'Pocket and Pot':
Hypothetical Bias in a No-Free-Riding Public Contribution Game

Tatiana Gubanova,[1] W. L. Adamowicz,[2] Melville McMillan[3]

[1]Economist, Economics and Markets Division, Alberta Department of Energy
9945 108 Street, Edmonton, Alberta, Canada T5K 2G6
Phone: (780) 644-2566; fax: (780) 422-9677; email: Tatiana.Gubanova@gov.ab.ca

[2]Professor and Associate Dean (Research), Department of Rural Economy
Faculty of Agricultural, Life and Environmental Sciences, University of Alberta
Edmonton, Alberta, Canada T6G 2H1
Phone: (780) 492-4603; fax: (780) 492-0268; email: Vic.Adamowicz@ales.ualberta.ca

[3]Professor, Department of Economics, Faculty of Arts, University of Alberta
Edmonton, Alberta, Canada T6G 2H4
Phone: (780) 492-7629; fax: (780) 492-3300; email: Melville.Mcmillan@ualberta.ca

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2009
AAEA & ACCI Joint Annual Meeting, Milwaukee, Wisconsin, July 26–29, 2009

‘Pocket and Pot’: Hypothetical Bias in a No-Free-Riding Public Contribution Game

Tatiana Gubanova,*
W. L. Adamowicz, Melville McMillan

April 29, 2009

Abstract

Hypothetical bias arises when values which people say they place on a good or service differ systematically from the values people reveal for the same good or service through actual, binding economic transactions. Studies of hypothetical bias with respect to public goods often use charitable contributions or other relatively unique goods and these studies employ a variety of mechanisms to elicit the stated and revealed values. This study proposes the inclusion of a free-rider barring random dictatorship mechanism in the standard public contribution game to investigate the issue of bias when a public good involves immediate monetary returns to subjects. Steps are taken to make the game have the look and feel of a real-world tradeoff between private investment and public good provision. Data for the experiment were collected using a sample of students from the University of Alberta. A statistically significant negative hypothetical bias is found for the first hypothetical and the first real rounds of the game. The bias decays in subsequent round pairs, oscillating around zero.

1 Introduction and Objectives

Economic valuation of private goods in market economies is generally based on market prices, from which willingness to pay (WTP) for the respective commodity can be inferred. Public goods do not have market prices. Markets fail to emerge for public goods because these goods are non-excludable and non-rival in consumption. The basic principle behind public good valuation is thus to make their values comparable with those of private goods. This requires attaching some sort of monetary value to public goods.

The hypothetical nature of contingent valuation surveys, a technique that can be used to measure the value of essentially any public good, can result in responses that differ systematically in magnitude from actual payments, if those were to be collected. This difference between stated and revealed values is often referred to as ‘hypothetical bias’ (Cummings, Brookshire, and Schulze 1986).[1] There is a common belief among non-market valuation researchers that hypothetical bias exists and is usually positive, that is WTP revealed from a real choice experiment is lower than that revealed from a hypothetical choice experiment (e.g., Cummings, Harrison, and Rutstrom 1995, Loomis et al. 1997, List and Shogren 1998). In addition, there is a growing literature that focuses on testing techniques designed to mitigate hypothetical bias.

*Corresponding author. Support for this project was provided by the Social Sciences and Humanities Research Council of Canada (SSHRC).

[1]The terms revealed, real and actual are used interchangeably and refer to situations in which an individual makes a consequential economic commitment. In experimental studies, this typically involves monetary payment for a good by the participant. Stated or hypothetical values refer to survey responses that lack any salient economic commitment (Murphy et al. 2005).

Examples of these techniques are 'cheap talk' scripts (Cummings and Taylor 1999) and 'consequential' survey (Bulte et al. 2005). There is no established theory explaining hypothetical bias despite numerous studies investigating this phenomenon and hypotheses suggested.

Two basic questions about hypothetical bias have become paramount (Murphy et al. 2005). First, what is its magnitude (both the existence and sign)? Second, what factors are responsible for this bias? The first objective of this study is to assess the magnitude of hypothetical bias in a laboratory experiment with a modified standard public goods game as an elicitation mechanism. We also attempt to evaluate the effect of several behavioral factors as well as certain socio-demographic factors on the degree of hypothetical bias.

Experiments are being used to 'testbed' valuation techniques in non-market valuation literature (Adamowicz and DeShazo 2006). Theoretically, hypothetical bias can be detected by conducting two sets of experiments which should be identical in everything except one aspect. In one experiment, people must actually buy a good after stating their WTP (real experiment), while the other experiment is purely hypothetical and subjects do not make any market transactions (hypothetical experiment).

We believe that the standard public goods game is the appropriate foundation for such a set of experiments. It offers a public good whose consumption is easy to monetize and which can be easily provided both hypothetically and for real. The abstract nature of the contribution game is not conducive to effects of warm-glow or any kind of 'do-gooder' bias. Also, the standard public goods game provides a unique mechanism for eliciting both hypothetical and real values of the good. But the standard public goods game has its own shortcomings which make it *per se* inappropriate for non-market valuation purposes. Accordingly, we modify the standard voluntary contribution game in two ways to ensure the compatibility of our experiment with a non-market valuation study. What emerges is a 'Pocket and Pot' game.

First, a risky lottery is introduced in addition to the public pool (the 'Pot') and the uninvested portion. The standard public goods game offers a public good whose consumption is easy to monetize and which can be easily provided both hypothetically and for real. However, the monetary value of any good only makes sense if a dollar spent on that good has the same value as the next best use. Second, the random dictatorship mechanism is introduced to determine the 'winning' contribution to the public Pot. The standard public goods game allows its players to practice strategic behavior which includes free-riding. Random dictatorship as a collective choice mechanism is introduced to eliminate the possibility of free-riding to the extent possible (Gibbard 1977 and Dutta, Peters, and Sen 2002). In other words, no free-riding is allowed in the suggested game, no one can be excluded from this redistribution and each player gets the same amount of tokens from the common pool.

In what follows, Section 2 provides a literature review on hypothetical bias. The rationale for the experiment and a hypotheses to test are covered in Section 3. Section 4 lays out the design of the modified public goods game and details its major components. Experiment results are presented in Section 5. Section 6 concludes with a summary of findings, addresses limitations of the game and outlines recommendations for future research.

2 Literature Overview

Bohm (1972) was one of the first to raise questions about hypothetical bias and its magnitude. Having compared several different approaches to estimating the demand for a public good and checked for the presence of hypothetical bias, Bohm (1972) establishes several requirements that must be met in order to get meaningful results about how much people would like to contribute to the output of a public good. First, the provision of a public good must be of a concern to the majority of the population. Secondly, the public good chosen must be well described and be realistic (quality, specific characteristics, the cost of production, etc.). Finally, for the experiment to mimic a real-world situation, it is necessary to have a large number of respondents involved.

Following Bohm's seminal paper, numerous non-market valuation studies had analyzed the hypothetical bias issue. Bishop and Heberlain (1986) found that hypothetical values for hunting permits exceed actual values by 30–130%.[2] In contrast, Dickie, Fisher, and Gerking (1987) did not find a statistically significant hypothetical bias. Subsequent research, however, has suggested that stated values typically exceed actual values. For example, Cummings, Harrison, and Rutstrom (1995) find hypothetical bias to be 163%; Loomis et al. (1997) find that hypothetical values exceed actual values by 86–200%; List and Shogren (1998) find hypothetical bias to be 119–247%. Sometimes the magnitude of hypothetical bias is very large; Neil et al. (1994) find that hypothetical values exceed actual values by 201–2,401%. Several studies fail to reject the null hypothesis of zero hypothetical bias (Johannesson, Liljas, and Johansson 1998, Sinden 1988, Smith and Mansfield 1998). There are also several private good valuation studies that report negative hypothetical bias (McClelland, Schulze, and Coursey 1993, Johannesson, Liljas, and Johansson 1998, and Griffin et al. 1995).

What are the reasons or factors explaining hypothetical bias? While numerous studies look at the bias' magnitude, the literature still lacks models which would put it in an economic framework. Champ et al. (1997) and Champ and Bishop (2001) suggest that hypothetical bias is manifested by an identifiable minority of subjects. Champ, Moore, and Bishop (2004) propose that the bias results from counting uncertain 'yeses' as firm ones when valuing the good hypothetically. Brown and Taylor (2000) investigate the role of gender as a possible explanation of hypothetical bias and find no significant difference in the percentage of females and males stating a positive value in both hypothetical and real treatments. However, the mean WTP for females in the study is significantly different from that for males in both the hypothetical treatment and the real one.

Although a large number of studies with different experimental protocols provide evidence that respondents tend to overstate their true WTP in hypothetical settings, other studies do not show any presence of hypothetical bias; see Sinden (1988), Johannesson, Liljas, and Johansson (1998), Smith and Mansfield (1998), Cummings and Taylor (1999), Balistreri et al. (2001).

In response to the observation that stated preferences are prone to hypothetical bias, researchers have proposed different ways to reduce the gap between stated and revealed preference estimates. For example, the National Oceanic and Atmospheric Administration's (NOAA) guidelines, NOAA (1994) and NOAA (1996), suggested calibrating hypothetical responses to actual choices, and several researchers proposed providing additional explicit information so that respondents could self correct their stated preferences during the survey or experiment. This additional information can be provided in the form of 'cheap talk' (Cummings and Taylor 1999) or in the form of a 'consequential' survey or experiment (Bulte et al. 2005).[3][4]

The reported efficacy of the cheap-talk design has not always been high. For example, Aadland and Caplan (2003) find that, although cheap talk is ineffective overall, it successfully reduces hypothetical bias for certain groups of respondents. Aadland, Caplan, and Phillips (2007) treat cheap talk as a mechanism that reduces uncertainty — respondents can ignore it completely or take it into account and use to update the uncertain value they place on the good. To test the theoretical result, Aadland, Caplan, and Phillips (2007) propose an experimental design that distinguishes the effects of anchoring and cheap talk on the valuation of a generic public good by mimicking the way contingent valuation surveys are typically implemented in practice.

There have been attempts to combine the numerous hypothetical bias studies together in order to establish what study design factors affect hypothetical bias and in what direction; see List and Gallet (2001), Murphy et al. (2005). List and Gallet (2001) conducted a meta analysis on data from numerous experimental studies in order to determine how experimental protocol affects hypothetical bias measured by the calibration factor.[5] There are several major results of the study. First, the data suggested that the calibration factor is not affected by whether an experiment takes place in the lab or field. It provides ev-

[2]Hanemann (1984) points out some deficiencies in this study.

[3]'Cheap talk' entails reading a script that explicitly highlights the hypothetical bias problem before participants make any decisions.

[4]It has been argued that respondents who perceive a survey or experiment to be 'consequential' will respond to questions truthfully regardless of the degree of perceived consequentiality (Mitchell and Carson 1989).

[5]The calibration factor is the ratio of the average hypothetical WTP and the average real WTP.

idence that nuances such as subject pools, social distance, and subtleties associated with the laboratory setting may not compromise the generality of empirical findings. Second, the calibration factor obtained from a WTP study is lower than a comparable calibration factor from a WTA study. Third, the calibration factor obtained from the analysis of a private good is lower than a comparable calibration factor from the analysis of a public good. Lastly, the calibration factors are not significantly different between within- and between-group treatments. Murphy et al. (2005) noticed a major flaw in the meta-analysis described above. Several of the preference valuation studies used in List and Gallet (2001) report different mechanisms for eliciting hypothetical and actual values. It means that hypothetical bias now is attributed not only to monetary consequences of a real choice, but is also contaminated by different elicitation mechanisms. Among other things, Murphy et al. (2005) report that both student pools and group experiments seem to reduce hypothetical bias. Also, choice-based elicitation mechanisms are reported to mitigate the difference between hypothetical and actual values. Finally, the primary factor that explains hypothetical bias is, according to Murphy et al. (2005), the magnitude of the hypothetical value. Nevertheless, the rationale of this finding is not clear without a well-developed theory of hypothetical bias.

3 Experiment Rationale and Hypotheses to Test

The literature review in the previous section shows that although some studies do not find any statistically significant hypothetical bias (Sinden 1988, Johannesson, Liljas, and Johansson 1998, Smith and Mansfield 1998, Cummings and Taylor 1999 and Balistreri et al. 2001) and several studies report a negative hypothetical bias (McClelland, Schulze, and Coursey 1993, Johannesson, Liljas, and Johansson 1998 and Griffin et al. 1995), people are likely to overstate their WTP in hypothetical situations (e.g., Cummings, Harrison, and Rutstrom 1995, Loomis et al. 1997, List and Shogren 1998). The evidence of the presence of positive hypothetical bias in valuation studies is so strong that, to counteract the hypothetical bias, researchers have proposed different methods ranging from 'numerical calibration' to 'verbal calibration' (e.g., Cummings and Taylor 1999, Bulte et al. 2005).

While numerous traditional valuation studies have addressed the issue of a hypothetical bias, several of its shortcomings seem to prevent one from dealing with the bias in an unambiguous way:

(1) Impossibility of comparing hypothetical and real values of a good on a *ceteris paribus* bases because the researcher is unable to actually provide the public good as required for the real choice part. Using different elicitation mechanisms, say, the contingent valuation method (CVM) and the hedonic price method (HPM) (a) may lead to unacceptable context differences and thus inference contamination regarding the magnitude of hypothetical bias, and (b) is only feasible with goods already being provided.

(2) Unfamiliarity of the good to respondents, lack of experience with it and/or its 'remoteness' (example of the latter is rainforest in Costa Rica) result in a questionable value of this good for most people and thus welfare changes from its consumption are hard to monetize (Diamond and Hausman 1994).

(3) Non-market valuation studies have often been criticized for framing contribution to the good as a 'good thing to do' (Diamond and Hausman 1994).

Non-market valuation researchers (e.g., Mitchell and Carson 1989; Sinden 1988; Fischhoff and Furby 1988) have sought to establish conditions under which contingent valuation surveys yield valid estimates of the valuation of goods. The following issues need to be addressed in order to obtain unbiased estimates of WTP values:

(1) For valuation purposes, the good must be well described and be meaningful, familiar, or of a concern for respondents (Mitchell and Carson 1989, p. 192; Sinden 1988, p. 102; Bohm 1972, p. 116).

(2) The property rights and the market for the good must be described in such a way that the respondents will accept the WTP format as plausible (Mitchell and Carson 1989, p. 192; Bohm 1972,

p. 116).

(3) Valid estimates of the value of a good can only be obtained in the environment where individuals are fully informed about possible choice alternatives, uncoerced, and able to identify their own best interests (Bohm 1972, p. 116; Fischhoff and Furby 1988, p. 1).

The standard public goods game which deals with a well-described mechanism of making investments in an abstract public good appears to satisfy all of these three criteria.

There is a history of using voluntary contribution games to improve the contingent valuation method (CVM). Prince et al. (1992) appear to be the first to investigate properties of a contribution game mechanism for the purpose of improving CVM. They simulate a field survey in the laboratory in order to investigate the performance of the contribution game mechanism. Bagnoli et al. (1992) show that the standard contribution game induces a downward bias in aggregate WTP values with respect to WTP reported in a standard contingent valuation framework. Research conducted by Prince et al. is directed at finding ways that the value elicitation questions in CVM may be designed, in light of the performance of the contribution game, and how the resulting data may be used to glean unbiased estimates of WTP values for public goods. Several modifications are introduced to the contribution game in order to make the laboratory setting more consistent with conditions that might exist in a survey used for valuing a local public good.

Following the practice of modifying voluntary contribution games depending on tasks under investigation, we introduce two major modifications to the standard public goods game, which are described in the next section. We design an experiment to rid of the aforementioned shortcomings and thus, to minimize the chances of contaminating the test of hypothetical bias. The following hypothesis regarding hypothetical bias will then be tested:

> The null hypothesis is that there is no difference between the hypothetical and real values of a laboratory public good in the experiment with the modified standard public goods game as an elicitation mechanism. The alternative hypothesis is that the hypothetical and real values of the good are different, with the direction of the difference unknown.

4 Design of Experiment

4.1 General Design

At the beginning of the game, each player receives ten tokens which can be invested in the private Pocket, the common Pot, or left 'uninvested'. The game is played in two steps. First, each player allocates her endowed tokens according to her own preferences and understanding of the game. Once each player submits her allocation, the computer determines what happens to the tokens. The computer does not make any choices of its own but acts according to a pre-programmed set of rules. The computer collects the choices of all players, draws random numbers, and reports game results to the players. The experiment is conducted in a computerized laboratory where subjects anonymously interact with each other. Players are requested to make token allocation decisions individually. No communication between players is allowed during the game.

