Centre for Decision Research and Experimental Economics

Discussion Paper Series

ISSN 1749-3293

Discussion Paper

CeDEx Discussion Paper No. 2007–14

Virtual World Experimentation: An Exploratory Study

Thomas Chesney, Swee-Hoon Chuah and Robert Hoffmann

December 2007

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of Public Economics, Individual Choice under Risk and Uncertainty, Strategic Interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/economics/cedex/ for more information about the Centre or contact

Karina Whitehead
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0) 115 95 15620
Fax: +44 (0) 115 95 14159
karina.whitehead@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/economics/cedex/papers/index.html

Virtual world experimentation: An exploratory study

Thomas Chesney[*], Swee-Hoon Chuah[*] and Robert Hoffmann[*†]

December 9, 2007

Abstract

We explore the scientific potential of virtual worlds for experimental economists. In particular, we report the results of a series of virtual world experiments designed to examine the suitability of (a) users as subjects and (b) the computer interface as an experimental platform. Formal results and informal observations from the sessions are discussed in terms of the methodological opportunities and challenges of virtual experimentation generally.

JEL-Classification: C72; C88; C99; Z13
Keywords: virtual worlds, laboratory experiments, human values survey

1 Introduction

Artificial or virtual worlds (VWs) are online communities in which individuals interact in simulated three-dimensional environments. The potential for research they offer has recently attracted the attention of social scientists (Bainbridge, 2007). VWs are significant for two related reasons. First, the growing number of users and the scope and nature of socio-economic acitivity between them increasingly make VWs social phenomena worthy of study (Castranova, 2005, e.g.). For sociologists, VWs present evolving cultures with their own social institutions that are becoming more significant to society at large (Noveck, 2004). VWs constitute increasingly sizeable economies that produce, consume and trade in their own convertible currencies, raising typical economic research questions. The developing scope and nature of corporate and individual entrepreneurship in VWs provide new business models and practices of interest to management researchers.

[*]Nottingham University Business School, Jubilee Campus, Nottingham NG8 1BB, United Kingdom
[†]Corresponding author. Email: robert.hoffmann@nottingham.ac.uk

Second, VWs provide not only new social phenomena, but also novel methods of studying them (Bainbridge, 2007). In particular, the computer technology underlying VWs may facilitate the economical and large-scale recruitment of individuals from different cultural-geographical and socio-economic groups for participation in interviews, focus groups, surveys or experiments. The IT interface affords detailed control of the environment in which individuals interact and decide. In this sense, VWs may bridge the gap between laboratory experiments and computer simulations, allowing researchers to use real persons to study the relationship between the conditions of interaction and the evolution of social institutions in a controlled manner. Ethnographic observation of social networks and institutions may draw on micro-data routinely recorded by VW-engines and proceed more economically, unobtrusively and over longer periods.

For economics, the greater part of the potential VWs hold may lie in these new methods. Experimental economics is the sub-discipline most likely to benefit. In the current paper, we are interested in the scientific opportunities VWs provide here. The purpose of this study is to assess and explore this potential through a specifically designed empirical study. We report on what we believe are the first economic experiments conducted in a VW and present preliminary results as well as methodological observations. The rest of the paper proceeds as follows. In the next section, we introduce VWs from the experimental viewpoint. In section 3, we discuss their methodological implications in more detail and the motivation of our own work and research questions. In section 4, we outline our resulting experimental design and its implementation. The experimental results are reported in section 5. Section 6 discusses our general observations from the experiment in terms of the methodological issues we consider. The final section contains concluding remarks.

2 Virtual Worlds

VWs are networked, computer-simulated environments in which individual users interact using *avatars* as their virtual representations. While there is considerable variation between the alternative VWs that exist, they typically reproduce features of the physical world such as a three-dimensional topography containing virtual objects obeying simulated physical laws and the possibility of communication, social interaction and economic exchange between inhabitants.

Second Life (SL) currently has over 11.2 million registered avatars and monthly growth rates in excess of 20%.[1] Accounting for multiple and dormant registrations, there are an estimated one million regular users who

[1]Economic and general statistics concerning SL are available at: http://secondlife.com/whatis/economy.php and http://blog.secondlife.com/.

spend over 20 million hours logged in per month. Between 20 and 30,000 users are online at any one time. In terms of demographics, the majority of these are from populous industrialised countries including the USA , UK, Germany, Brazil, France and Japan. The median age of users is 36, with 57% being male.

