.

References

Al-Ubaydli, O. and J. List (2012). On the generalizability of experimental results in economics, in Frechette, G. and A. Schotter (eds) *The Methods of Modern Experiments*, Oxford University Press.

Alatas, V., L. Cameron, A. Chaudhuri, N. Erkal, and L. Gangadharan (2009). Subject pool effects in a corruption experiment: A comparison of Indonesian public servants and Indonesian students. *Experimental Economics* 12(1): 113-132.

Andersen, Steffen, Glenn W. Harrison, Morten Igel Lau, and Elisabet Rutström (2009). Preference Heterogeneity in Experiments: Comparing the Field and Laboratory, *Journal of Economic Behavior and Organization,* forthcoming.

Anderson, Jon, Stephen V. Burks, Jeffrey Carpenter, Lorenz Götte, Karsten Maurer, Daniele Nosenzo, Ruth Potter, Kim Rocha and Aldo Rustichini (2010). Self Selection Does Not Increase Other-Regarding Preferences among Adult Laboratory Subjects, but Student Subjects May Be More Self-Regarding than Adults, *IZA Discussion Paper* No. 5389.

Andreoni, James and Charles Sprenger (2010). Estimating Time Preferences from Convex Budgets, *NBER Working Paper* No. 16347

Baran, Nicole, Paola Sapienza, and Luigi Zingales (2010). Can We Infer Social Preferences from the Lab? Evidence from the Trust Game, *CEPR Discussion Paper* No. DP7634.

Bellemare, Charles, and Sabine Kröger (2007). On Representative Social Capital, *European Economic Review*, 51(1), 181-202.

Bellemare, Charles, Sabine Kröger and A. van Soest (2008), Measuring Inequity Aversion in a Heterogeneous Population using Experimental Decisions and Subjective Probabilities, *Econometrica*, 76(4), 815-839.

Bellemare, Charles, Sabine Kröger and A. van Soest (2011), Preferences, Intentions, and Expectations Violations: a Large-Scale Experiment with a Representative Subject Pool, *Journal of Economic Behavior and Organization*, 78(3), 349-365.

Belot, M., R. Duch, and L. Miller. (2010). Who Should be Called to the Lab? A comprehensive comparison of students and non-students in classic experimental games. *Nuffield centre for experimental social sciences, University of Oxford, Discussion paper series*

Benz, Matthias and Stephan Meier (2008). Do people behave in experiments as in the field? Evidence from Donations, *Experimental Economics*, 11 (3), 268-281

Berg, Joyce, John Dickhaut and Kevin McCabe (1995). Trust, Reciprocity, and Social History, *Games and Economic Behavior*, 10(1) 122-142.

Binswanger, Hans P. (1980). Attitudes toward risk: Experimental measurement in rural India. *American Journal of Agricultural Economics,* 62, 395–407.

Binswanger, Hans P. (1981). Attitudes toward risk: Theoretical Implications of an Experiment in Rural India. *Economic Journal*, 91, 867–890.

Bolton, Gary E and Axel Ockenfels (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *The American Economic Review*, 90(1), 166-193.

Bowles, Samuel, and Sandra Polania-Reyes (2011). Economic Incentives and Social Preferences: Substitutes or Complements?, *Journal of Economic Literature,* forthcoming.

Brañas-Garza, Pablo, Filippos Exadaktylos and Antonio Espín Martín (2011). Experimental subjects are not special. *Working paper*

Burks, S., J. Carpenter, and L. Goette. (2009). "Performance pay and worker cooperation: Evidence from an artefactual field experiment." *Journal of Economic Behavior & Organization*, 70:3, pp. 458-69.

Camerer, Colin (2011). The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. *Working paper*

Cameron, Lisa, Ananish Chaudhuri, Nisvan Erkal, and Lata Gangadharan (2009). Propersities to Engage in and Punish Corrupt Behavior: Experimental Evidence from Australia, India, Indonesia and Singapore, *Journal of Public Economics*, 93 (7-8), 843-851.

Cardenas, J. (2011). "Social Norms and Behavior in the Local Commons as Seen Through the Lens of Field Experiments." *Environmental and Resource Economics*, pp. 1-35.

Cardenas, J. C. (2005). "Groups, Commons and Regulations: Experiments with Villagers and Students in Colombia," in *Psychology, Rationality and Economic Behaviour: Challenging Standard Assumptions*. Agarwal and Vercelli eds: International Economics Association.

Carpenter, Jeffrey, Cristina Connolly and Caitlin Myers (2008). Altruistic Behavior in a Representative Dictator Experiment, *Experimental Economics*, 11(3), 282-298.

Carpenter, Jeffrey, Stephen Burks and Eric Verhoogen (2004). Comparing Students to Workers: The Effects of Social Framing on Behavior in Distribution Games, *IZA Discussion Papers* No. 1341

Casari, Marco, John Ham and John Kagel (2007). Selection bias, demographic effects and ability effects in common value auctions experiments, *American Economic Review*, 97 (4), 1278-1304.