The game is constructed to reflect a real-life situation of making a choice between private purchases (investments) and investing in a public good. The 'Pocket' represents the private sector of an economy or can be thought of as a private good. The 'Pot' represents a public good. The standard public goods game provides only two options for investing endowed tokens. They are usually named a 'personal account' and a 'public good'. In a standard public goods game, the personal account is not risky and is similar to the 'uninvested' option in our experiment. This option is somewhat similar to a 'choose none' option in contingent valuation. Introducing the risky private Pocket where an investor, with stochastic probabilities, can gain or lose makes the game more realistic. The Pot is also risky in the experiment because the level of the public good that an individual desires can not be guaranteed to that individual. That is, the individual may not attain her desired allocation of public and private goods due to collective decision making. The same risk is present in both standard public goods games and other valuation studies.

There is a problem common to the valuation of those actual public goods which are imprecisely defined or unusual. In order to participate in a public goods game with a real good, subjects need to have sufficient experience with this good and to know their own WTP beforehand, without the necessity to construct it or guess it on the fly. Charitable contributions or 'certificates' of stewardship of environmental resources, which are common public goods in valuation experiments, provide examples to the contrary (Diamond and Hausman 1994). These define unusual and/or abstract goods and may be of value only to a limited number of people from the general population. Money or tokens are easily understandable by any set of participants. According to neoclassical microeconomics, the marginal utility of money is positive; that is, players consider money to be a desirable good. The underlying assumption of our experiment is that marginal utility of both money and tokens is positive. The research question is whether players differentiate between tokens in real rounds which can be exchanged for money (the exchange rate is one token for one Canadian dollar) and tokens in hypothetical rounds which are not exchangeable for money.[6]

Six consecutive rounds of the game are played by each player. A multiple-round game is chosen over a one-round game because many real-world activities, including investment decisions, that have aspects of risk, public good provision, and other social dilemmas are typically not one-time encounters, but rather repeated undertakings (Levitt and List 2007).

Table 1: Description of Four Treatments

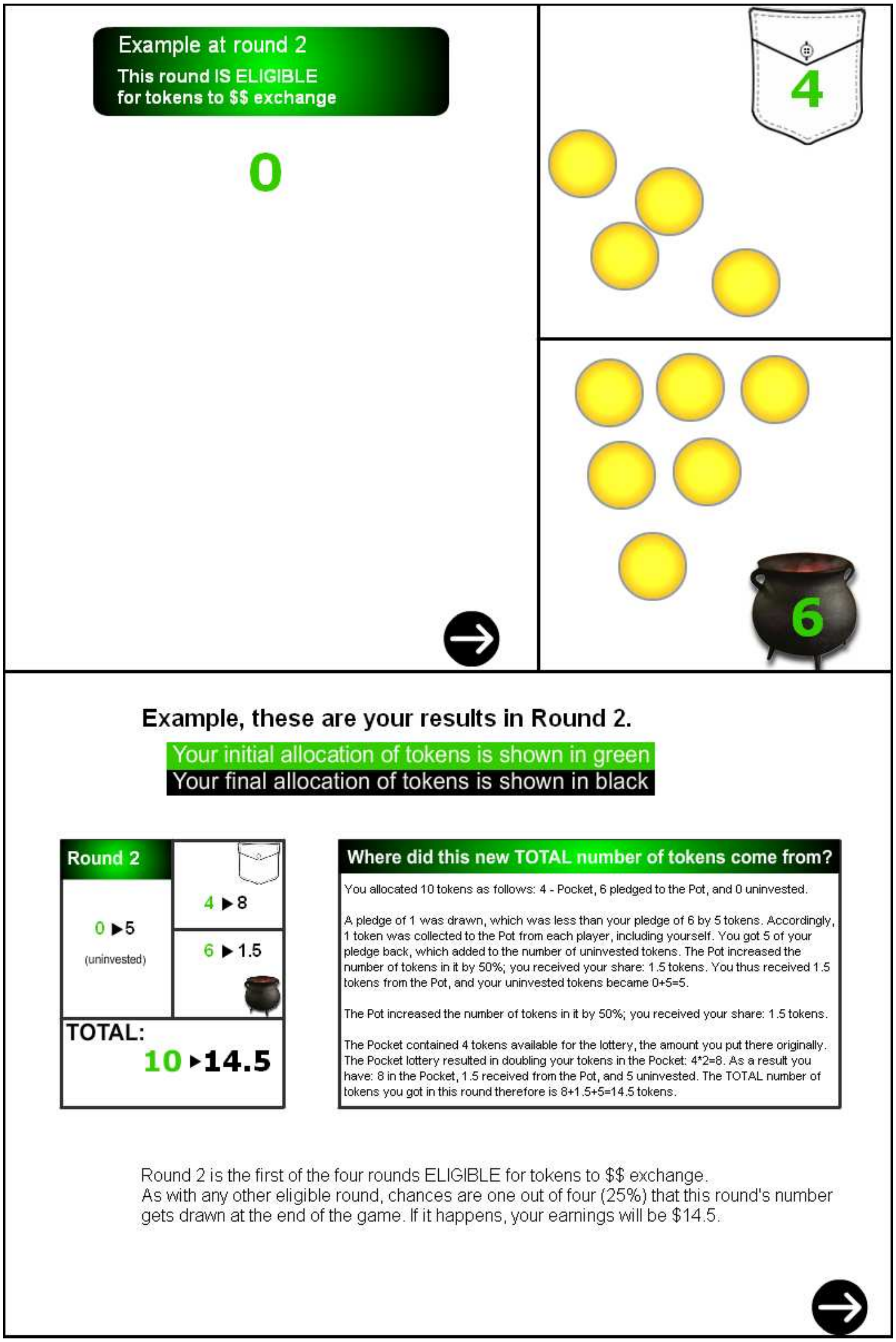
	HRC	RHC
Total number of rounds	6	6
Hypothetical	3	3
Real	3	3
Order of Rounds:		
Round 1	**H**ypothetical	**R**eal
Round 2	**R**eal	**H**ypothetical
Rounds 3–6	random	random

There are two treatments used in the game, that we denote by HRC and RHC. Differences between the treatments are explained in Table 1, which is discussed below. Each player participates in one treatment only. The first six rounds consist of three hypothetical (H) and three real (R) rounds where players have an option to invest any integer amount of tokens from 0 to 10 (in the 'complete' (C) version) or where players have an option to invest only 0, 5, or 10 tokens (in the 'short' (S) version). Finally, the seventh round is always real but the complexity is changed.[7] For the complete version, players face the reduced investment flexibility — that is, they are allowed to invest only 0, 5, or 10 tokens in the last round of the game. For the short version, players are offered the full investment flexibility — that is,

[6]One common complaint about laboratory models is that they are not sufficiently realistic to engage fully the attention of the participants. To mitigate this problem, we follow the dictates of experimental economics in using real monetary gains and losses. Admittedly, these losses are trifling compared to real-world losses. Nevertheless, observations in the laboratory indicate that small monetary losses are large enough so that everyone clearly prefers that the loss not happen. For example, subjects in the experiment reacted with visible unhappiness whenever they lost money in the Pocket lottery. The presence of both tokens and dollars mimics a traditional non-market valuation approach where hypothetical and real settings are distinguished by participants. This experiment could have been done with monetary units only.

[7]This is done to compare real outcomes under different complexity scenarios forcing players' homogeneity (each player makes choices under both complexity levels). The results of the complexity treatments are not discussed in this paper.

Figure 1: Screen Shots from the Game



they are allowed to invest any integer number of tokens, from 0 to 10, in the last round of the game. At the beginning of the first round, players are aware that there are only four rounds where tokens are eligible for money exchange (real rounds) and three rounds where they are not (hypothetical rounds). In addition, players are aware which round is which when they start playing the rounds. They are not aware of the change in complexity in the last round, which is needed to overcome the 'rational expectations' phenomenon (Muth 1961). If players had been fully informed about the change in complexity levels in the last round then, according to rational expectations theory, they could have adjusted their strategies beforehand and no difference in a person's behavior would have been detected between the two different complexity levels. For technical reasons, we chose to specify the order of rounds as follows. The first round is always hypothetical and the second round is always real in the HRC and HRS treatments, and the first round is always real and the second round is always hypothetical in the RHC and RHS treatments. Results of the first round become available for players only after the second round is played, together with the second round results. By doing this we ensure that (1) subsequent hypothetical and real rounds start with the same information (in other words, we ensure that no learning and/or strategy-adjustment is possible between the first two rounds), and (2) the first two rounds are completely identical except for the fact that the real round can be binding. The latter is crucial for the detection of hypothetical bias. The order of the subsequent rounds 3–6 is determined randomly by the computer for both complete and short versions of the game.

Each player starts each round with ten endowed tokens, and rounds are independent of each other. Players may reallocate their endowed tokens in each round. Subjects are informed that the final amount of tokens that is redeemable for money is that from only one of the three real rounds. This restriction is a standard means to reduce the costs of conducting an experiment. At the same time, we hope to minimize any possible influence of strategic behavior on game results, which can arise when subjects are aware in advance which real round is binding, by not revealing which real round will be used as the basis for redeeming tokens.

Results for each round are generated instantaneously by the computer once everyone submits one's allocation. The result screen (available starting with the second round) shows each player's submitted allocation of tokens, payoffs for the Pocket, for the Pot, and for the 'uninvested' portion as well as the total payoff (sum of the payoffs for the Pot, for the Pocket, and for the 'uninvested' portion), and a step-by-step description of how the computer determined the payoffs in this round; see Figure 1. No information on other players' investment decisions or outcomes is provided for the following two reasons. First, it can bias respondents' allocation decisions in subsequent rounds. Secondly, a person in a real-life choice situation is usually unaware of other people's contributions for a public good.[8] We chose to not provide subjects with their own history of results from previous rounds because we believe that when such a history is provided, respondents are further motivated towards strategic behavior. The goal of this experiment is to extract respondents' preferences towards a private good (the Pocket) and a public good (the Pot) which, according to the conventional economic theory, are known to subjects before the game begins. If this is true, the subjects should behave consistently throughout the experiment, in accordance with their preferences. If this is not true, subjects may try to behave strategically. In either case, providing the history of final payoffs for each round risks introducing additional bias to the experiment.

4.2 Pocket and Pot Rules and Payoffs

The experiment has two investment options: the private Pocket and the common Pot. After players submit their allocations, the computer uses a preprogrammed set of rules to determine what happens with payers' investments. Rules for the Pocket and the Pot are described below.

The Pocket represents a private sector of the economy or can be considered to be a private good. There are equal chances (50:50) that the number of tokens in the Pocket will double, or that a player will

[8]The pledge drawn by the computer is the only piece of information which is revealed to all players and is related to other player's choices. Players potentially may adjust their investment decisions in subsequent rounds based on the pledge drawn by the computer in previous rounds.

lose half of the tokens she puts there. The lottery outcome is determined randomly, by the computer. The risk associated with putting tokens in the Pocket is intended to parallel the investment risk in a real-life situation. Pocket lottery outcome probabilities and outcomes (a stake doubles with probability 0.5 and reduces by half with probability 0.5) were subjectively chosen, based on the following considerations:

- There is a need to motivate players to allocate all tokens between the Pocket and Pot rather than leave tokens 'uninvested'. It can jeopardize the objective of this study — to investigate the difference between the Pot/Pocket investments in hypothetical and real rounds — when the majority of players leave their tokens 'uninvested'. The expected payoff of the Pocket lottery is 1.25 times higher than the amount invested. At the same time, the Pocket is characterized by stochastic risk (that is, ambiguity). According to the theory of uncertainty, risk-loving and risk-neutral players should prefer the option of investing their tokens to the Pocket rather than keeping them 'uninvested'.
- Rules for computing the Pocket lottery outcome should be easy for players to understand and to remember.
- The primarily goal of the experiment is to investigate the presence of hypothetical bias in the Pot contributions. Therefore, one would want players' contributions to the Pot to be in the interval (0, 10), conditional on the presence of the risky Pocket and risk-less 'uninvested' portion. As such, the Pocket lottery should not be too attractive relative to the Pot. For this reason, the expected payoff of the Pocket lottery is 1.25 which is 20% lower than the expected Pot payoff (1.5), given that the Pot contribution gets drawn by the computer.[9]

Another investment option is the Pot, which represents a public good. After each participant personally chooses how many tokens to put in her Pot, this amount becomes her pledge to the common pool. The computer randomly draws one pledge out of all players' pledges, collects this number of tokens from each player, and places them in the common pool. In other words, the computer acts as a random dictator. Each pledge submitted by players has an equal probability of being drawn by the computer. Players are informed that pledging is like voting: the chances of a particular pledge being drawn increases with the number of players who pledged this amount. Once the pledge is determined, the computer increases the number of tokens in the pool by 50%. This bonus imitates the production process.[10] Finally, each player gets back an equal share of tokens from the common pool.

If a player pledged more than the pledge drawn by the computer, the difference is reimbursed to the player in the form of 'uninvested' tokens. When tokens are reimbursed to the player, they cannot be used for the Pocket lottery. If the pledge submitted equals the pledge drawn, no reimbursement nor additional tokens are provided/needed. If a player pledged less than the pledge drawn, the player must provide the difference. First, 'uninvested' tokens are used to cover the deficit. Any remaining balance is transferred from the Pocket. Whatever remains in the Pocket is eligible for the Pocket lottery.

The Pot is characterized by non-excludability and non-rivalry in consumption. In this case, after a pledge is drawn by the computer, each player gets an equal amount of tokens depending on the pledge drawn. No one can be excluded from this redistribution. Everyone gets the same amount of tokens, that is one's consumption of this good does not reduce other players' consumption.

[9] In the first version of this game, the Pocket lottery odds were two to one, that is 67% chance of success versus 33% chance of failure. Results obtained from the first focus group showed that most of the players used the investment options fully, that is invested either in the Pocket or in the Pot. The average percent of tokens invested in the Pocket was quite high (56%) relative to the Pot contributions (41%). We decided to equate chances of success and failure in order to reduce the attractiveness of the Pocket as an investment option.

[10] The production technology multiplier for the Pot contributions is chosen such that it is not too close to unity; 1.5 seems like a reasonable number. At the same time, the production multiplier (1.5) is higher than the expected Pocket lottery multiplier (1.25). Such a difference is needed to make the Pocket less attractive as an investment option compared to the Pot.

4.3 Random Dictatorship Mechanism

Introducing a random dictator is another step distancing the present game from the standard public goods game. Introducing the random dictator eliminates the possibility of free-riding. In this game, the random dictator is introduced in a sense that once the pledge is drawn by the computer, everyone must provide this amount of tokens to the common pool. In other words, no free-riding is allowed. The random dictatorship mechanism (RDM) is not implemented in the standard public goods games because one of the goals of these games is to study free-riding or strategic behavior of players. In this study, we use a modified standard public goods game as a tool (rather than the focus of the study) to investigate hypothetical bias in public good contributions in the familiar-public-good and no-free-riding environment. RDM is a strategy-proof mechanism (Gibbard 1977) in a one-round game: RDM does not offer an individual incentives to misrepresent her preferences to secure a better personal outcome. Gibbard's proof regarding a one-round game can be extended to a multiple-round game similar to ours. The knowledge of the pledge drawn by the computer in the previous round is the only information that can motivate a player to misrepresent her preferences in the next round. If the player misrepresents her preferences by mimicking her Pot contribution in the next round according to the pledge drawn by the computer in the previous round, she gets the second best outcome which results in a lower utility level and this person is not a utility maximizer. The utility maximizer would stick to her preferences and if her Pot contribution gets drawn by the computer, she gets the first best outcome which results in the highest utility level. Under these conditions, RDM is strategy-proof in a several-round game such as ours. In addition, Dutta, Peters, and Sen (2002) show that any strategy-proof (in Gibbard's sense) and unanimous mechanism must be a random dictatorship.

In addition, the random dictatorship mechanism, as opposed to majority voting and median voter mechanisms, is characterized by the following aspects that are useful for this game:

- Compared to the majority voting mechanism, RDM is more convenient for games with many levels of provision of a public good. The majority voting mechanism works well in the presence of two levels of provision of a public good. In this game, there can be eleven different levels of provision (from 0 to 10) in the complete version, and there can be three levels of provision (0, 5, or 10) in the short version.
- Compared to the median voter mechanism, RDM appears to be easier to explain to participants. The concept of the median may be hard to grasp for a subject not familiar with statistical concepts.

It should be noted that a random-dictator mechanism not only eliminates free-riding but also imposes a non-voluntary contribution environment. But a non-voluntary contribution environment is not that totally alien to the real world: in the real world, public goods are financed using tax money. Taxes are mandatory and the only way to free ride is tax evasion. Also, a random-dictator mechanism opens a possibility to overrule the individual decision about one's private investment (when a pledge drawn by the computer is more than one's pot contribution and uninvested tokens are not enough to cover the deficit) which is a limitation of the game.

4.4 Debriefing Questionnaire

After the game is played, each subject is given a questionnaire which consists of two parts. The first part contains questions about the strategies adopted by players during the game and their risk attitudes. The second part contains standard socio-demographic questions. In the first part of the questionnaire, players are asked to rate their attitude towards risk on the 11-point attitude rating (Likert) scale. The lower end-point of 0 means that the respondent is extremely risk averse. The upper end-point of 10 means that the respondent is extremely risk loving.