SL is divided into individual sectors with topographical features in which avatars can operate, including oceans, rivers, mountains and beaches as well as flora. A typical location is displayed in figure 1. Avatars are capable of locomotion, including walking, running and flying and are immune to destruction. They communicate using instant text messaging (IM) and can signal voice intonation such as whispering and shouting as well as use gestures and body language. Public IM can be received by all avatars in the vicinity, while private IM is transmitted only between two avatars irrespective of location. Internet telephony has recently been introduced to SL. Users can edit the appearance of avatars in terms of physical features, clothing and assessories. As a result, avatars can assume the form of humans, animals, fantasy creatures or objects. Avatars are associated with user accounts that include money balances in Linden dollars (L$) which can be bought from or sold to Linden Lab, the creators and owners of SL, at a relatively stable exchange rate of about 260 L $per 1 U.S. dollar. A total of 3.85 billion L $(U.S. $14.8 million) are in circulation. SL provides an interface feature that allows immediate and direct account-to-account transfers. These balances can be used to purchase a portfolio of tradable virtual objects including land, buildings, vehicles, clothing, accessories and tools.

3 Virtual Experimentation

VWs such as SL may have potential as powerful new platforms for designing and conducting experimental research. Bainbridge (2007) makes the following case:

> Virtual worlds such as SL provide environments and tools that facilitate creating online laboratories that can automatically recruit potentially thousands of research subjects, over a period of months, at low cost. SL offers scripting and graphics tools that allow anyone to build a virtual laboratory building, functioning equipment to run the experiment, and incentives to motivate participation. (p. 473)

Traditional experimental economics involves testing economic theories by observing subject decisions under choice conditions systematically manipulated in laboratory settings. The field has recently begun to broaden its scope by exploring new methods and applications outside the standard laboratory environment commonly populated by Western student subjects.

Figure 1: Typical SL-screenshot showing the user's avatar(male foreground figure), the surrounding SL-environment and interface controls along the bottom.

There are two related ways in which experimentalists are trying to improve the realism of the decisions they observe. First, field studies in naturalistic settings are being proposed as a way of avoiding the distorting effects artificial laboratory settings may have on subject behaviour (Harrison and List, 2004). Secondly, new recruitment techniques and sampling locations are being used to overcome the reliance of experimentalists on Western university students to generate results (Anderhub et al., 2001; Henrich et al., eds, 2004, e.g.). VWs may give an opportune impetus to both of these concerns. First, due to their computerised interfaces, they may provide relatively controlled environments for conducting experiments while remaining within a naturalistic setting familiar to subjects. Secondly, VWs may be inhabited by a wider cross section of people such that sampling from different cultures and more heterogeneous backgrounds may be possible in a single location accessible to experimentalists.

Conversely, however, the very technology that generates these advantages may give rise to a number of *a priori* concerns about virtual experimentation. Principally, experimenters know little about the identity or state of the subjects behind the avatars that participate. This may make it difficult to recruit appropriate subjects, to ensure discipline in the virtual laboratory, to prevent repeat participation and subject collusion and to engender

subjects' trust and confidence in the experiment. There is a possibility of demographic or cultural idiosyncrasies of VW-residents generally. This may generate a sample bias that renders virtual experimentation inappropriate to test general economic theories. VW-subjects may have more hedonistic or short-term tendencies or show less conformity than the average person. In addition, virtual behaviour is not moderated by physical presence and may therefore not be comparable to traditionally-generated results.

We decided to investigate all these theoretical considerations by designing exploratory virtual experiments with two specific objectives in mind. The first is to use experimental results to assess the suitability of SL-residents as subjects representative of the economic decision makers to whom standard economic theories relate. The second is to use the practical experience of conducting a virtual experiment to assess the relative merits of this method. We now report how we approached these issues.

4 Procedure

All our experiments were conducted in SL. Although more popular VWs exist, SL provides a number of comparative advantages for experimenters. First, SL does not have a thematic focus or game features that may reduce the representativeness of its users. Second, its sophisticated interface has been designed for general social networking and is well-suited for experimental work. As a result, SL has a diverse population with relatively well-developed formal and informal social and economic institutions. We now describe how we designed and conducted the experimental sessions. Due to the exploratory nature of this project, our procedure evolved slightly as we gained more experience in SL-experimentation. In the following, we outline the general working procedure that we developed and deployed over the course of our experiments in terms of five stages of which individual experimental sessions consisted.