Cleave, Blair Llewellyn, Nikos Nikiforakis and Robert Slonim (2010). Is There Selection Bias in Laboratory Experiments?, *Univ. of Melbourne Dept. of Economics Working Paper* No. 1106

Cooper, David J., John Kagel, Wei Lo and Qing Liang Gu (1999). Gaming Against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers, *American Economic Review*, 89 (4), 781-804.

Dixon, P. N. (1978), 'Subject recruitment incentives, personality factors, and attitudes towards experimentation in a simultaneous intentional-incidental learning task', Journal of General Psychology 99, 99–105.

Eckel, Catherine, and Philip Grossman (1996). Altruism in Anonymous Dictator Games, *Games and Economic Behavior, 16*, 181-191.

Eckel, Catherine, and Philip Grossman (2000). Volunteers and Pseudo-Volunteers: The Effect of Recruitment Method in Dictator Experiments, *Experimental Economics*, 3, 107–120.

Eckel, Catherine, and Philip Grossman (2008). Forecasting Risk Attitudes: An experimental Study Using Actual and Forecast Gamble Choices, *Journal of Economic Behavior & Organization*, 68(1) 1-17.

Falk, Armin, Stephan Meier and Christian Zehnder (2011). Did We Overestimate the Role of Social Preferences? The Case of Self-Selected Student Samples, *Journal of the European Economic Association*, forthcoming.

Fehr, Ernst and John List (2004). The Hidden Costs and Rewards of Incentives – Trust and Trustworthiness among CEOs, *Journal of the European Economic Association*, 2, 741-771.

Fehr, Ernst and Klaus M. Schmidt (1999). A Theory of Fairness, Competition, and Cooperation, *The Quarterly Journal of Economics*, 114(3), 817-868

Frederick, Shane (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4): 25–42.

Garbarino, Ellen and Robert Slonim (2009). The Robustness of Trust and Reciprocity Across a Heterogeneous Population, *Journal of Economic Behavior and Organization*, 69(3), 226-40.

Garbarino, Ellen, Robert Slonim and Justin Sydnor (2011), Digit Ratios (2D:4D) as Predictors of Risky Decision Making, *Journal of Risk and Uncertainty*, 42(1), 1-26.

Glaeser, Edward, David Laibson, Josè A. Scheinkman and Christine L. Soutter (2000), Measuring Trust, *Quarterly Journal of Economics* 115(3): 811-846.

Greiner, Ben (2004). The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. *University of Cologne, Working Paper Series in Economics* 10.

Grether, David M. and Charles R. Plott (1979). Economic Theory of Choice and the Preference Reversal Phenomenon, *The American Economic Review* 69(4) 623-638

Groves, Roberts M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70, 646-675.

Güth, Werner, Carsten Schmidt and Matthias Sutter (2007). Bargaining outside the lab – a newspaper experiment of a three-person ultimatum game. *The Economic Journal*, 117: 449–469.

Güth, Werner, Rolf Schmittberger and Bernd Schwarze (1982). The Robustness of Trust and Reciprocity Across a Heterogeneous Population, *Journal of Economic Behavior and Organization*, 3(4), 367-388.

Hannah, Lynn, John Kagel and Don Moser (2002). Partial Gift Exchange in an Experimental Labor Market: Impact of Subject Population Differences, Productivity Differences and Effort Requests on Behavior, *Journal of Labor Economics*, 20 (4), 923-951.

Harbaugh, William T., Kate Krause, Steven G. Liday and Lise Vesterlund (2003). Trust in Children, in *Trust, Reciprocity and Gains from Association* edited by Elinor Ostrom and James Walker, Russell Sage Foundation, New York.

Harrison, Glenn W. and John A. List (2004). Field Experiments, *Journal of Economic Literature*, 42(4) 1009-1055

Harrison, Glenn W., Morten Lau and Elisabet Rutström (2009). Risk Attitudes, Randomization to Treatment, and Self-Selection Into Experiments, *Journal of Economic Behavior and Organization* 70, 498–507.

Heckman, James (1979). Sample Selection Bias as a Specification Error, *Econometrica*, 47 153–61.

Heckman, James, Hidehiko Ichimura, J. Smith and Petra Todd (1998). Characterizing Selection Bias Using Experimental Data, *Econometrica,* 66, 1017-98.

Henrich, Joseph, Robert Boyd, Sam Bowles, Colin Camerer, Herbert Gintis, Richard McElreath and Ernst Fehr (2001). In Search of Homo economicus: Experiments in 15 Small-Scale Societies, *American Economic Review*, 91(2) 73-79.

Henrich, J., S. Heine and A. Norenzayan (2010). Most People are not WEIRD. *Nature*, 446: 29.

Herrmann, Benedikt, Christian Thöni and Simon Gächter (2008). Antisocial Punishment Across Societies, *Science*, 319(5868), 1362-7.

Hoffman, M. H. and J. Morgan (2011). Whos Naughty? Whos Nice? Social Preferences in Online Industries. *Working paper*.