Six questions are asked about strategies adopted by players during the game. Again, an 11-point Likert scale was used. The lower end-point of 0 means that the respondent completely disagreed with the respective question. The upper end-point of 10 means that the respondent fully agreed with the respective question.[11] Subjects are asked to rate the following statements: whether they were primarily interested in maximizing the Pocket payoff; whether they were primarily interested in maximizing the Pot payoff; whether they were interested in keeping their money safe; whether they used several strategies during the game depending on the outcome of the previous round(s) (this question is asked to check the consistency of preferences as well as the presence of strategic behavior); whether their choices were purely random; and whether subjects found the Pocket to be riskier than the Pot (this question is asked to collect information on risk perception towards the Pot).

Finally, players are asked whether, while listening to the game instructions, they had an impression that the Pot had been presented as an alternative preferable to the Pocket on social, moral, or ethical grounds. This question is aimed at capturing whether or not subjects exhibit 'warm glow' and/or compliance bias while contributing to the Pot.[12] To the best of our knowledge, instructions to the game are written and presented in such a way to avoid any description of the Pot as a socially preferable alternative compared to the Pocket. In other words, no bias is expected.

In the second part of the questionnaire, players are asked several socio-demographic questions regarding their gender, student/non-student status, department (if applicable), monthly living expenses, presence of a part-time job, volunteer experience, and experiment participation experience.

5 Analysis of Experiment Results

5.1 Experiment Participants

In total, eight sessions of the experiment were held, two of each treatment. The number of participants in a session ranged between 7 and 11, depending on how many of the invited subjects showed up for the experiment. Each participant had an option to discontinue her participation at any time during the study, although nobody opted out. Each subject participated in only one of the four experimental treatments. All sessions were held in February and March 2008, at the University of Alberta. Subjects were recruited with the use of ORSEE — Online Recruitment System for Economic Experiments, maintained by the Department of Rural Economy, University of Alberta. This data base is used by the Department of Rural Economy researchers for recruiting subjects for various experiments, games, and surveys. Experimental sessions lasted about one hour and subjects earned, on average, CA$20–25, which included a show-up fee of CA$5.

Subjects were either undergraduate or graduate students at the University of Alberta, with supposedly little knowledge of economics. There may be a few disadvantages of using students for laboratory experiments. Rosenthal and Rosnow (1969) note that student subjects are likely to be 'scientific do-gooders', interested in the research, or students who readily cooperate with the experimenter and seek social approval. On the other hand, List and Gallet (2001) as well as Murphy et al. (2005) find that student pools do not necessarily compromise the generality of empirical results. We believe that, for this simple behavioral experiment with hard-to-guess objectives, the use of a student pool is legitimate and should not negatively affect its results.

5.2 Token Allocations

Initial token allocations and round results averaged across players are presented in Table 2. For the complete version, players on average choose to invest more tokens in the Pot than the Pocket, by 1.53 and 2.26 tokens in hypothetical and real rounds, respectively. For the short version, players on average

[11] Note that the rating scale for these questions differs from that for the risk attitude question. As a result, the risk attitude ratings will not be standardized, whereas responses for the other rating questions will be.

[12] Warm glow represents the purchase of moral satisfaction while contributing to the public good (Andreoni 1990). Players in the present study may have contributed to the Pot because it is 'a good thing to do'. Compliance bias would be committed by individuals who did not find the Pot attractive for investment purposes and had no intentions to contribute to the Pot but would still contribute a positive amount to it in a belief that this is what experimenters expect them to do.

choose to invest by 1.58 and 1.09 tokens more in the Pocket than the Pot in hypothetical and real rounds, respectively.

To determine what could have attributed to the differences in investment behavior of players in different versions of the game, it would be useful to look at a detailed picture of players' initial token allocations across rounds and game versions, which is discussed later in this section.

Round results — UnOut, PockOut, and PotOut (see Table 2) — calculated for each player are affected by the initial token allocation made by the player as well as other players' Pot contributions. Therefore, round results are not mutually independent due to the experiment construction. The violation of the assumption of mutually independent observations makes it impossible to analyze the results of each round using statistical tests. Initial allocations, on the other hand, are mutually independent, which places no restriction on the statistical tests implementation.

For the purpose of analyzing experimental results, non-parametric techniques are chosen over parametric methods. Application of parametric methods requires a number of assumptions regarding distributions of variables under consideration. These assumptions can hardly be provided without a formalized theory on hypothetical bias. In addition, there are several reasons for the ineffectiveness of regression analysis applied to the obtained data. First, the linearity in explanatory variables is a convenient but often unrealistic assumption. In case of a non-linear regression, which non-linear function should be used to model the data collected in the experiment? There is no theory to help us find the correct non-linear form of the model. Second, the size of our data set is quite small for regression purposes (forty two players for the complete version and thirty seven players for the short version). Finally, regression analysis deals with the partialing-out effect of one explanatory variable under consideration, keeping other variables constant. The lack of variation in the latter due the the small sample size makes partialing-out rather questionable (yet computationally possible).

A detailed picture of Pot contributions, Pocket investments, and uninvested tokens across round pairs for different game versions is presented in Figures 2, 3, and 4, respectively.

Figure 2 provides information on the complete and short version strategies regarding contributions to the Pot across round pairs. Players' Pot investment behavior is somewhat similar in the first pair of hypothetical/real rounds for both versions: five tokens is the most popular Pot contribution in both hypothetical and real rounds and in both complete and short versions of the experiment. But it seems that the complete version players are more likely to invest a non-zero number of tokens in the Pot than are the short version players.

As the game proceeds, the short version players appear to lose their interest in the Pot. For the second pair of hypothetical/real rounds in the complete version, a majority of players prefer to put ten or seven or six tokens in the Pot in a hypothetical round, whereas they put five or ten tokens there in a real round. In the third pair of rounds, the bulk of the players put zero or six tokens in the Pot in a hypothetical round and seven or ten tokens in a real round. For the short version, a majority of the players prefer to put zero or five tokens in the Pot in both hypothetical and real rounds up until the end of the game.

Figure 3 shows the complete and short version strategies with respect to Pocket investments across round pairs. For the first pair of rounds in the short version, a majority of players put five tokens in the Pocket in a hypothetical round, whereas they put zero tokens there in a real one. A similar picture is observed for the complete version players: a majority of them put two or five tokens in the Pocket in a hypothetical round and zero tokens in a real round.

Players' investment behavior appears to change starting with the second pair of hypothetical/real rounds. The majority of the short version players put five or ten tokens in the Pocket in both hypothetical and real rounds until the end of the game. The bulk of the complete version players put from zero to five tokens in the Pocket in hypothetical and real rounds until the end of the game. In other words, there is a striking difference in players' attitude towards the Pocket in the complete and short versions of the game that emerges in the second pair of hypothetical/real rounds and then goes on. The short version players seem to prefer making high Pocket investments whereas the complete version players are less likely to do so.

Figure 4 shows that there is not much of a difference in players' preferences towards uninvested

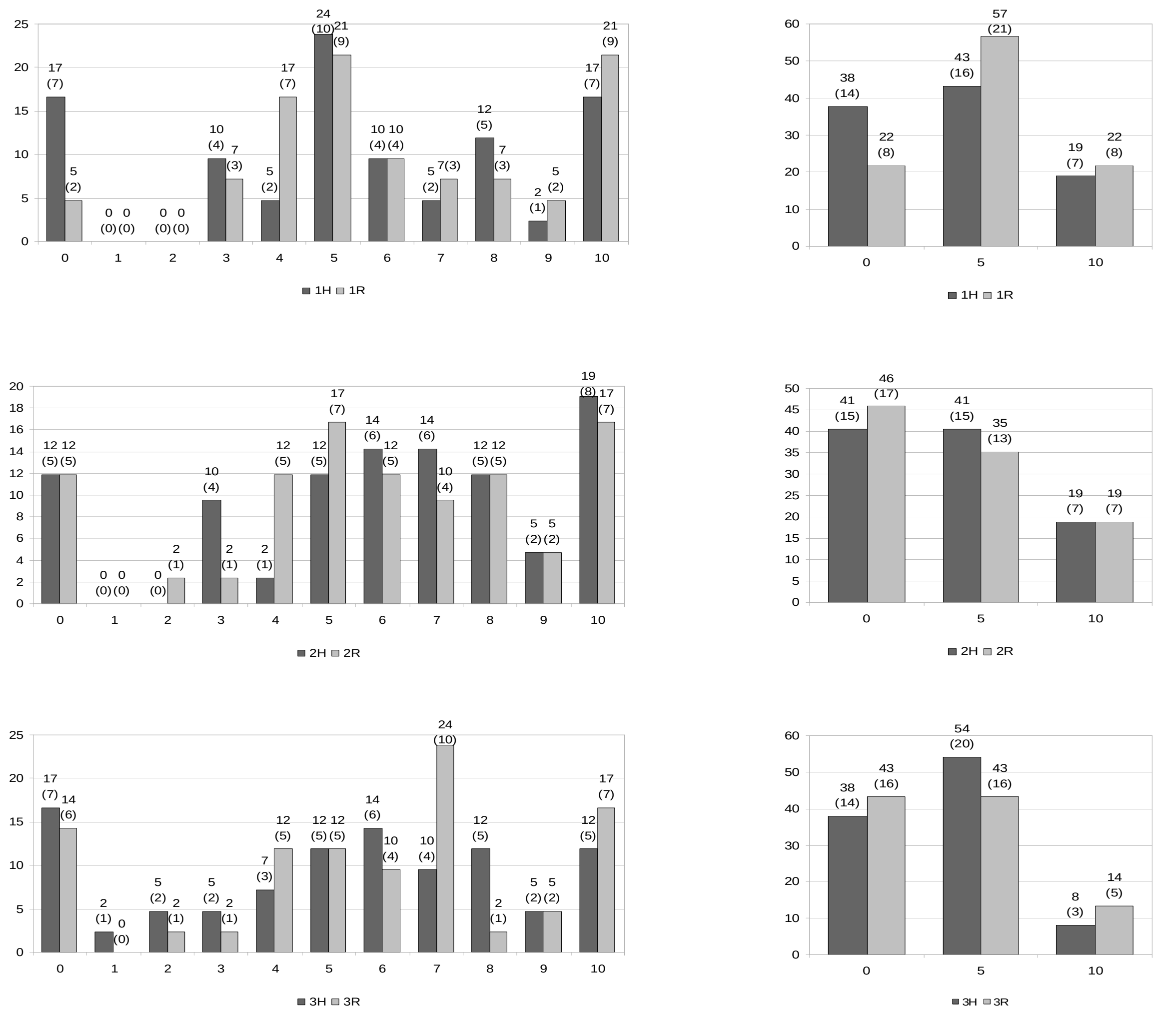
Table 2: Experiment Results Averaged Across Players and Rounds

	UnIn[a]	PockIn	PotIn	(PotIn–PockIn)	UnOut	PockOut	PotOut	Total[b]
	Complete Version (# of players = 42)							
Average investments in a hypothetical round	0.35	4.06	5.59	1.53	2.20	3.88	7.15	13.23
Average investments in a real round[c]	0.44	3.65	5.91	2.26	2.05	2.95	8.50	13.50
	Short Version (# of players = 37)							
Average investments in a hypothetical round	0.77	5.41	3.83	−1.58	1.13	5.11	7.50	13.74
Average investments in a real round[c]	0.81	5.14	4.05	−1.09	2.97	6.13	3.92	13.02

[a] 'UnIn', 'PockIn' and 'PotIn' denote players' initial investments in the uninvested portion, the Pocket and the Pot; 'PotIn-PockIn' is the difference between the Pot and the Pocket investments; 'UnOut', 'PockOut' and 'PotOut' indicate token allocations as a result of each round.
[b] Final payoff of each round is the sum of 'UnOut', 'PockOut' and 'PotOut'. Real payoff can be binding.
[c] Excluding round 7 which is characterized by the reversed complexity level.

Figure 2: Pot Contributions Across Round Pairs

(a) Complete version (b) Short version

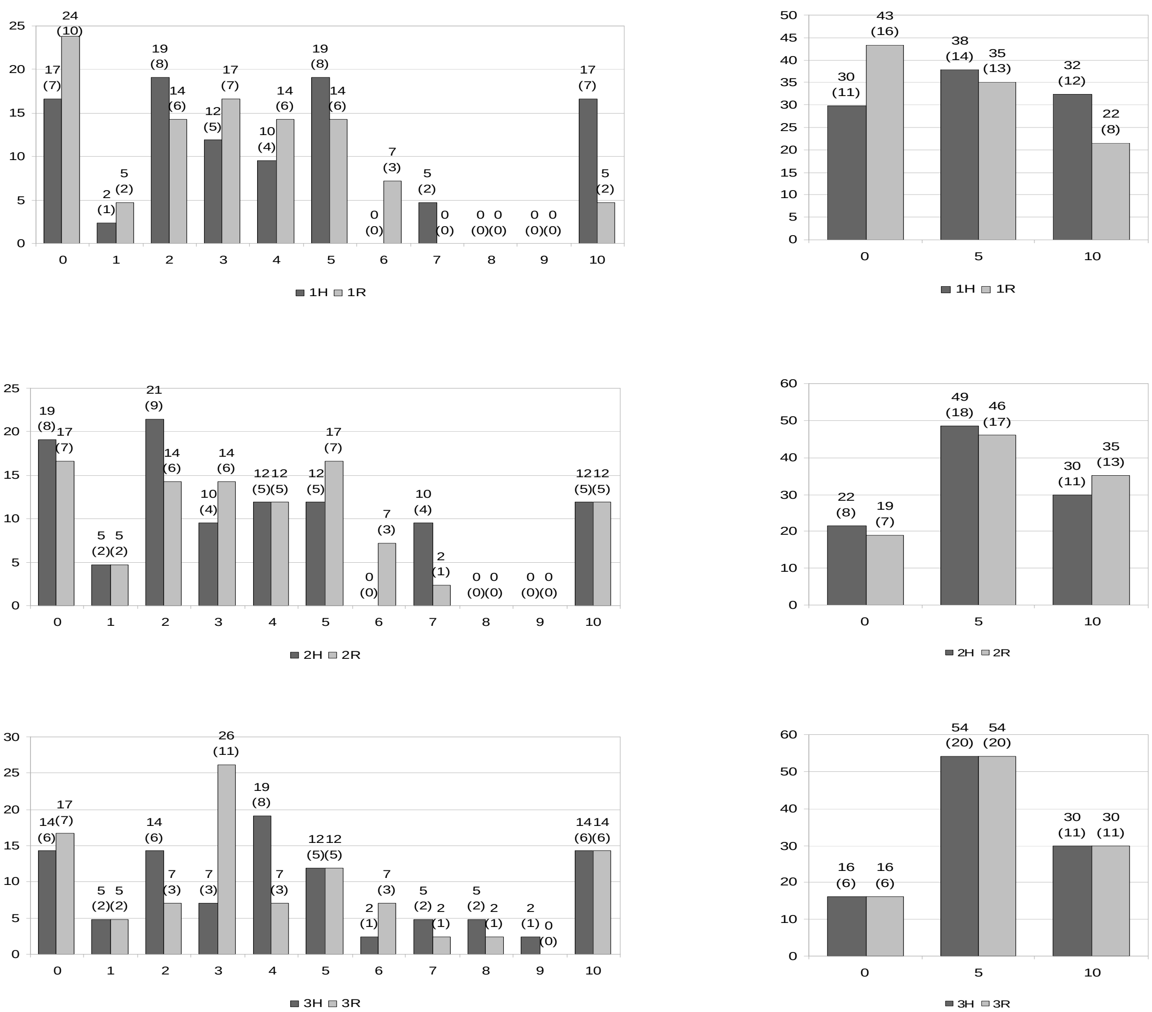


The horizontal axis shows all possible numbers of tokens to be allocated to the Pot; the percentage breakdown of the total of 42 (complete version) and 37 (short version) players is shown on the vertical axis. The numbers denote the percentages of players who contributed the respective numbers of tokens to the Pot. The numbers below, in parentheses, are actual token counts. 1H/1R, 2H/2R, and 3H/3R denote first, second, and third pairs of (H)ypothetical/(R)eal rounds, respectively.