In general, we found that centralised recruitment calls using SL-related Internet message boards and Web fora did not generate interest. Instead, recruiting subjects within SL immediately prior to experimental sessions proved fruitful. Recruitment, the first stage of our sessions, therefore involved using the experimenters' avatars to enter the busiest SL-locations where we addressed individuals or groups of residents using a standard public IM invitation in English. We answered questions about the experiment, and *teleported* volunteers to our virtual experimental laboratory. In the briefing stage, subjects who arrived (typically in groups between two and seven depending on the task) were given note-cards with general information about the experiment concerning etiquette, anonymity, confidentiality and incentivisation as well as the URL for a website containing experimental instructions and a comprehension quiz. The decision task stage of experiment

Figure 2: A typical experimental session in progress. The experimenters' avatars are standing.

commenced after all subjects completed the quiz successfully. Subjects communicated decisions to the experimenter and received feedback via private IM. Next, in the survey stage, subjects were sent the URL of a webform containing the value survey and demographic questions. In the final, payment stage of the experimental session, subjects were paid earnings in $L on the spot using the SL payment transfer feature. A typical experimental session in progress is shown in figure 2.

5 Results

Our choice of tasks was guided by our objective to assess whether a VW-subject pool may be appropriate in testing economic theories. In particular, we wanted to examine whether virtual behaviour conforms to established results generated in conventional experimentation. As a result, we chose a number of standard decision tasks for the experiment, the ultimatum (UG), dictator (DG), public good (PGG), guessing (GG) and minimum effort (MEG) games. Previous results for all of these popular experimental games abound for a variety of conditions as well as demographic and cultural groups and provide ready benchmarks for our own results. Table 1 provides some general information about the decision task stages of our experiments.

Task	UG	DG	GG	PGG	MEG	ESS
Subjects (N)	64	30	31	32	31	113
Subjects per session (n)	4-5	4-5	4-7	4	5-7	n/a
Average pay (U.S. $)	5.25	1.95	2.30	20.15	8.25	3.85
Duration (minutes approx.)	25	10	25	35	20	10
Rounds or questions	1	1	10	10	10	21

Table 1: Summary statistics for experimental games and survey.

The second avenue for testing subject pool suitability is to survey and compare our subjects' values and demographics to those of standard experimental subjects and general populations. Values provide a measurement of a respondent's cultural orientation and are known to affect behaviour (Rokeach, 1973; Chuah et al., 2006). We used the human values survey designed by Shalom Schwartz for the European Social Survey (ESS) project (Schwartz, 2002). Likewise, a number of demographics such as gender, age, and nationality are known to affect behaviour (see Camerer (2003) for an overview). In the following sections, we report the results we obtained from the game tasks and survey.

5.1 Ultimatum Game

To conduct the UG experiments, we ran separate sessions with proposers and responders respectively. In the proposer sessions, subjects were given the task to decide how to share L$3000 (U.S. $11.50) with a randomly-chosen co-player from a responder session who had the choice to accept or reject the split, resulting in the proposed shared being paid out or neither player receiving anything. We recruited a total of 32 pairs.

UG behaviour reflects a mixture of altruism and strategic thinking on the part of subjects (Forsythe et al., 1994). Because instrumentally-rational responders should accept any share of the stake, rational proposers should offer the minimum. In experiments, subjects tend to reject offers below about 20% of that stake and offer in the region of 20-50%, reflecting proposer expectations of rejections and/or altruistic concerns. As noted, there is a host of previous UG-results under different conditions including repetition, subject information and anonymity. Our results are comparable to those of experiments with standard conditions as outlined in our design. Standard experiments have been conducted with diverse experimental populations that permit the benchmarking of observed SL-bargaining to assess the suitability of users as subjects. In particular, UG-behaviour has been shown as sensitive to the cultural and demographic characteristics of subjects (Camerer, 2003; Oosterbeek et al., 2004). Buchan et al. (1997) and Chuah et al. (2007) (CHJW) identified slightly but significantly higher of-

	SL	RPOZ 1	RPOZ 2	RPOZ 3	CHJW
n	32	24	27	29	40
Stake	11.50	10	10	10	16
Offers					
Mean	45.73	44	45	45	44
Mode	50	50	50	50	50
St. Dev.	18.6	7.2	9.6	21.00	9.5
Rejections					
% of offers $< 20\%$	33.33	-	-	50	-
% of all offers	6.25	8.3	22	24	15