Holt, C. (2006). *Markets, Games, and Strategic Behavior: Recipes for Interactive Learning*. Addison-Wesley.

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

Horton, John, David Rand and Richard Zeckhauser (2011), The online laboratory: conducting experiments in a real labor market, *Experimental Economics,* forthcoming.

Jackson, J. M., Procidano, M. E. & Cohen, C. J. (1989), 'Subject pool sign-up procedures: a threat to external validity', *Social Behavior and Personality* 17, 29–43.

Jung, J. (1969). Current practices and problems in the use of college students for psychological research, *The Canadian Psychologist* 10, 280–290.

Kagel, J., R. Battalio, and J. Walker (1979). Volunteer Artifacts in Experiments in Economics: Specification of the Problem and Some Initial Data from a Small Scale Field Experiment, *Research in Experimental Economics*, 1.

Kahneman, Daniel and Amos Tversky (1979). Prospect Theory: An Analysis of Decision under Risk, *Econometrica* 47(2) 263-292

Krawczyk, Michael (2011). What brings subjects to the lab? *Experimental Economics*, 14(4), 482-489.

Levitt, Steven D., and John A. List (2007a). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World, *Journal of Economic Perspectives*, 21(2), 153-174.

Levitt, Steven D., and John A. List (2007b). Viewpoint: On the Generalizability of Lab Behaviour to the Field, *Canadian Journal of Economics,* 40(2), 347-370.

List, J. 2004. Young, Selfish, and Male: Field evidence on social preferences. *Economic Journal*, 114, pp. 121-49.

List, John (2006), The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions, *Journal of Political Economy*, 114(1), 1-37.

Machina, Mark J. (1987). Choice Under Uncertainty: Problems Solved and Unsolved, *The Journal of Economic Perspectives* 1(1) 121-154

McKelvey, Richard D. and Thomas R. Palfrey (1992). An Experimental Study of the Centipede Game, *Econometrica* 60(4) 803-836

Miller, L. E., & Smith, K. L. (1983). Handling Nonresponse Issues. *Journal of Extension*, 21(5), 45-50.

Nagel, Rosemarie (1995). Unraveling in Guessing Games: An Experimental Study, *The American Economic Review* 85(5) 1313-1326

Orne, Martin (1962). On the Social Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications, *American Psychologist*, 17(10), 776-83.

Remoundou, Kyriaki, Andreas Drichoutis and Phoeben Koundouri (2010). Warm glow in charitable auctions: Are the WEIRDos driving the results? *MPRA Paper No. 25553*

Roe, Brian E., Timothy C. Haab, David Q. Beversdorf, Howard H. Gu, Michael R. Tilley (2009). Risk-attitude selection bias in subject pools for experiments involving neuroimaging and blood samples. *Journal of Economic Psychology*, 30(2), 181–189

Rosen, Ephraim (1951). Differences between Volunteers and Non-Volunteers for Psychological Studies, *Journal of Applied Psychology*, 35, 185-193.

Rosenthal, Robert, and Ralph Rosnow (1969). *Artifacts in Behavioral Research*, Oxford University Press.

Rosenthal, Robert, and Ralph Rosnow (1973). *The Volunteer Subject*, New York: John Wiley and Sons.

Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara and Shmuel Zamir (1991). Bargaining and market behavior in Jerusalem, Ljubiljana, Pittsburgh, and Tokyo: An experimental study, *American Economic Review*, 81, 1068-1095.

Rush, M. C., Phillips, J. S. & Panek, P. E. (1978). Subject recruitmemt bias: The paid volunteer subject, *Perceptual and Motor Skills* 47, 443–449.

Saunders, D. M., Fisher, W. A., Hewitt, E. C. & Clayton, J. P. (1985). A Method For Empirically Assessing Volunteer Selection Effects: Recruitment Procedures and Responses to Erotica, *Journal of Personality and Social Psychology* 49, 1703–1712.

Senn, C. Y. & Desmarais, S. (2001). Are Our Recruitment Practices for Sex Studies Working Across Gender? The Effect of Topic and Gender of Recruiter on Participation Rates of University Men and Women, *Journal of Sex Research* 38, 111–117.

Silverman, I. & Margulis, S. (1973). Experiment tilte as a source of sampling bias in commonly used 'subject pool' procedures, *The Canadian Psychologist* 14, 197–201.

Slonim, Robert and Ellen Garbarino (2006). Increases in trust and altruism from partner selection: experimental evidence, *Experimental Economics*, 11(2), 134-153.

Stahl, Dale (1996). Boundedly Rational Rule Learning in a Guessing Game, *Games and Economic Behavior*, 16, 303-330.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization, *Review of Economic Studies* 23(3), 165-80.

Tversky, Amos and Daniel Kahneman (1992). Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and uncertainty* 5(4) 297-323

Wagner, M. E. & Schubert, D. S. P. (1976). Increasing volunteer representativeness by recruiting for credit or pay, *The Journal of General Psychology* 94, 85–91.