Figure 3: Pocket Investments Across Round Pairs

(a) Complete version

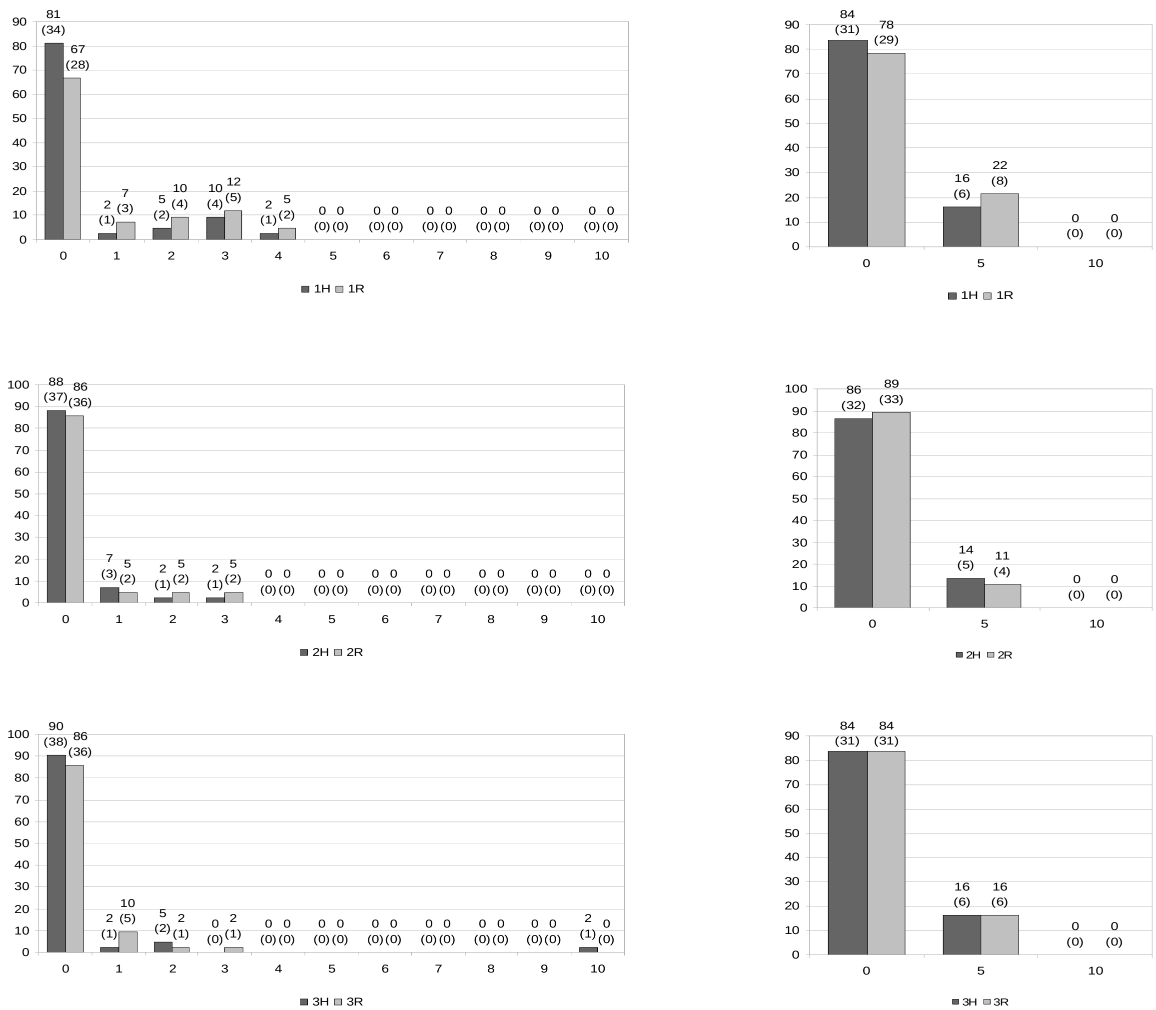
(b) Short version



The horizontal axis shows all possible numbers of tokens to be allocated to the Pocket; the percentage breakdown of the total of 42 (complete version) and 37 (short version) players is shown on the vertical axis. The numbers denote the percentages of players who contributed the respective numbers of tokens to the Pocket. The numbers below, in parentheses, are actual token counts. 1H/1R, 2H/2R, and 3H/3R denote first, second, and third pairs of (H)ypothetical/(R)eal rounds, respectively.

Figure 4: Uninvested Tokens Across Round Pairs

(a) Complete version (b) Short version



The horizontal axis shows all possible numbers of tokens to be left uninvested; the percentage breakdown of the total of 42 (complete version) and 37 (short version) players is shown on the vertical axis. The numbers denote the percentages of players who left the respective numbers of tokens uninvested. The numbers below, in parentheses, are actual token counts. 1H/1R, 2H/2R, and 3H/3R denote first, second, and third pairs of (H)ypothetical/(R)eal rounds, respectively.

tokens for the complete and short versions of the game. For the short version, players prefer to leave zero or five tokens uninvested in both hypothetical and real rounds. For the complete version, the number of uninvested tokens varies from zero to three, depending on the round.

Figures 2 to 4 show that token allocations in the Pot and the Pocket are very similar for the first pair of hypothetical/real rounds for the complete and short version players. It appears that the short version players focused on the Pocket (lost interest in the Pot) starting with the second pair of hypothetical/real rounds. There is a possible explanation to this observation. Players are allowed to invest 0, 5, or 10 tokens in the short version. These investment options may not fully reflect players' preferences which, combined with the trade-off between the stochastic risk in the Pocket and the risk associated with other players' choices for the Pot, may have resulted in relatively low Pot investments and relatively high Pocket investments. Also, players invest the largest portion of their initial endowments in the Pocket and the Pot, leaving none or almost no tokens uninvested in both versions of the game. In addition, the short version players seem to be able to find the preferable strategy faster and thus achieve 'convergence' faster (possibly because of the limited investment flexibility). It takes longer for the complete version players to achieve 'convergence' (possibly because of the high investment flexibility which is exploited by the players at an increasing rate as they proceed through the game).

Analysis of Table 2 and Figures 2 to 4 can be used to examine the difference in players' Pot and Pocket investment behavior between the complete and short versions of the game. Is the group playing the complete version different from the group playing the short version? In an attempt to answer this question, the following null hypothesis was tested using the Friedman test for nine factors:[13]

$$\mathrm{H}_0 : \text{Complete factor} = \text{Short factor} \quad \text{versus} \quad \mathrm{H}_1 : \text{Complete factor} \neq \text{Short factor}.$$

The null hypothesis is that the two groups are not different in terms of their rating responses for a specific factor; the alternative hypothesis is that there is a difference. The nine factors considered are:

- *risk* which denotes player's attitude towards risk;
- *maxPock* which represents a strategy that focuses on maximizing the Pocket lottery payoff;
- *maxPot* which is a strategy that focuses on maximizing the Pot/pool payoff;
- *safe* which is a strategy that focuses on keeping a player's money safe;
- *strMany* which is a strategy that focuses on applying many different strategies during the game;
- *random* which is a strategy of making random choices;
- *riskPock* which indicates risk perception towards Pocket/Pot;
- *complex* that denotes which investment option — complete or short — players found to be easier to make; and
- *bias* which represents if players found any social, moral or ethical bias in the Pot instructions.

Before proceeding with the hypothesis testing, it should be noted that forty two players participated in the complete version of the game and thirty seven participated in the short version. To perform the Friedman test, the number of players participating in both versions should be equal. Thus we reduced the number of players who participated in the complete version of the game down to thirty seven by performing a random sampling without replacement. Then, the Friedman tests were conducted to check the null hypotheses stated above. To minimize the possibility of sampling errors, we performed the Friedman test 1,000 times for each factor (each time with a new sample of thirty seven observations out of forty two for the complete version). As a result, a distribution of the Friedman test p-values for each factor considered was obtained; see Figure 5. Unfortunately, no rigorous statistical analysis is possible at

[13] The Friedman test is a nonparametric statistical test used to detect differences in treatments (in this case, game versions) across multiple mutually independent observations. The procedure involves ranking each observation within rows, then considering the sums of ranks by columns. The approximate distribution of the Friedman test statistic is χ^2 with degrees of freedom equal to the number of treatments used in the test minus one (Conover 1999).

this stage because the distribution of p-values, even if it exists, is unknown. For each factor considered, we computed the number of p-values that are less than or equal to 0.01, 0.05, and 0.10 (Table 3). This gives us an idea of how many Friedman tests out of 1,000 conducted for each factor rejected the null hypothesis for the three different significance levels. The conclusion of whether a particular number of rejections of the null hypothesis is large or 'normal' is subjective. Table 3 shows a suspiciously high number of p-values falling in the rejection region at 5% and 10% significance levels for the three factors: maxPock, maxPot, and complex. For maxPock, the number of p-values falling in the rejection region at 1% significance level seems to be very high. Thus, it can be suspected that players adopt different strategies regarding their Pocket and Pot investments in the complete and the short versions of the game. In addition, there is a chance that choice flexibility in terms of the number of tokens allowed to be invested is viewed differently by players who participated in different game versions.

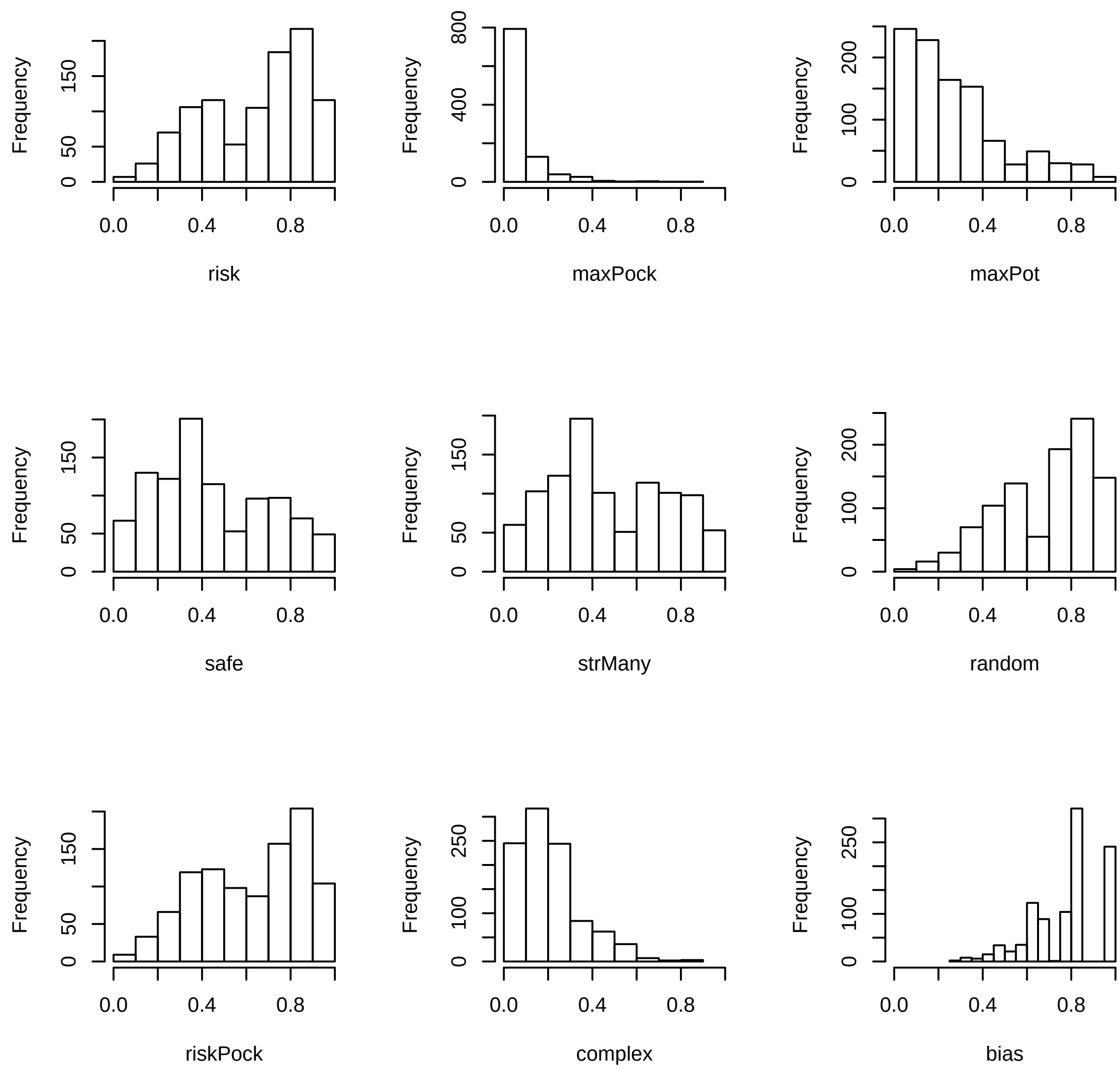
Although these findings are not based on a statistical test, they nonetheless have implications that need to be taken into consideration for further analysis. Analyzing hypothetical bias, the difference in investment behavior for the complete and the short versions of the game must be explained.

5.3 Heterogeneity in Rating Responses

Even though each player was given the same number of endowed tokens at the beginning of each round, we assume that people differ across their preferences, decision-making abilities, and other factors. Player homogeneity is not to be expected in an experiment like this; therefore, any inter-personal comparisons should be implemented with caution. In an attempt to control to some degree for the presence of such differences, each player was asked to complete a debriefing questionnaire after the game. This questionnaire contains several questions involving the use of a rating scale. Any inter-personal comparisons made using scale ratings should implement controls for scale usage heterogeneity.[14] No corrections for the scale usage heterogeneity or players' responses are needed when intra-personal comparisons are made.

[14] It has been known for a long time that respondents vary in their usage of the scale (Cronbach 1946; Lentz 1938). For instance, some subjects tend to use the upper portion of the scale, others use its lower or middle portions. These content-irrelevant factors of responding are referred to as 'scale usage heterogeneity'. The heterogeneity brings additional noise to models describing data and can significantly contaminate results.

Figure 5: Distributions of 1,000 Friedman Test P-values for Nine Factors



The Friedman test on the equality of factor effects for the complete version and the short version of the game is conducted 1,000 times. Nine factors are considered: 'risk' denotes players' attitude towards risk, 'maxPock' represents a strategy that focuses on maximizing the Pocket lottery payoff; 'maxPot' is a strategy that focuses on maximizing the Pot/pool payoff; 'safe' is a strategy that focuses on keeping a player's money safe; 'strMany' is a strategy that focuses on applying many different strategies during the game; 'random' is a strategy of making random choices; 'riskPock' indicates risk perception towards Pocket/Pot; 'complex' denotes which investment option — complete or short — players found to be easier to make; and 'bias' represents if players found any social, moral or ethical bias in the Pot instructions.

Table 3: Number of P-values No Greater Than 0.01, 0.05 and 0.10 for Nine Factors

Factor[a]	Number of p-values $\leq$ 0.01	0.05	0.10
risk	0	0	7
maxPock	193	567	793
maxPot	6	96	246
safe	0	22	67
strMany	1	25	60
random	0	0	4
riskPock	0	2	9
complex	1	90	245
bias	0	0	0

[a] Mnemonic 'risk' denotes players' attitude towards risk, 'maxPock' represents a strategy that focuses on maximizing the Pocket lottery payoff; 'maxPot' is a strategy that focuses on maximizing the Pot/pool payoff; 'safe' is a strategy that focuses on keeping a player's money safe; 'strMany' is a strategy that focuses on applying many different strategies during the game; 'random' is a strategy of making random choices; 'riskPock' indicates risk perception towards Pocket/Pot; 'complex' denotes which investment option — complete or short — players found to be easier to make; and 'bias' represents if players found any social, moral or ethical bias in the Pot instructions.

The standard approach to control for scale usage heterogeneity is to standardize responses within each subject:

$$\frac{R_{ij} - \bar{R}_i}{\sigma(R_i)}, \tag{1}$$

where R_{ij} is a rating response of i-th subject on j-th question, $\bar{R}_i$ and $\sigma(R_i)$ are the average rating response and the standard deviation of responses of i-th subject, respectively.

The rating scale for the risk attitude question (risk) differs from that for the other rating questions in the questionnaire. For this question, players are asked to rate their attitude towards risk on the 11-point attitude rating (Likert) scale. The lower end-point of 0 means that the respondent is extremely risk averse. The upper end-point of 10 means that the respondent is extremely risk loving. Players are asked to rate six more questions regarding strategies they adopted during the game. An 11-point Likert scale is used again. The lower end-point of 0 means that the respondent completely disagreed with the respective question. The upper end-point of 10 means that the respondent fully agreed with the respective question. Note that the rating scale for these questions differs from that for the risk attitude question. As a result, rating responses to the risk attitude question cannot be standardized, while ratings from the other rating questions can.

The distributions of rating responses for the risk attitude question in both game versions are presented in Figure 6. Although graphs in Figure 6 do not look alike, analysis of quantiles from the distribution of responses shows that players have similar risk attitudes in both complete and short versions of the game;

see Table 4.[15]

Table 4 shows that the complete version players are very similar to the short version players in terms of responses to the three questions: application of different strategies during the game (strMany), practicing random choices (random), and players' risk perception towards Pocket/Pot (riskPock). But, when comparing the other three questions — maximizing the Pocket (maxPock) and the Pot (maxPot) payoffs and keeping money safe (safe) — the short version players prefer maximizing the Pocket lottery payoff and keeping their money safe whereas, the complete version players prefer maximizing the Pot payoff. For the graphical representation of this conclusion, see Figures 7 and 8.