Table 2: Summary statistics of ultimatum game offers (in % of the U.S. $stake) and rejections for n subject pairs in SL as well as in three locations reported by RPOZ and in the UK reported by CHJW.

fers of Asian subjects potentially linked to their collectivist values. Henrich et al., eds (2004) found a much wider range of offers (between 25-57%) in a series of experiments with traditional, small-scale societies in South America, Central and South-East Asia as well as Sub-Saharan Africa. Demographic influences on bargaining include age (Hoffmann and Tee, 2006) and gender (Solnick, 2001), which again can have small but significant effects. However, UG-results are relatively robust when played by standard university subjects in industrialised nations. Roth et al. (1991) (RPOZ) found little difference between offers made by urban subjects recruited in the U.S. (Tucson, RPOZ 1 and Pittsburgh, RPOZ 2), Tokyo (RPOZ 3), Yugoslavia and Israel. Henrich et al., eds (2004) identify proposed shares in the 42-48% range as typical for these subjects pools (see also table 2.2. in Camerer (2003)).

Table 2 reports summary statistics of UG bargaining by SL-subjects compared with behaviour reported in RPOZ for three locations and by CHJW for UK subjects. The SL mean offer is 45.73% of the stake with a modal offer of half. These central tendencies in the proposals are very similar to those reported for comparable samples. Figure 3 shows the distributions of offers of subjects in all these experiments. With the exception of a small number of hyper-fair outliers among SL-subjects, the distribution we found is also very similar to those in the previous studies.

5.2 Dictator Game

DGs were again conducted in separate sessions for proposers and responders, except that responders were not given the opportunity to accept or reject offers. The stake size was 1000 $L (U.S. $3.90). The DG was originally conceived as a way of separating strategic and altruistic motives in UG-offer behaviour (Forsythe et al., 1994). While instrumentally rational

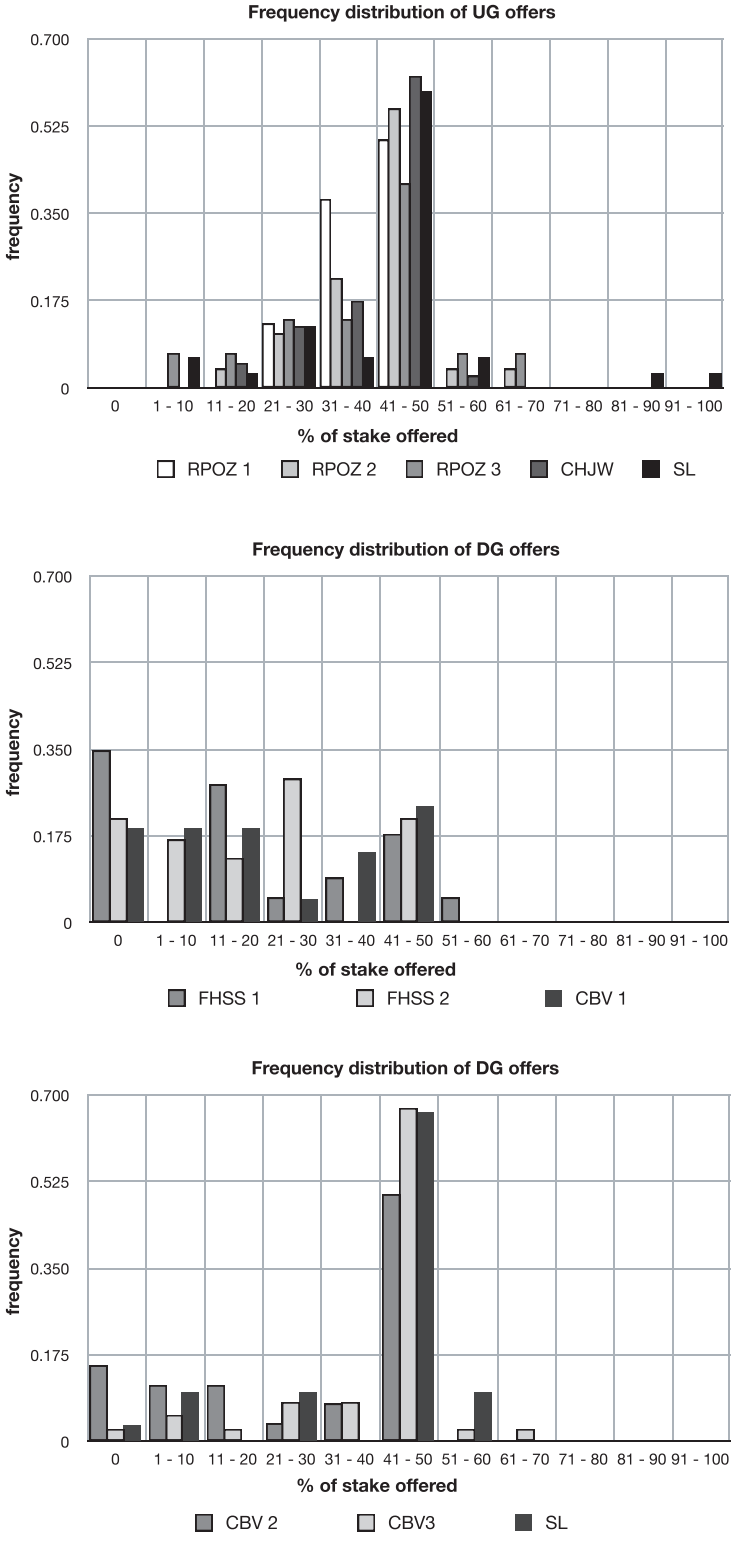


Figure 3: Distribution of UG and DG offers in SL as well as in selected previous studies.

	SL	FHSS 1	FHSS 2	CBV 1	CBV 2	CBV 3
n	30	24	45	21	37	26
Stake	3.90	10	5	100	100	100
Offers						
Mean	43.00	24	24	25	45	33
Mode	50	30	0	50	50	50
Median	50	25	20	20	50	45
St. Dev.	16.17	17.68	20.44	19	12	20

Table 3: Summary statistics of dictator game offers (in % of the U.S. $stake) for n subject pairs in SL as well as reported in previous experiments.

players should keep all of the stake, experimental subjects offer in the region of 20-35% to responders, reflecting altruistic preferences. DG-behaviour is sensitive to a host of experimental conditions such as anonymity, source and destination of the stake (see Camerer (2003) for an overview). In addition, subject demographics influence offers. We compare our results to those generated in previous studies using standard conditions and subjects. Table 3 reports summary statistics of SL-dictator behaviour compared to subjects in comparable studies by Forsythe et al. (1994) (FHSS) and Carpenter et al. (2005) (CBV). Figure 3 displays the distribution of offers in the experiments reported there. The first two of these studies (centre panel of the figure) report offers made by standard college student subjects which tend to be in the region of 23-24% of the stake (see also Hoffman et al. (1996); Cason and Mui (1998)), although some studies, such as Schotter et al. (1996), have found offers close to 40%. Of particular interest to us is the study by CBV, who identified marked differences in DG offer levels based on age and experimental location. In their study, they compare offers made by students (average age: 19.44 years) in standard college settings (CBV 1), by older community college students (26.91, CBV 2) and by workers in a warehouse setting (37.13, CBV 3).

The data show the DG offers made by SL-subjects to be higher than those reported in standard college settings, but similar to those made by older subjects in CBV. These results (bottom panel in figure 3) reflect the greater average age of our subjects (see section 5.6) and the fact that DG-offers are sensitive to age (Harbaugh et al., 2003). It is also noteworthy that in our experiment, proposers communicated their offers to the experimenter directly using private IM rather than using forms collected and delivered in stacks by monitors as tends to be practiced in physical locations. Our treatment provides more scope for social influence and demand effects that would be expected to raise offers.

5.3 Public Good Game

In the PGG sessions, subjects in groups of $n = 4$ asked to divide L$400 (U.S. $1.50) between a private and a group fund and explained that their total earnings would be their private allocation plus $a = 0.4$ times the total of all group allocations. This process was repeated ten times. The game is a n-person version of the prisoner's dilemma and pits subjects' self-serving motives against their desire to further the benefit of the group. Instrumentally-rational play involves complete free-riding and allocating the whole endowment to the individual fund. In repeated PGGs, players decisions may be guided both by strategic considerations of reciprocation and purely altruistic motives.

Again, a large literature exists that identifies the experimental conditions that elicit cooperative behaviour. In general, subjects contribute positive amounts to the public good that steadily decline when the game is repeated. The studies that conduct PGG games under standard conditions serve as benchmarks for the behaviour of our SL-subjects. We compare the behaviour of SL-subjects with those in experiments with comparable conditions reported by Andreoni (1988, 1998) as well as Fehr and Gächter (2000), who used values for parameter a of 0.5, 0.5 and 0.4 respectively.