The distribution of the standardized rating responses for different questionnaire questions for the complete version and the short version of the game are presented in Figure 7 and Figure 8, respectively. While analyzing these figures, we are looking to see whether bimodality is present in responses. This knowledge will help us to validate a certain conjecture put forward by Champ et al. (1997) and Champ and Bishop (2001). Some authors suggested that hypothetical bias was manifested by an identifiable group of subjects, hence bimodality in their data. If bimodality is suspected in the rating responses provided by players participated in our experiment, it can be conjectured that the players can be split in two groups with different behavioral patterns. These behavioral differences can result in different patterns of rating responses which, in turn, may be causing the seeming bimodality.

Unfortunately, the exact distribution of the rating responses, even if it exists, is unknown. As a result, no statistically valid bimodality check is possible here. For the complete version (Figure 7), it appears that rating responses to three questions — maxPock, random, and riskPock — may exhibit bimodality. For the short version (Figure 8), rating responses to two questions — random and riskPock — may exhibit bimodality.

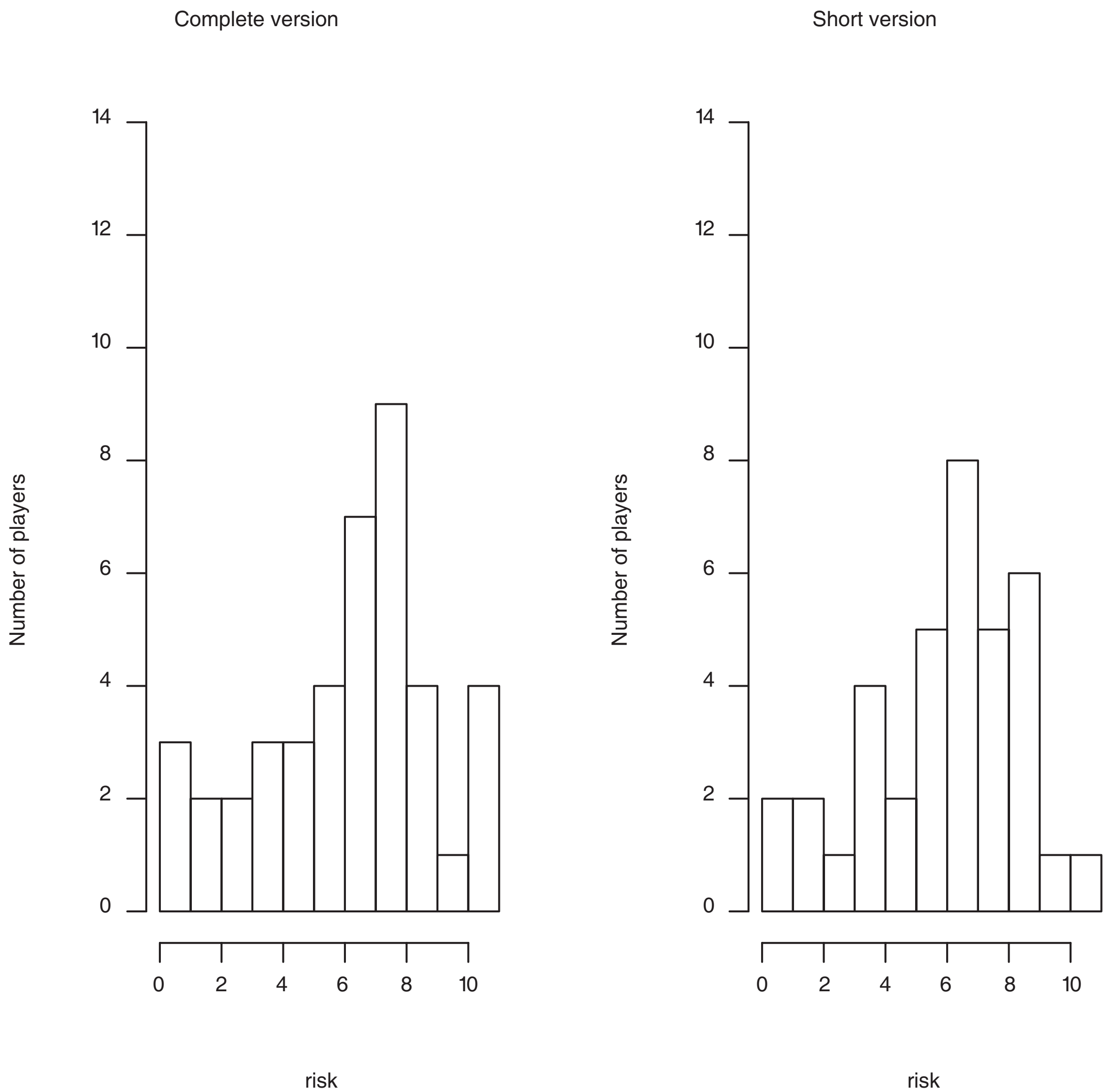
In addition to the scale usage heterogeneity, non-contingent responding should be addressed. Finally, although somewhat belatedly, we are interested in determining whether rating responses obtained from players are meaningful and so the analysis of it (as above) potentially useful. We need to be confident that there is not non-contingent responding (NCR) when NCR, as characterized by Baumgartner and Steenkamp (2001), is the tendency to respond to questions carelessly, randomly, or non-purposefully. There is no conventional way to check for NCR as it depends on the nature of data collected. To provide an analysis here, we chose to compute Kendall's concordance measure between the average difference in the Pot and Pocket investments and rating responses (Table 5).[16] If signs and significance of concordance coefficients are consistent with the logic of common sense (see discussion below), we can conclude that responses are meaningful and can be used for further analysis.

Results presented in Table 5 demonstrate that players' responses follow common sense: more risk-loving players as well as those attracted to Pocket lottery outcomes preferred to invest more in the Pocket than the Pot; those who wanted to keep their money safe as well as those who wanted to cash in on Pot outcomes preferred investing more in the Pot relative to the Pocket. The signs and significance of the concordance coefficients between hypothetical and real rounds show the consistency of players' behavior across round types. Real round coefficients for some factors exhibit a stronger statistical significance than those from hypothetical rounds, implying that players distinguished between the two types of rounds and may have behaved differently (e.g., players may have used hypothetical rounds to experiment in attempts

[15] Quantile is the percent (or fraction) of points below the given value. That is, the 30% (or 0.3) quantile is the point at which 30% percent of the data fall below and 70% fall above that value. Quantiles are points taken at regular intervals from the cumulative distribution function of a random variable.

[16] Kendall's concordance τ is a nonparametric association measure for a pair of variables, similar in concept to correlation (Conover 1999). The measure ranges between -1 and 1. Positive values (concordance) indicate that greater values of one variable (X) correspond to greater values of the other one (Y). Negative values (discordance) indicate the opposite. The advantage of using Kendall's τ over Pearson's correlation measure r is that the distribution of r depends on the bivariate distribution function of (X,Y). Therefore r has no value as a test statistic in nonparametric tests unless the distribution of (X,Y) is known. Kendall's τ is based on the order (ranks) of the observations rather than the numbers themselves, and the distribution of the measure does not depend on the distribution of (X,Y). Additionally, correlation is a measure of linear dependence, whereas Kendall's concordance only measures association, no matter what form it takes.

Figure 6: Distribution of Non-Standardized Rating Responses for the Risk Attitude Question



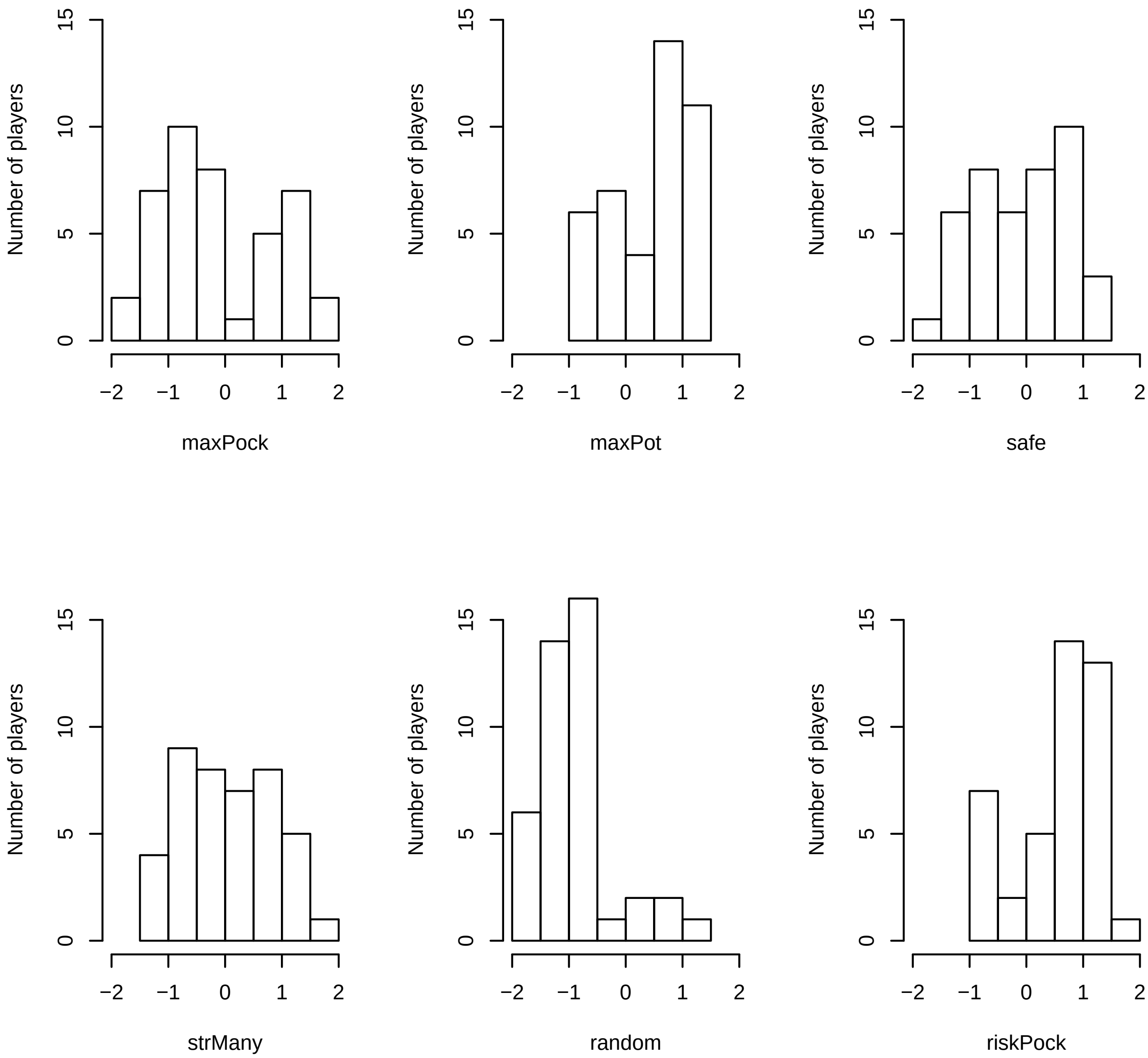
An 11-point attitude rating (Likert) scale is the horizontal axis (0 means that the respondent is extremely risk averse; 10 means that the respondent is extremely risk loving). Numbers of players are shown on the vertical axis. The total number of players is 42 for the complete version and 37 for the short version.

Table 4: Quantile Analysis for Seven Questionnaire Variables

Factors	Game version	0% min	10%	20%	30%	40%	50% median	60%	70%	80%	90%	100% max
risk	complete	0.0	1.1	3.0	4.3	5.4	6.0	7.0	7.0	7.8	8.9	10.0
	short	0.0	1.6	3.0	4.8	5.0	6.0	6.0	7.0	7.8	8.0	10.0
maxPock	complete	−1.98	−1.26	−1.03	−0.85	−0.71	−0.35	−0.12	0.62	1.07	1.19	1.95
	short	−1.55	−1.01	−0.75	−0.26	0.27	0.46	0.60	0.75	0.91	1.29	1.85
maxPot	complete	−0.90	−0.69	−0.25	0.02	0.47	0.85	0.91	0.97	1.04	1.17	1.46
	short	−1.01	−0.81	−0.58	−0.43	−0.22	0.06	0.39	0.59	0.80	1.09	1.97
safe	complete	−1.69	−1.10	−0.89	−0.67	−0.38	0.06	0.35	0.52	0.80	0.91	1.28
	short	−1.54	−1.12	−1.02	−0.85	−0.64	−0.30	−0.13	0.29	0.53	0.89	1.77
strMany	complete	−1.45	−0.98	−0.84	−0.56	−0.27	0.00	0.25	0.56	0.68	1.02	1.59
	short	−1.15	−0.70	−0.64	−0.42	−0.24	−0.10	0.20	0.65	0.78	0.95	1.31
random	complete	−1.70	−1.52	−1.25	−1.14	−1.06	−0.98	−0.89	−0.81	−0.69	0.19	1.20
	short	−1.99	−1.65	−1.35	−1.23	−1.09	−0.97	−0.84	−0.69	−0.16	1.10	2.04
riskPock	complete	−0.92	−0.75	−0.22	0.44	0.71	0.89	0.94	1.01	1.09	1.19	1.52
	short	−1.81	−0.79	−0.45	0.14	0.60	0.77	0.96	1.06	1.11	1.24	1.76

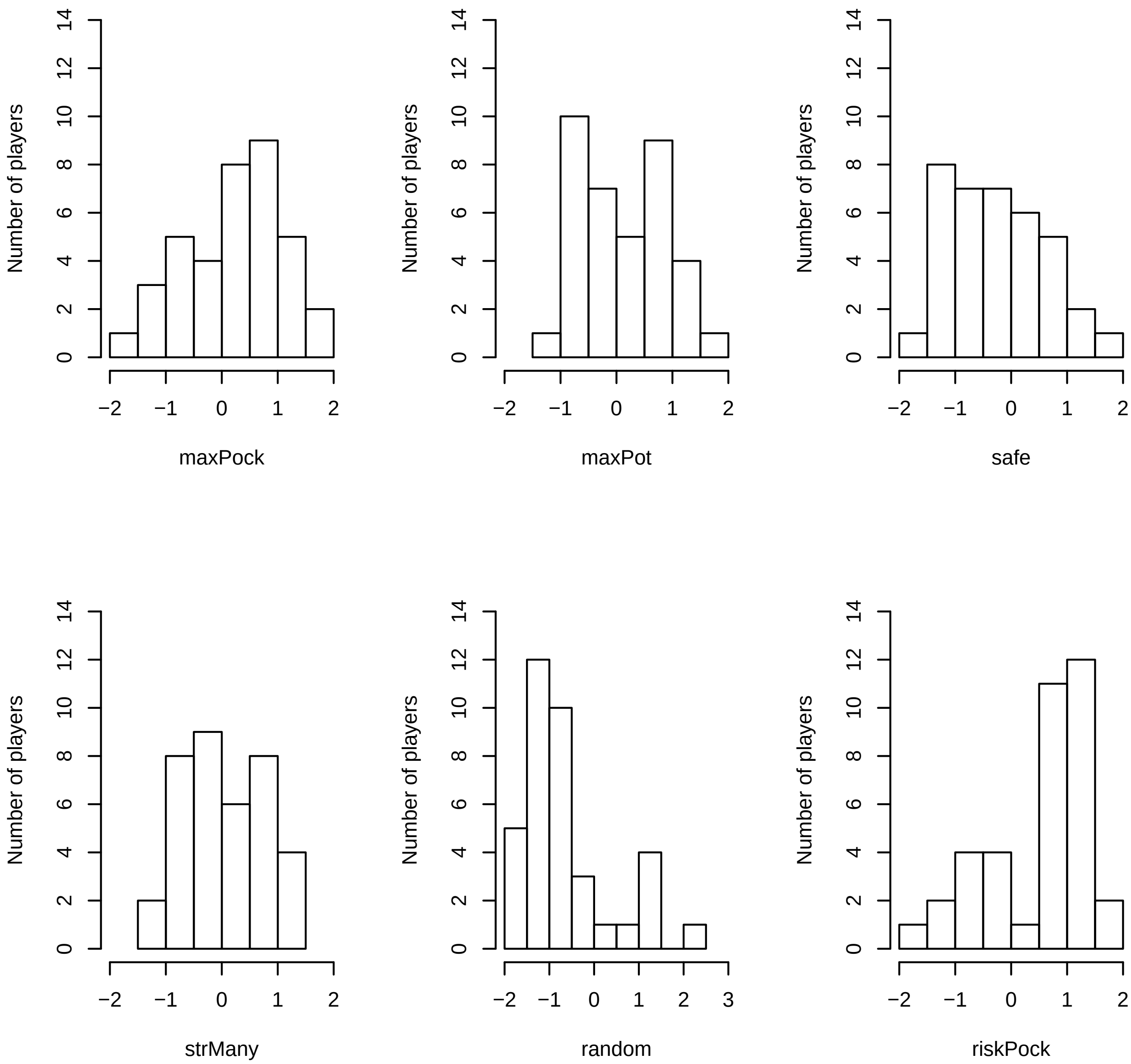
‘risk’ denotes players’ attitude towards risk; ‘maxPock’ represents a strategy that focuses on maximizing the Pocket lottery payoff; ‘maxPot’ is a strategy that focuses on maximizing the Pot/pool payoff; ‘safe’ is a strategy that focuses on keeping a player’s money safe; ‘strMany’ is a strategy that focuses on applying many different strategies during the game; ‘random’ is a strategy of making random choices; and ‘riskPock’ indicates risk perception towards Pocket/Pot. Non-standardized scale for the $risk$ factor and standardized scales for the other factors are used here.