Panel 2 in figure 7 shows the average contribution to the group fund subjects made in SL and the three previous studies we use as benchmarks. SL-subjects contribute marginally more than subjects in the other pools in all rounds. The average contribution decays over rounds in similar ways in all studies. The higher average we find is not unusual within the context of findings made using variegated subject pools. For instance, Henrich et al., eds (2004) report on PGGs played with traditional society subjects in many continents and find mean contribution rates to vary between 22 and 65%. These subjects differ from standard college subjects in a number of ways, age being one. Our result may also be due to the apparent greater altruism of SL subjects we observed in the DG.

5.4 Minimum Effort Game

In the MEG sessions, groups of $n = 5$ to 7 subjects were asked to choose an integer in the interval $[1, 7]$ and informed payoffs would be determined by the smallest number chosen within the group according to the payoffs adapted from Van Huyck et al. (1990) (VBB) and shown in the matrix in table 4. Each group played this game a total of ten times. The game has multiple equilibria in which all players make the same choice, which payoff dominate each other in turn with a unique Pareto-efficient equilibrium in every player choosing 7. The game represents situations where a group's ability to coordinate on the individually as well as collectively best outcome may be undermined by individuals' pessimistic expectations of others' rea-

	Smallest choice in group						
	7	6	5	4	3	2	1
7	390	330	270	210	150	90	30
6		360	300	240	180	120	60
5			330	270	210	150	90
4				300	240	180	120
3					270	210	150
2						240	180
1							210

Table 4: MEG payoff matrix (in L$). The first column represents player choices which, combined with the smallest choice in the group determines payoffs.

soning. A typical example is punctuality (Camerer, 2003). While everyone arriving on time for a meeting is mutually the best outcome, an individual may arrive late to avoid a wait expecting others to be late also. After a number of meetings, such expectations may become increasingly self fulfilling as general punctuality disintegrates.

The experimental evidence shows this kind of convergence on payoff-dominated outcomes depending on the size of the group, the size of payoffs and information players receive about the choices of others.

Figure 4 shows the distribution of choices and minimum choices in SL and comparable previous studies of Knez and Camerer (1994) (KC) and Bornstein et al. (2002) (BGN). In all these, the MEG was played in groups between five and seven subjects with the same payoffs we based ours on. There are no obvious differences in the distributions of average choices between the three studies. The minima in SL appear somewhat greater than in the other two studies. A similar view arises from the round-to-round changes in the choices and minimum choices averaged over experimental groups (figure 5). The figure also includes data from experiments reported in VBB, Riechmann and Weimann (2008) (RW) and Devetag (2005) (D), who used similar payoffs and groups of n = 14-16, 7 and 7 and respectively. In all studies, similar declines in the choices are visible.

5.5 Guessing Game

In our GG sessions, $n = 4$ to 7 subjects were asked to choose integers in the interval [1,100] and informed that the subject with a response closest to $p = 0.7$ times the average of all choices would receive L$200 (U.S. $0.75). Ties were resolved by dividing this sum among the winners. This procedure was repeated ten times for each group of subjects.

The GG (sometimes known as the beauty contest game) is used as a tool

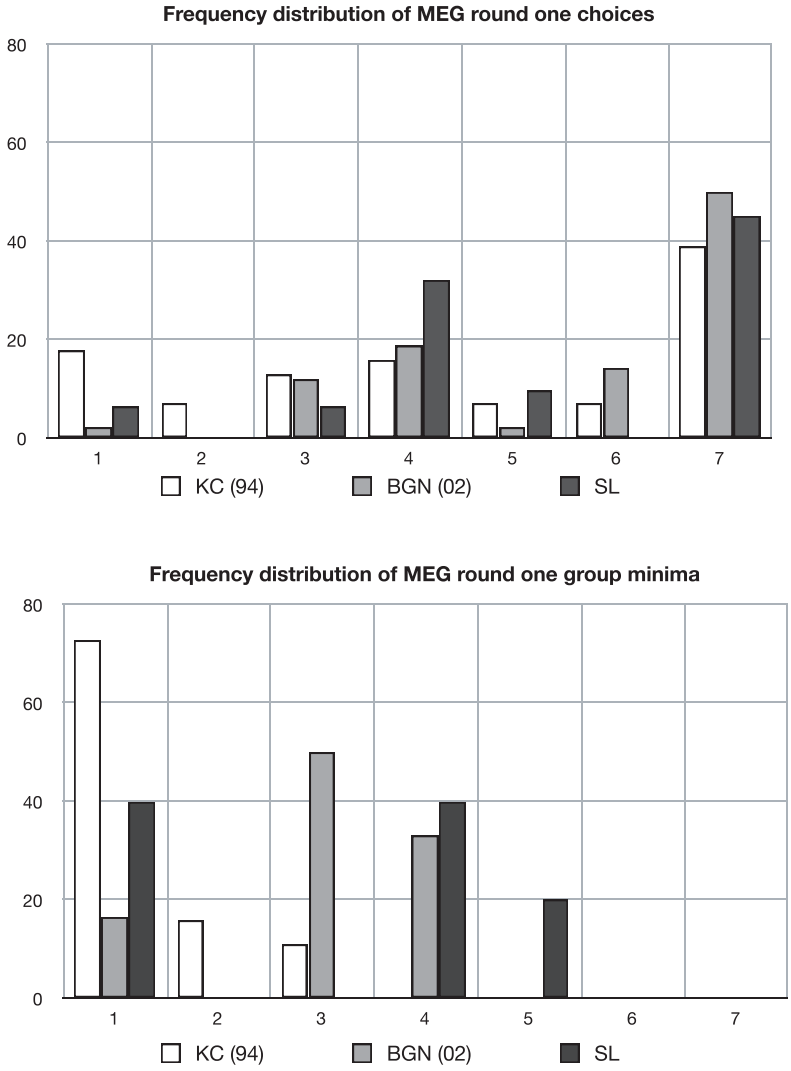
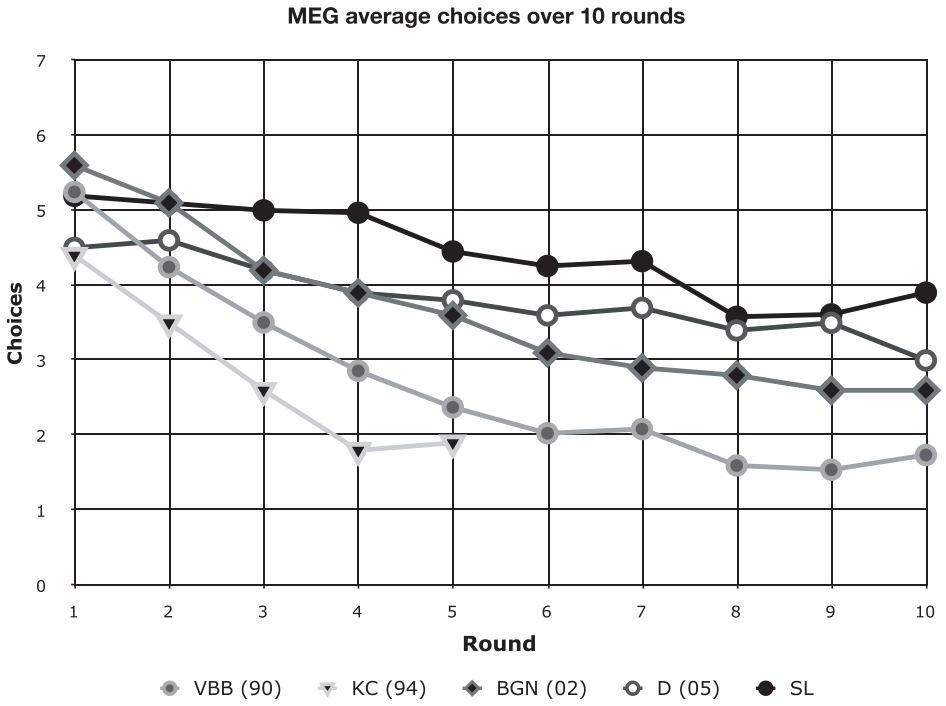


Figure 4: Distribution of round one MEG choices and minimum choices in SL and selected previous studies.