Figure 7: Distribution of Standardized Rating Responses for Six Behavioral Questionnaire Questions, Complete Version



Standardized 11-point attitude rating (Likert) scale is the horizontal axis (the lower end-point means that the respondent completely disagreed with the respective question; the upper end-point means that the respondent fully agreed with the respective question). Numbers of players are shown on the vertical axis. The total number of players is 42. ‘maxPock’ represents a strategy that focuses on maximizing the Pocket lottery payoff; ‘maxPot’ is a strategy that focuses on maximizing the Pot/pool payoff; ‘safe’ is a strategy that focuses on keeping a player’s money safe; ‘strMany’ is a strategy that focuses on applying many different strategies during the game; ‘random’ is a strategy of making random choices; ‘riskPock’ indicates risk perception towards Pocket/Pot.

Figure 8: Distribution of Standardized Rating Responses for Six Behavioral Questionnaire Questions, Short Version



Standardized 11-point attitude rating (Likert) scale is the horizontal axis (the lower end-point means that the respondent completely disagreed with the respective question; the upper end-point means that the respondent fully agreed with the respective question). Numbers of players are shown on the vertical axis. The total number of players is 37. ‘maxPock’ represents a strategy that focuses on maximizing the Pocket lottery payoff; ‘maxPot’ is a strategy that focuses on maximizing the Pot/pool payoff; ‘safe’ is a strategy that focuses on keeping a player’s money safe; ‘strMany’ is a strategy that focuses on applying many different strategies during the game; ‘random’ is a strategy of making random choices; ‘riskPock’ indicates risk perception towards Pocket/Pot.

Table 5: Non-Contingent Responding. Kendall's Concordance Coefficients Between Average (Pot−Pocket) Investments and Rating Responses

	Complete Version		Short Version	
Factors	Hypothetical rounds	Real rounds	Hypothetical rounds	Real rounds
risk	−0.2928***	−0.3014***	−0.1621	−0.0736
	(0.0097)	(0.0079)	(0.2021)	(0.5567)
maxPock	−0.3885***	−0.5388***	−0.1354	−0.3790***
	(0.0003)	(0.0001)	(0.2639)	(0.0015)
maxPot	0.2132**	0.4079***	0.0999	0.3411***
	(0.0495)	(0.0001)	(0.4096)	(0.0043)
safe	0.2321**	0.3629***	0.1805	0.1106
	(0.0325)	(0.0008)	(0.1363)	(0.3554)
strMany	−0.1577	−0.2505**	0.0999	0.1390
	(0.1458)	(0.0212)	(0.4096)	(0.2453)
random	0.0436	−0.1099	0.0451	−0.0853
	(0.6880)	(0.3122)	(0.7096)	(0.4759)
riskPock	0.1342	0.1630	−0.2869**	−0.0663
	(0.2160)	(0.1337)	(0.0179)	(0.5792)
bias	−0.1793	−0.0118	−0.0703	0.0511
	(0.1699)	(0.9284)	(0.6296)	(0.7223)

'risk' denotes players' attitude towards risk; 'maxPock' represents a strategy that focuses on maximizing the Pocket lottery payoff; 'maxPot' is a strategy that focuses on maximizing the Pot/pool payoff; 'safe' is a strategy that focuses on keeping a player's money safe; 'strMany' is a strategy that focuses on applying many different strategies during the game; 'random' is a strategy of making random choices; 'riskPock' indicates risk perception towards Pocket/Pot; and 'bias' represents if players found any social, moral or ethical bias in the Pot instructions. p-values are in parentheses. ***, **, and * denote estimates significant at 1%, 5%, and 10% significance levels, respectively.

to learn about how the game worked or other players behaved). One possible explanation of the many insignificant concordance coefficients in the short version is the insufficient variation in the Pocket/Pot investments due to the limited number of possible outcomes. The overall result indicates that allocations are contingent on rating responses, that is, players seem to have answered the questionnaire questions carefully and meaningfully.

5.4 Hypothetical Bias

Result 1 *A statistically significant negative hypothetical bias is found between the first hypothetical round and the first real round. The bias in the subsequent second and third hypothetical/real pairs is not statistically significant. The hypothetical bias magnitude oscillates around zero and converges to zero as the experiment progresses.*

Given the data collected in the experiment, hypothetical bias for the Pot contributions can be measured in two ways: as an absolute measure and as a relative measure that is relative to the Pocket investments. The absolute hypothetical bias of the i-th player (AHB_i) is the following.

$$AHB_i = Pot_i^H - Pot_i^R, \tag{2}$$

where H and R denote hypothetical and real rounds, respectively, and Pot stands for the Pot contributions.

The relative hypothetical bias of the i-th player is the hypothetical bias for the difference of Pot and Pocket investments (RHB_i):

$$RHB_i = (\text{Pot} - \text{Pocket})_i^H - (\text{Pot} - \text{Pocket})_i^R, \tag{3}$$

where H and R indicate hypothetical and real rounds, respectively, and (Pot$-$Pocket) is the difference between the Pot and Pocket investments.

We conducted analysis using both measures of hypothetical bias and find that they always produce similar results (see Figure 9 as an example). The relative measure of hypothetical bias seems more natural because it shows how substitution between public (the Pot) and private (the Pocket) investments changes from hypothetical to real choices. Thus, results for the RHB measure are presented in the paper; results pertaining to the AHB measure are presented whenever they are crucial to one's understanding of a task under consideration or are different compared to the RHB measure. The choice in favor of the difference between the Pot and Pocket investments rather than their ratio in Equation (3) can be explained by the fact that both Pot and Pocket investments can take zero values.

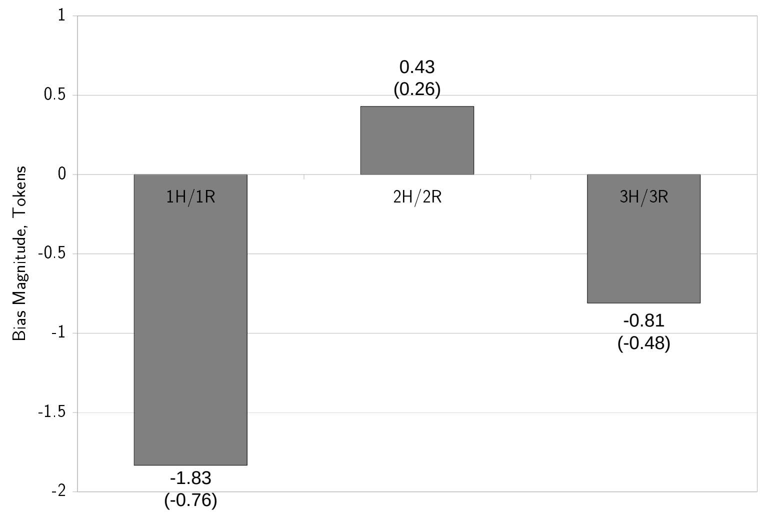
Evidence in support of Result 1 is presented in Figure 9 and Table 6. In Figure 9, the presence of negative hypothetical bias for the first pair of hypothetical/real rounds in the complete as well as the short version of the game is observed. The bias sign is reversed for the second pair of rounds in both versions. For the third pair of hypothetical/real rounds, hypothetical bias disappears in the short version but its negative sign is restored in the complete version. In other words, it looks like hypothetical bias oscillates around zero and converges to some steady-state (possibly zero). This observation is true for both measures of hypothetical bias as well as both versions of the game. The resemblance between AHB and RHB magnitudes and signs tells us that the presence/absence of hypothetical bias is robust to the change in the hypothetical bias measure.

To check the statistical significance of the bias detected for each pair of rounds, the following null and alternative hypotheses were formulated and the former tested using the Friedman test:

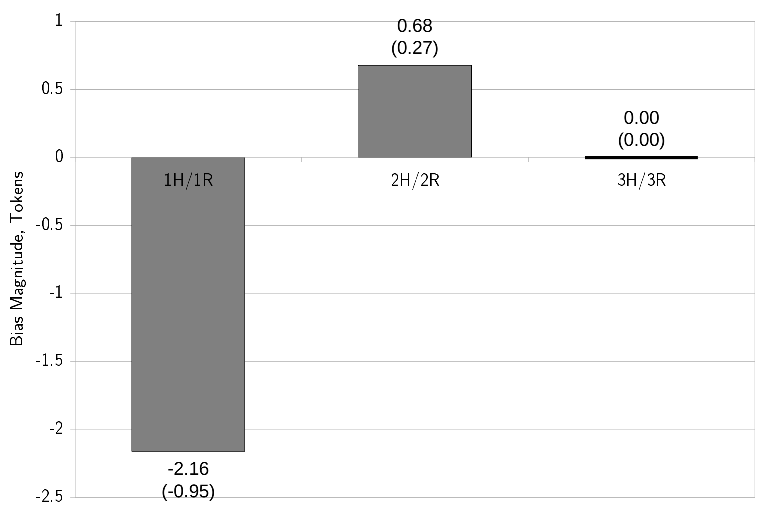
$$\text{H}_0 : \text{Hypothetical (Pot}-\text{Pocket)} = \text{Real (Pot}-\text{Pocket)}$$
$$\text{versus}$$
$$\text{H}_1 : \text{Hypothetical (Pot}-\text{Pocket)} \neq \text{Real (Pot}-\text{Pocket)}.$$

Figure 9: Relative and Absolute Hypothetical Bias Across Round Pairs

(a) Complete Version



(b) Short version



Shown numbers are relative hypothetical bias values averaged across players. Numbers in parenthesis are absolute hypothetical bias values averaged across players. '1H/1R', '2H/2R', and '3H/3R' denote first, second, and third pairs of (H)ypothetical/(R)eal rounds, respectively.

Table 6: Test for the Difference between Hypothetical (Pot−Pocket) Investments and Real (Pot−Pocket) Investments Across Round Pairs

	Complete Version			Short Version		
Factors	Hypoth. Pot−Pocket	Real Pot−Pocket	Friedman Statistic[a]	Hypoth. Pot−Pocket	Real Pot−Pocket	Friedman Statistic
First hypothetical and first real	1.310	3.143	3.000 (0.0833)	−1.081	1.081	5.000 (0.0254)
Second hypothetical and second real	2.428	2.000	0.571 (0.4497)	−1.486	−2.162	0.154 (0.6949)
Third hypothetical and third real	0.833	1.643	0.500 (0.4795)	−2.162	−2.162	0.048 (0.8273)

[a] The Friedman test p-values are in parentheses.

The null hypothesis is that there is no difference between (Pot−Pocket) investments in hypothetical and real rounds; the alternative hypothesis is that there is a difference.[17]

Results of the Friedman test are presented in Table 6. These show the presence of a statistically significant negative hypothetical bias only for the first pair of hypothetical/real rounds in both versions of the game. RHB is significant at the 10% significance level for the complete version whereas, for the short version, it is significant at the 5% significance level. For the second and third round pairs, the bias is not statistically significant.

The above findings trigger two questions. First, why does statistically significant hypothetical bias only exist for the first hypothetical/real round pair and then disappear in the consecutive rounds tending to converge towards zero (Figure 9)? Second, why is the bias negative in the first pair? That is, why do players systematically put more tokens in the Pot in the first real round relative to those in the first hypothetical round?

Result 2 *The possible explanation of the presence of hypothetical bias based on players' limited experience with the game in the first pair of rounds does not seem to be supported by the data. This result is conditional on the choice of variance/standard deviation as an experience measure.*

One possible explanation for the absence of any statistically significant hypothetical bias in the second and third hypothetical/real round pairs is that players might have gained some experience with the game, which affected their token allocations. A similar explanation was offered in an experiment conducted by McClelland, Schulze, and Coursey (1993). This study involves risk and purchases of insurance polices and does not refer to public good contributions. The authors found that inexperienced hypothetical bids can underestimate actual auction bids (hence a negative hypothetical bias), whereas experienced hypothetical bids can either overestimate actual bids or be equal (that is, hypothetical bias is positive with those players or there is no bias at all). Put otherwise, practice with the auction mechanism and experience with risk may eliminate hypothetical bias. The reader is reminded that, in our experiment, no results are reported to players until both hypothetical/real or real/hypothetical rounds are completed in the first pair of rounds (where hypothetical bias has been detected). Players' experience thus has the associated learning limit.

Since the experiment setup did not contain any facility to track changes in experience (learning) of players as they moved through the game, there is a need to come up with a proxy measure for the experience/learning factor. Tumer and Wolpert (2004) and Wolpert and Tumer (2001) suggest a 'learnability' property that measures the sensitivity of an agent's utility to the agent's own strategy versus the strategies of others. A low learnability (characteristic to inexperienced players) would result in a smaller role of one's own decision criteria and a high sensitivity to the perceived strategies of others. In a dynamic setting, this should arguably cause higher volatility in choices made by that agent. On the other hand, experimentation with one's own preferences should also lead to an agent's low learnability, now because of the smaller role of the utility part. In either case, variance (or standard deviation) as a measure of dispersion in the group of players should then work as a reasonable measure of players' experience.

To explore the supposition that players, as a group, became more game experienced as they moved through the game, the following two hypotheses were put forward:

$$\mathrm{H}_0 : \text{Hypothetical } \sigma_{12} = \text{Hypothetical } \sigma_{23} \quad \text{versus} \quad \mathrm{H}_1 : \text{Hypothetical } \sigma_{12} \neq \text{Hypothetical } \sigma_{23};$$

$$\mathrm{H}_0 : \text{Real } \sigma_{12} = \text{Real } \sigma_{23} \quad \text{versus} \quad \mathrm{H}_1 : \text{Real } \sigma_{12} \neq \text{Real } \sigma_{23}.$$

The null hypothesis is that there is no difference between the standard deviation (σ) of the respective factor calculated for the first and the second hypothetical rounds and that calculated for the second and

[17] In addition, the following null hypothesis was tested by the Friedman test:

$$\mathrm{H}_0 : \text{Hypothetical Pot} = \text{Real Pot} \quad \text{versus} \quad \mathrm{H}_1 : \text{Hypothetical Pot} \neq \text{Real Pot}.$$

The null hypothesis is that there is no difference between Pot contributions in hypothetical and real rounds; the alternative hypothesis is that there is a difference. The test results (not presented here) are in line with the results presented in Table 6.

the third hypothetical rounds; the alternative hypothesis is that there is a difference (Friedman test). The second null hypothesis is that there is no difference between the the respective factor's σ calculated for the first and second real rounds and that calculated for the second and the third real rounds; the alternative hypothesis is that there is a difference. Three factors are considered: Pot contributions, Pocket investments, and the difference between the Pot and Pocket investments. The Friedman test is a pair-wise test, thus, for each player, σ_{12} is computed between investments made in the first and the second rounds by a player. Similarly, σ_{23} is computed between investments made in the second and the third rounds by each player. Results, presented in Table 7, suggest that the null hypothesis for all tests should not be rejected. In other words, there is no evidence that hypothetical as well as real standard deviations of the three factors considered diminish as players proceed through the game. Thus, the conjecture that players become more focused during the game is not supported by the data.[18] On the other hand, it is possible that the variance measure does not really capture the idea of focusing and learning as players proceed through the game.

Table 7: Test for Differences in Standard Deviations in Consecutive Hypothetical and Real Rounds for Three Factors

Factors	Complete Version St.dev. 1:2[a]	Complete Version St.dev. 2:3[b]	Complete Version Friedman Statistic[c]	Short Version St.dev. 1:2	Short Version St.dev. 2:3	Short Version Friedman Statistic
Pot Contributions:						
Hypothetical rounds	1.58	1.46	0.8621 (0.3532)	2.58	2.39	0.1667 (0.6831)
Real rounds	1.09	0.98	0.0400 (0.8415)	1.53	1.82	0.4737 (0.4913)
Pocket Investments:						
Hypothetical rounds	1.56	1.58	0.0345 (0.8527)	2.87	2.68	0.4286 (0.5127)
Real rounds	1.09	0.96	0.0526 (0.8185)	1.53	1.62	0.0000 (1.0000)
Pot−Pocket:						
Hypothetical rounds	3.08	3.05	0.1250 (0.7237)	5.45	5.06	0.1667 (0.6831)
Real rounds	2.09	1.90	0.3913 (0.5316)	3.06	3.44	0.0476 (0.8273)

[a] Standard deviation is computed between the first and the second rounds.
[b] Standard deviation is computed between the second and the third rounds.
[c] The Friedman test p-values are in parentheses.

[18] In addition, it was tested if the standard deviation of the considered factors decreases as players proceed through the game regardless of the round type. That is, the following hypothesis was tested:

$$\mathrm{H}_0 : \ \sigma_i = \ \sigma_{i+1} \quad \text{versus} \quad \mathrm{H}_1 : \ \sigma_i \neq \ \sigma_{i+1},$$

where $i \ \in [1, 6]$ and $i + 1$ indicate the current and the next round, respectively, regardless of whether it is a hypothetical or a real round. Test results (not presented) do not support the conjecture that players, as a group, become more focused during the game.

To test whether players might have systematically experimented with investments in hypothetical rounds and were more careful with those in real rounds, the following null hypothesis is tested:

$$\mathrm{H}_0 : \text{Hypothetical } \sigma = \text{Real } \sigma \quad \text{versus} \quad \mathrm{H}_1 : \text{Hypothetical } \sigma \neq \text{Real } \sigma.$$

This null hypothesis is that there is no difference between the respective factor's σ in hypothetical and real rounds; the alternative hypothesis is that there is a difference (Friedman test). Three factors are considered: Pot contributions, Pocket investments, and the difference between the Pot and Pocket investments. Results, presented in Table 8, suggest that the standard deviation of both Pot and Pocket investments as well as their difference is greater in hypothetical rounds than that in real rounds. The implication is that players, as a group, varied their investment decisions more in hypothetical rounds than real rounds throughout the game. It can be conjectured that players experimented more with the Pot/Pocket investments in hypothetical rounds compared to real rounds, including situations when they tried to 'outsmart' or mislead other players; or players simply focused less in hypothetical rounds making more random choices.

Table 8: Test for Differences in Standard Deviation in Hypothetical and Real Rounds for Three Factors

	Complete Version			Short Version		
Factors	Hypoth. st.dev.[a]	Real st.dev.[b]	Friedman Statistic[c]	Hypoth. st.dev.	Real st.dev.	Friedman Statistic
Pot Contributions	1.88	1.23	10.939 (0.0009)	2.97	2.04	8.048 (0.0046)
Pocket Investments	2.03	1.21	15.114 (0.0001)	3.21	2.09	9.800 (0.0017)
Pot−Pocket	3.79	2.33	16.000 (0.0001)	6.15	4.03	8.167 (0.0043)

[a] Standard deviation is computed across three hypothetical rounds.
[b] Standard deviation is computed across three real rounds.
[c] The Friedman test p-values are in parentheses.

To summarize, the statistical tests on standard deviations conducted in this section give an impression that players, in both versions of the game, chose statistically different Pot/Pocket investment strategies in hypothetical and real rounds but did not significantly change these strategies throughout the game.

5.5 Exploratory Factor Analysis for Hypothetical Bias

Result 3 *The grouping of players by their risk perception toward the Pocket/Pot reveals that those who found the Pot to be riskier than the Pocket demonstrate a significant hypothetical bias regardless of what round of the game being played. Those players who found the Pocket to be riskier than the Pot did not demonstrate any hypothetical bias at all. In addition, the hypothesis that people exhibit hypothetical bias as a result of purchasing moral satisfaction and/or of compliance bias does not seem to be fully supported by the data.*

In order to find factors responsible for the presence of hypothetical bias in the first hypothetical/real pair of rounds and its absence from subsequent round pairs, we calculate Kendall's pair-wise concordance coefficients for the relative hypothetical bias (Equation 3) and socio-demographic variables as well as for the absolute hypothetical bias (Equation 2) and socio-demographic variables for the complete and short versions. Results presented in Table 9 indicate that, for the complete version of the game, the factor of part-time job and the frequency of volunteering are significantly concordant with both relative and absolute hypothetical bias for the third pair of hypothetical/real rounds. Concordance results for the short version are presented in Table 10. The table shows that the frequency of volunteering is significantly concordant with both relative and absolute hypothetical bias in the second pair of hypothetical/real rounds, and living expenditures is significantly concordant with both measures of the bias in the third pair of hypothetical/real rounds. No concordance coefficients are significant for the first pair of hypothetical/real rounds for both game versions, which means little can be inferred from these results that would help to explain the presence of hypothetical bias in the first pair of rounds and its absence later on.

To investigate further the occurrence of hypothetical bias, concordance between the relative or absolute hypothetical bias and scale ratings of attitude towards risk and different strategies (as were reported by players in the exit questionnaire) was computed for both complete and short versions. For the complete version (Table 9), two factors — the risk perception towards Pocket/Pot (riskPock) and the presence of social, moral, or ethical bias (bias) — are significantly concordant with both measures of hypothetical bias in the first hypothetical/real pair and marginally concordant or not concordant with the bias in the later rounds. Kendall's concordance coefficients between these factors and any measure of hypothetical bias are not statistically significant for the short version of the game (Table 10).

Notice that the significance of concordance coefficients for the short version is measure-dependent, that is, there are a few coefficients — expenditures, risk, maxPock, and maxPot — that are significant for the relative hypothetical bias measure but are not for the absolute measure. Conventionally, whenever implications of some results depend on the measure chosen to obtain those, the results have limited reliability. The difference in significance of the concordance coefficients between the complete version and the short version of the game, observed in Tables 9 and 10, may be due to the limited number of ways that tokens can be allocated in the short version. For this reason, the discussion below is based only on the more robust results obtained for the complete version.

Champ et al. (1997) and Champ and Bishop (2001) suggest that hypothetical bias is manifested by an identifiable minority of subjects. Following upon this conjecture, we wanted to test whether complete version players grouped by their risk perception towards Pocket/Pot would exhibit different behavior in terms of the hypothetical bias presence within each group. Additionally, we conducted the same analysis for the players grouped by the presence of social, moral, or ethical bias.

After grouping the complete version players by their risk perception with regard to Pot/Pocket (riskPock), we checked to see if there was any difference in investment behavior between hypothetical and real rounds within each group. Specifically, the following null hypothesis is tested:

$$\mathrm{H}_0 : \text{Hypothetical (Pot}-\text{Pocket)}_i^k = \text{Real (Pot}-\text{Pocket)}_i^k$$

$$\text{versus}$$

$$\mathrm{H}_1 : \text{Hypothetical (Pot}-\text{Pocket)}_i^k \neq \text{Real (Pot}-\text{Pocket)}_i^k,$$

where $k \in [1, 2]$ indicates two groups of players: the first group consists of players who found the Pot to be riskier than the Pocket, while players in the second group found the Pocket to be riskier than the Pot; $i \in [1, 3]$ denotes the round pairs' number; and (Pot$-$Pocket) is the difference in Pot contributions relative to Pocket investments.

The null hypothesis is that there is no difference between the (Pot$-$Pocket) investments in hypothetical and real rounds within a group; the alternative hypothesis is that there is a difference (Friedman test). Table 11 contains Friedman's test results for the difference in behavior across the three pairs of hypothetical/real rounds for each group. Results indicate that those who found the Pot to be riskier than the Pocket (eight players out of forty two), exhibit highly significant negative relative hypothetical bias in the first hypothetical/real rounds and significant or marginally significant positive relative hypothetical bias in the second and third pairs of hypothetical/real rounds. Those who found the Pocket to be riskier than

Table 9: Kendall's Concordance Coefficients Between Relative/Absolute Hypothetical Bias and Respondent Characteristics Across Round Pairs, Complete Version

Factors[b]	RHB[a]			AHB		
	1H/1R	2H/2R	3H/3R	1H/1R	2H/2R	3H/3R
Gender	0.1373	−0.0174	0.0342	0.1193	−0.0441	0.0417
	(0.3141)	(0.8978)	(0.7992)	(0.3923)	(0.7475)	(0.7594)
Student	0.1172	−0.0181	0.0711	0.0857	−0.0137	0.1036
	(0.3592)	(0.8870)	(0.5732)	(0.5120)	(0.9148)	(0.4166)
Expenditures	0.0000	−0.1450	0.0136	0.1090	−0.1471	0.0192
	(1.0000)	(0.2339)	(0.9105)	(0.3831)	(0.2322)	(0.8746)
Part-time job	0.0334	−0.1741	−0.2877**	0.0000	−0.1713	−0.2587*
	(0.8063)	(0.1989)	(0.0326)	(1.0000)	(0.2116)	(0.0573)
Volunteering	−0.0636	−0.2049	−0.2858**	−0.0102	−0.2357*	−0.2849**
	(0.6251)	(0.1133)	(0.0263)	(0.9389)	(0.0718)	(0.0284)
Game Experience	−0.0230	0.0768	−0.0699	0.0330	0.1096	−0.0581
	(0.8525)	(0.5322)	(0.5668)	(0.7942)	(0.3780)	(0.6375)
risk	−0.0477	0.0590	0.0671	0.0028	0.0745	0.0445
	(0.6858)	(0.6148)	(0.5647)	(0.9817)	(0.5299)	(0.7056)
maxPock	0.0252	0.1444	0.1530	−0.0144	0.1855	0.1340
	(0.8239)	(0.1990)	(0.1705)	(0.9006)	(0.1026)	(0.2348)
maxPot	−0.0164	−0.2553**	−0.1225	0.0354	−0.2652**	−0.0981
	(0.8850)	(0.0232)	(0.2729)	(0.7592)	(0.0197)	(0.3848)
safe	−0.1598	−0.1121	−0.1556	−0.1496	−0.1048	−0.1391
	(0.1576)	(0.3189)	(0.1638)	(0.1955)	(0.3566)	(0.2179)
strMany	−0.0126	0.1070	0.1090	−0.0171	0.1098	0.1489
	(0.9114)	(0.3409)	(0.3292)	(0.8827)	(0.3339)	(0.1868)
random	−0.0214	0.2515**	0.1850*	−0.0105	0.2640**	0.1689
	(0.8500)	(0.0253)	(0.0978)	(0.9276)	(0.0203)	(0.1346)
riskPock	0.3472***	−0.1867*	−0.1163	0.3109***	−0.2007*	−0.1539
	(0.0021)	(0.0967)	(0.2976)	(0.0071)	(0.0774)	(0.1725)
bias	−0.2256*	−0.1116	−0.0348	−0.2651*	−0.0844	−0.0177
	(0.0980)	(0.4102)	(0.7957)	(0.0573)	(0.5380)	(0.8967)

[a] RHB and AHB are relative and absolute hypothetical bias, respectively. 1H/1R, 2H/2R, and 3H/3R denote first, second, and third pairs of (H)ypothetical/(R)eal rounds, respectively.
[b] Test p-values are in parenthesis. ***, **, and * denote that estimates are significant at 1%, 5%, and 10% significance level, respectively.

Table 10: Kendall's Concordance Coefficients Between Relative/Absolute Hypothetical Bias and Respondent Characteristics Across Round Pairs, Short Version

Factors[b]	RHB[a]			AHB		
	1H/1R	2H/2R	3H/3R	1H/1R	2H/2R	3H/3R
Gender	−0.1939	−0.1855	−0.1201	−0.1544	−0.1928	−0.1105
	(0.2057)	(0.2118)	(0.4244)	(0.3349)	(0.2015)	(0.4724)
Student	−0.0022	0.1196	0.1490	0.0434	0.1144	0.1383
	(0.9880)	(0.3929)	(0.2926)	(0.7736)	(0.4210)	(0.3397)
Expenditures	−0.2085	0.2185*	0.3414**	−0.1453	0.1996	0.3933***
	(0.1310)	(0.1027)	(0.0117)	(0.3133)	(0.1422)	(0.0045)
Part-time job	0.0200	−0.1096	−0.1039	0.0305	−0.1221	−0.1087
	(0.8961)	(0.4605)	(0.4895)	(0.8489)	(0.4185)	(0.4799)
Volunteering	−0.0444	0.3317**	−0.0249	−0.0908	0.3427**	−0.0473
	(0.7635)	(0.0206)	(0.8638)	(0.5564)	(0.0186)	(0.7499)
Game Experience	−0.0290	0.0387	0.1984	−0.0207	0.0417	0.1929
	(0.8365)	(0.7767)	(0.1507)	(0.8880)	(0.7635)	(0.1721)
risk	−0.1147	−0.0263	0.2326*	−0.0855	−0.0216	0.2034
	(0.3887)	(0.8485)	(0.0750)	(0.5383)	(0.8691)	(0.1280)
maxPock	0.2273*	0.2800**	0.0721	0.1674	0.2806**	0.0682
	(0.0736)	(0.0232)	(0.5635)	(0.2068)	(0.0251)	(0.5929)
maxPot	−0.2060*	−0.2435**	−0.1304	−0.1517	−0.2568**	−0.1436
	(0.1050)	(0.0483)	(0.2959)	(0.2528)	(0.0403)	(0.2603)
safe	−0.0924	0.0149	0.0172	−0.0256	0.0323	−0.0108
	(0.4674)	(0.9038)	(0.8906)	(0.8469)	(0.7964)	(0.9327)
strMany	0.0568	−0.0215	−0.0995	0.1477	−0.0017	−0.1221
	(0.6547)	(0.8614)	(0.4250)	(0.2653)	(0.9892)	(0.3387)
random	0.0391	0.0348	0.1510	−0.0650	0.0017	0.1759
	(0.7585)	(0.7779)	(0.2261)	(0.6240)	(0.9892)	(0.1679)
riskPock	0.0355	−0.1309	−0.0377	0.0098	−0.1276	−0.0323
	(0.7798)	(0.2886)	(0.7622)	(0.9408)	(0.3085)	(0.8000)
bias	0.1925	−0.1749	−0.1763	0.1747	−0.1700	−0.1643
	(0.2089)	(0.2390)	(0.2407)	(0.2751)	(0.2600)	(0.2856)

[a] RHB and AHB are relative and absolute hypothetical bias, respectively. 1H/1R, 2H/2R, and 3H/3R denote first, second, and third pairs of (H)ypothetical/(R)eal rounds, respectively.
[b] Test p-values are in parenthesis. ***, **, and * denote that estimates are significant at 1%, 5%, and 10% significance level, respectively.

the Pot did not demonstrate any statistically significant hypothetical bias in any pair of hypothetical/real rounds.

Note the peculiarities in behavior of those players who find the Pot to be riskier than the Pocket. Even though these players believe that the Pot is riskier, they put 1.25 more tokens in the Pot than in the Pocket in the first real round. In the first hypothetical round, this group of players behaves in line with their assessment of the Pot riskiness: they invest 4.625 more tokens in the Pocket than in the Pot. In the second and third pairs of round, these players put more in the Pocket than in the Pot in real rounds and more in the Pot than in the Pocket in hypothetical rounds. Those players who find the Pocket to be riskier than the Pot, on the other hand, behave consistently throughout the game by investing more in the Pot than in the Pocket.

Table 11: Pocket/Pot Riskiness Grouping. Test for the Difference between Hypothetical and Real (Pot−Pocket) Investments, Complete Version

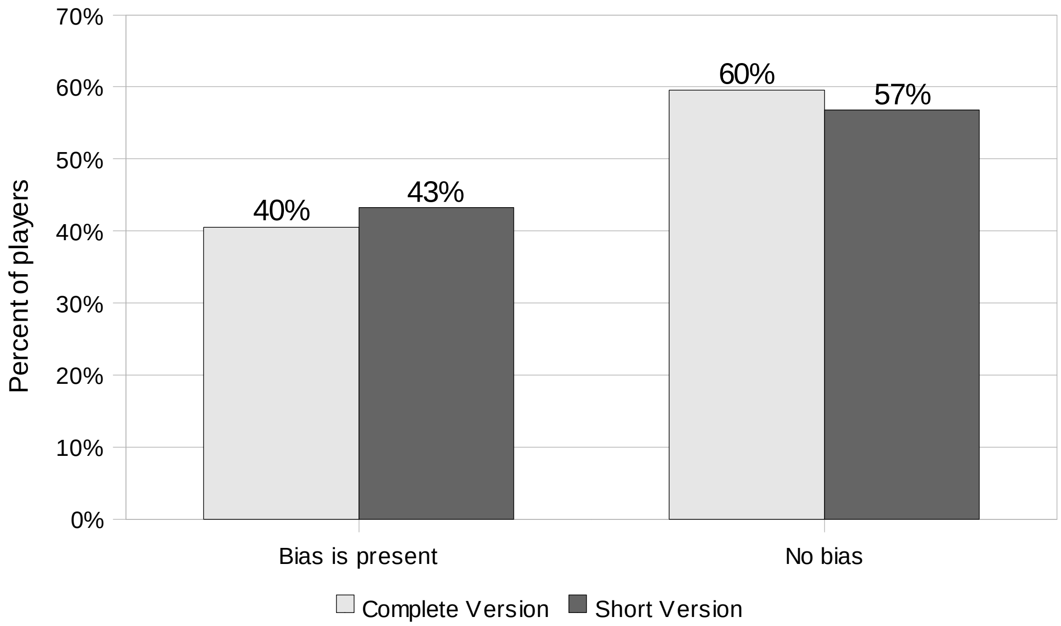
Rounds	'Pot is Riskier' Group	'Pocket is Riskier' Group
Number of players	8	34
First hypothetical and first real:		
Hypothetical (Pot−Pocket) Investments[a]	−4.625	2.706
Real (Pot−Pocket) Investments[b]	1.250	3.588
Friedman Statistic[c]	5.0000	0.7273
	(0.0254)	(0.3938)
Second hypothetical and second real:		
Hypothetical (Pot−Pocket) Investments	1.125	2.735
Real (Pot−Pocket) Investments	−4.625	3.559
Friedman Statistic	5.0000	0.0435
	(0.0254)	(0.8343)
Third hypothetical and third real:		
Hypothetical (Pot−Pocket) Investments	0.125	1.000
Real (Pot−Pocket) Investments	−2.375	2.588
Friedman Statistic	2.6670	0.0000
	(0.1025)	(1.0000)

[a] Average hypothetical (Pot−Pocket) investments are computed across all players for the corresponding hypothetical round.
[b] Average real (Pot−Pocket) investments are computed across all players for the corresponding real round.
[c] The Friedman test p-values are in parentheses.

Instructions to the game were prepared with an intention that, while listening to the instructions, subjects should not have gotten an impression that the Pot was being presented as an alternative preferable to the Pocket on social, moral, or ethical grounds. Nonetheless, 40–43% of all respondents purported the experimenter's intention to promote the Pot; see Figure 10. This suggests that some people perceived a hidden message in the experiment concerning Pot contributions.

If one assumes that people exhibit hypothetical bias as a result of purchasing moral satisfaction and/or of compliance bias, then, by eliminating warm glow and/or compliance bias, one can reduce the hypothetical bias. If this conjecture is true, then the group of subjects that purported the experimenter's

Figure 10: Purported Bias Toward the Pot



bias toward the Pot should exhibit hypothetical bias for all three pairs of hypothetical/real rounds, and the other group that did not sense such an intention should not demonstrate the presence of hypothetical bias. The following null hypothesis was formulated and tested to this end:

$$\mathrm{H}_0 : \text{Hypothetical (Pot}-\text{Pocket)}_i^k = \text{Real (Pot}-\text{Pocket)}_i^k$$

versus

$$\mathrm{H}_1 : \text{Hypothetical (Pot}-\text{Pocket)}_i^k \neq \text{Real (Pot}-\text{Pocket)}_i^k,$$

where $k \in [1, 2]$ indicates the group of players: the first group consists of players that purported the experimenter's bias toward the Pot, and players in the second group did not sense such an intention; $i \in [1, 3]$ denotes the round number; and (Pot$-$Pocket) is the difference in Pot contributions relative to Pocket investments.

The null hypothesis is that there is no difference between the (Pot$-$Pocket) investments in hypothetical and real rounds within a group; the alternative hypothesis is that there is a difference. Results of Friedman's test for the difference in behavior across the three pairs of hypothetical/real rounds due to the bias towards the Pot are presented in Table 12. Results indicate that those who found some social, moral, or ethical bias in the game instructions, demonstrate a significant negative hypothetical bias in the first pair of hypothetical/real rounds only. No bias is detected for the subsequent rounds. Those players who did not find any Pot promotion bias, do not demonstrate any significant hypothetical bias. Thus, the above conjecture that hypothetical bias is caused to some extent by the purchase of moral satisfaction is not fully supported by the data.

One can also try and group players by both risk perception and social bias (to obtain four groups), but having too many categories with the small sample size would result in a very few observations in a category. Any statistical test done in a setting like that would have virtually no power.

To summarize, the validity of the short version conclusions regarding the factor analysis for hypothetical bias may be questionable because of the limited investment options in the short version. Thus exploratory factor analysis on hypothetical bias is based on the complete version of the game. It is revealed that the small group of players who found the Pot to be riskier than the Pocket (the 'Pot is riskier'

Table 12: Sense of Pot Promotion Bias Grouping. Test for the Difference between Hypothetical and Real (Pot−Pocket) Investments, Complete Version

Rounds	'Bias' Group	'No Bias' Group
Number of players	17	25
First hypothetical and first real:		
Hypothetical (Pot−Pocket) Investments[a]	−0.412	2.480
Real (Pot−Pocket) Investments[b]	3.529	2.880
Friedman Statistic[c]	4.5714	0.0769
	(0.0325)	(0.7815)
Second hypothetical and second real:		
Hypothetical (Pot−Pocket) Investments	0.882	3.480
Real (Pot−Pocket) Investments	1.529	2.320
Friedman Statistic	0.0909	1.4706
	(0.7630)	(0.2253)
Third hypothetical and third real:		
Hypothetical (Pot−Pocket) Investments	0.059	1.360
Real (Pot−Pocket) Investments	1.176	1.960
Friedman Statistic	0.2857	0.2222
	(0.5930)	(0.6374)

[a] Average hypothetical (Pot−Pocket) investments are computed across all players for the corresponding hypothetical round.
[b] Average real (Pot−Pocket) investments are computed across all players for the corresponding real round.
[c] The Friedman test p-values are in parentheses.

group), exhibit significant or marginally significant hypothetical bias in all round pairs. Those who found the Pocket to be riskier than the Pot (the 'Pocket is riskier' group) did not demonstrate any statistically significant hypothetical bias in any pair of hypothetical/real rounds. In addition, the conjecture that hypothetical bias is caused to some extent by the purchase of moral satisfaction is not fully supported by the data.

6 Conclusions

The objective of the present study was to design an experiment, based on the public contribution game, which would shed some light on the magnitude of the hypothetical bias and help to reveal factors responsible for it. The key novelty in our study was the inclusion of a free-riding barring mechanism in a standard public goods game; also important were the steps we took to make the game have a look and feel of a real-world tradeoff between private investment and public good provision.

As a result of the laboratory experiment on budget allocation/investment behavior in different environments, several results have emerged. First, a statistically significant *negative hypothetical bias* is found between the first hypothetical and the first real rounds of the game. The bias in the subsequent second and third hypothetical/real pairs alternates in sign but is not statistically significant, despite having this peculiar oscillating pattern. The finding of a negative hypothetical bias is particularly intriguing in light of several recent studies which took steps to mitigate it without first establishing its existence. For example, Bulte et al. (2005) and Aadland, Caplan, and Phillips (2007) assume that hypothetical bias exists and is positive. Having based their results on this assumption, they conclude that a 'cheap talk' script and a 'consequential' survey or experiment may have a mitigating effect. The experiment presented in this study may provide a useful platform to investigate whether stated public good demand would still be lower in the bias mitigating settings even if there is no hypothetical bias to begin with.

The reasons for the negative sign of hypothetical bias detected in the first pair of hypothetical/real rounds remain open to speculation. This non-conventional result requires further investigation. One possible explanation is that what seems to be a negative hypothetical bias with respect to the Pot is in fact more of a positive hypothetical bias from the Pocket perspective. That is, players put more tokens to the Pocket when their choices were hypothetical — no consequences to risk — but their real Pocket contributions were more moderate because the risk of losing tokens became real. This conjecture could be examined by changing the odds of the Pocket lottery from 50/50 to 67/33, or even 75/25, where first number denotes the probability of success (tokens double) and the second one that of failure (tokens are decimated by half).

Additionally, a source of negative hypothetical bias can lie in the modified public goods mechanism itself. When players make their token allocations, they are seeking a return on their investments rather than an increase in public well-being (when subjects purchase a public good in a valuation study). This might explain in part why the difference in hypothetical versus real investments turns out to be negative in the experiment. What follows is that the experiment presented in this study may provide a useful platform to investigate how different experimental designs affect the presence and sign of hypothetical bias.

The reasons why we detect hypothetical bias in the first pair of hypothetical/real rounds and not in consecutive rounds are not clear. This issue requires further investigation. A possible explanation based on the fact that players might have gained some experience with the game (McClelland, Schulze, and Coursey 1993) does not seem to be supported by the data. We checked the hypothesis that people exhibit hypothetical bias as a result of purchasing moral satisfaction and/or of compliance bias. Data collected during the experiment do not seem to fully support this hypothesis.

The grouping of players by their risk perception toward the private and public investment — the Pocket and the Pot — reveals that those who found the Pot to be riskier than the Pocket, that is, those who may consider the reliance on other peoples' choices to be risky, demonstrate a significant hypothetical bias regardless of what round of the game being played. Those players who found the Pocket to be riskier than the Pot did not demonstrate any hypothetical bias at all. We believe that risk associated with other players' choices as a factor contributing to the existence of hypothetical bias requires further investigation.

The experiment design of this study has its limitations. First, the absence of a practice round in the experiment may have resulted in players being confused or unclear about the game environment. Second, unrealistically low initial endowment in the amount of ten tokens does not necessarily motivate players to reveal their 'true' preferences regarding the investment options offered in the game. Third, the random dictatorship mechanism is intended to eliminate free-riding but if players are confused about the game they may have still behaved strategically (at least in the first rounds) and the finding of a statistically significant negative hypothetical bias in the first pair of rounds is compromised. Fourth, the external validity of a situation when the random-dictator overrules the individual decision about one's private investment (when a pledge drawn by the computer is more than one's pot contribution and uninvested tokens are not enough to cover the deficit) is questionable. Finally, by the game construction, it is impossible to distinguish between the complexity result and the status quo bias.

Perhaps the most important recommendation that can be made for stated choice/contingent valuation studies is that one should be cautious in making a priori assumptions about the existence of hypothetical bias and, especially, in using rules-of-thumb for calibration (like 'divide by two') as the magnitude of hypothetical bias seems to be sensitive to the experiment design. Also, when respondents are queried on their preferences through a sequence of choice situations, the researcher is advised to take into account possible temporal changes. It may be wise to apply some sort of 'burn-in' approach to the series of responses or to use a system of weights. Finally, no a priori assumptions should be made with respect to choice complexity, since some respondents feel better with more varied/less restrictive choice sets while others prefer simpler but more restrictive setups.

References

Aadland, D. and A. Caplan (2003, May). Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias. *American Journal of Agricultural Economics 85*(2), 492–502.

Aadland, D., A. Caplan, and O. Phillips (2007, October). A Bayesian examination of information and uncertainty in contingent valuation. *Journal of Risk and Uncertainty 35*(2), 149–178.

Adamowicz, W. and J. DeShazo (2006, May). Frontiers in stated preference methods: An introduction. *Environmental and Resource Economics 34*(1), 1–6.

Andreoni, J. (1990, June). Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal 100*(401), 464–477.

Bagnoli, M., S. Ben-David, and M. McKee (1992, February). Voluntary provision of public goods : The multiple unit case. *Journal of Public Economics 47*(1), 85–106.

Balistreri, E., G. McClelland, G. Poe, and W. Schulze (2001, March). Can hypothetical questions reveal true values? A laboratory comparison of dichotomous choice and open-ended contingent values with auction values. *Environmental and Resource Economics 18*(3), 275–292.

Baumgartner, H. and J.-B. Steenkamp (2001, May). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research 38*(2), 143–156.

Bishop, R. and T. Heberlain (1986). *Valuing Environmental Goods: A State of the Art assessment of the Contingent Valuation Method*, Chapter Does contingent valuation work?, pp. 123–147. Totowa, NJ: Rowman and Allenheld.

Bohm, P. (1972). Estimating demand for public goods: An experiment. *European Economic Review 3*(2), 111–130.

Brown, K. and L. Taylor (2000, September). Do as you say, say as you do: Evidence on gender differences in actual and stated contributions to public goods. *Journal of Economic Behavior and Organisation 43*(1), 127–139.

Bulte, E., S. Gerking, J. List, and A. de Zeeuw (2005, March). The effect of varying the causes of environmental problems on stated WTP values: Evidence from a field study. *Journal of Environmental Economics and Management 49*(2), 330–342.

Champ, P. and R. Bishop (2001, August). Donation payment mechanisms and contingent valuation: An empirical study of hypothetical bias. *Environmental and Resource Economics 19*(4), 383–402.

Champ, P., R. Bishop, T. Brown, and D. McCollum (1997, June). Using donation mechanisms to value nonuse benefits from public goods. *Journal of Environmental Economics and Management 33*(2), 151–162.

Champ, P., R. Moore, and R. Bishop (2004, August). Hypothetical bias: The mitigating effects of certainty questions and cheap talk. Presented at 2004 AAEA Annual Meeting, Denver, CO. Available at http://ageconsearch.umn.edu/bitstream/123456789/15885/1/sp04ch21.pdf.

Conover, W.J. (1999). *Practical Nonparametric Statistics* (3rd ed.). New York, NY: John Wiley & Sons.

Cronbach, L. (1946, Winter). Response set and test validity. *Educational and Psychological Measurement 6*(4), 475–494.

Cummings, R., D. Brookshire, and W. Schulze (Eds.) (1986). *Valuing Environmental Goods–An Assessment of the Contingent Valuation Method*. Rowman and Allanheld, Totowa, NJ.

Cummings, R., G. Harrison, and E. Rutstrom (1995, March). Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review 85*(1), 260–266.

Cummings, R. and L. Taylor (1999, June). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. *American Economic Review 89*(3), 649–665.

Diamond, Peter A. and Jerry A. Hausman (1994). Contingent valuation: Is some number better than no number? *The Journal of Economic Perspectives 8*(4), 45–64.

Dickie, M., A. Fisher, and S. Gerking (1987, March). Market transactions and hypothetical demand data: A comparative study. *Journal of the American Statistical Association 82*(397), 69–75.

Dutta, B., H. Peters, and A. Sen (2002, October). Strategy-proof probabilistic mechanisms in economies with pure public goods. *Journal of Economic Theory 106*(2), 392–416.

Fischhoff, B. and L. Furby (1988, June). Measuring values: A conceptual framework for interpreting transactions with special reference to contingent valuation of visibility. *Journal of Risk and Uncertainty 1*(2), 147–184.

Gibbard, A. (1977, April). Manipulation of schemes that mix voting with chance. *Econometrica 45*(3), 665–681.

Griffin, C., J. Briscoe, B. Singh, R. Ramasubban, and R. Bhatia (1995, September). Contingent valuation and actual behavior: Predicting connections to new water systems in the state of Kerala, India. *World Bank Economic Review 9*(3), 373–395.

Hanemann, M. (1984, August). Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics 66*(3), 332–341.

Johannesson, M., B. Liljas, and P. Johansson (1998, May). An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions. *Applied Economics 30*(5), 643–647.

Lentz, T. (1938, November). Acquiescence as a factor in the measurement of personality. *Psychological Bulletin 35*(9), 659.

Levitt, S. and J. List (2007, Spring). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspective 21*(2), 153–174.

List, J. and C. Gallet (2001, November). What experimental protocol influence disparities between actual and hypothetical stated values? Evidence from a meta-analysis. *Environmental and Resource Economics 20*(3), 241–254.

List, J. and J. Shogren (1998, December). Calibration of the difference between actual and hypothetical valuations in a field experiment. *Journal of Economic Behavior and Organization 37*(2), 193–205.

Loomis, J., T. Brown, B. Lucero, and G. Peterson (1997, September). Evaluating the validity of the dichotomous choice question format in contingent valuation. *Environmental and Resource Economics 10*(2), 109–123.

McClelland, G., W. Schulze, and D. Coursey (1993, August). Insurance for low-probability hazards: A bimodal response to unlikely events. *Journal of Risk and Uncertainty 7*(1), 95–116.

Mitchell, R. and R. Carson (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Resources for the Future, Washington D.C.

Murphy, J., P. Allen, T. Stevens, and D. Weatherhead (2005, March). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics 30*(3), 313–325.

Muth, J. (1961, July). Rational expectations and the theory of price movements. *Econometrica 29*(3), 315–335.

Neil, H., R. Cummings, P. Ganderton, G. Harrison, and T. McGuckin (1994, May). Hypothetical surveys and real economic commitments. *Land Economics 70*(2), 145–154.

NOAA (1994, May). Natural resource damage assessment: Proposed rules. Federal Register 59, National Oceanic and Atmospheric Administration. pp. 23098-23111.

NOAA (1996, January). Natural resource damage assessment: Final rules. Federal Register 61, National Oceanic and Atmospheric Administration. pp. 439.

Prince, R., M. McKee, S. Ben-David, and M. Bagnoli (1992, July). Improving the contingent valuation method: Implementing the contribution game. *Journal of Environmental Economics and Management 23*(1), 78–90.

Rosenthal, R. and R. Rosnow (1969). *Artifact in Behavioral Research*. New York: Academic Press.

Sinden, J. (1988, August-December). Empirical tests of hypothetical biases in consumers' surplus surveys. *Australian Journal of Agricultural Economics 32*(2-3), 98–112.

Smith, V. and C. Mansfield (1998, November). Buying time: Real and hypothetical offers. *Journal of Environmental Economics and Management 36*(3), 209–224.

Tumer, K. and D. Wolpert (2004). *Collectives and the Design of Complex Systems*. New York, NY: Springer-Verlag.

Wolpert, D. and K. Tumer (2001). Optimal payoff functions for members of collectives. *Advances in Complex Systems 4*, 265–279.