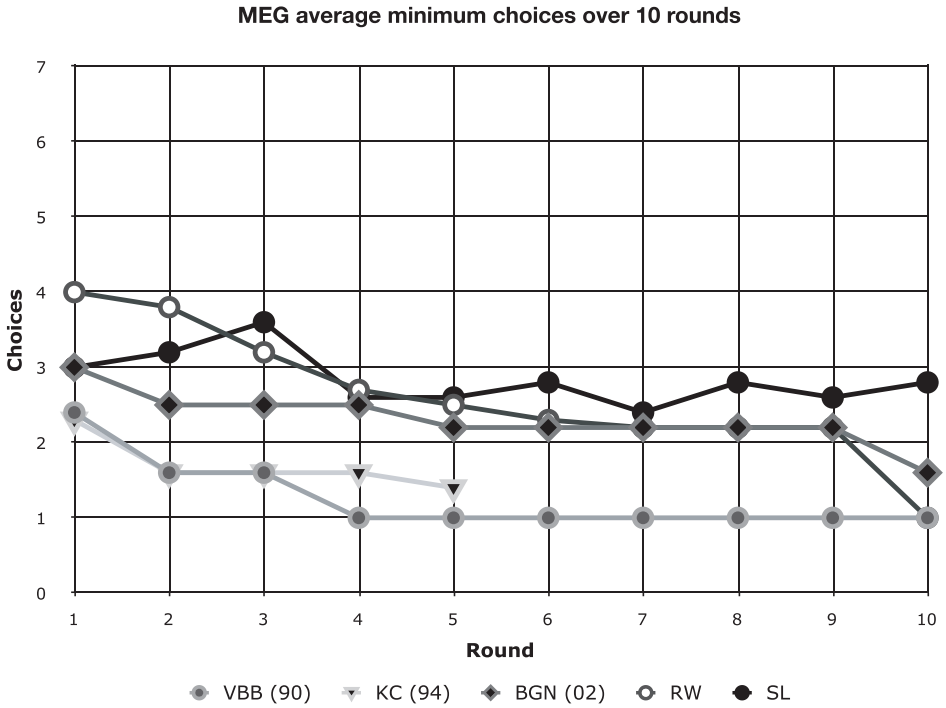


Figure 5: Average and average minimum MEG choices over ten rounds in SL and selected previous studies.

Subjects	Mean	Median	St. Dev.	% 0	n
SL	50.00	56.00	27.10	0.00	31
Portfolio managers	24.31	24.35	16.15	0.08	26
Economics PhDs	27.44	30.00	18.69	0.13	16
University CEOs	37.81	36.50	18.92	0.03	73
Caltech students	21.88	23.00	10.35	0.07	27
German students	36.73	33.00	20.21	0.03	67
Singaporean students (HCW 1)	36.45	35.00	24.28	0.00	21
Singaporean students (HCW 2)	58.27	50.00	26.98	0.05	22
Singaporean students (HCW 3)	39.78	35.00	25.46	0.02	46
Wharton students	37.92	35.00	18.84	0.00	35
U.S. high school students	32.45	28.00	18.61	0.04	52
College students (KCGPA 1)	35.00	35.00	12.86	0.00	51
Senior citizens (KCGPA 2)	37.00	33.00	17.46	0.00	50

Table 5: Summary statistics for round 1 GG choices in n-subject pools in SL as well as reported by Camerer (2003), Camerer (1997) and **?**. The percentage of subjects choosing 0 is given by %0.

to identify what levels of reasoning subjects employ in strategic thinking (Nagel, 1995; Duffy and Nagel, 1997; Camerer, 1997). A zero-order (i.e. unstrategic) player may choose randomly or use a focal point such as the median of the interval (50 in our case). First-order choosers may take others into consideration but assume these to be of order 0. An optimal first-order choice would be in the interval [0,70] accounting for the impossibility of the group average to exceed 70. In particular, a choice of 35 (0.7×50) may reflect a belief that zero-order guessers choose 50 on average. Second-order players who assume others to use order 1 will not choose above 49 (0.7×70), and may opt for 25 (0.7×35) believing order 1 choices to average 35 and so forth. The iterative application of increasingly higher levels of reasoning will eventually yield an equilibrium choice of 0.

The average and distribution of GG-choices therefore provides insight not only to what levels of reasoning subjects use, but also what levels they attribute to others. Equilibrium choices may reflect higher orders of reasoning but be ineffective when other players operate at lower levels. In addition, repeated GGs show to what extent subjects learn to adjust their choices on the basis of previous rounds' results. Table 5 shows statistics about subjects' choices in single or first rounds of repeated games played in groups of different sizes with a parameter p=0.7. The Singaporean student data are from 10-round GG-experiments reported in Ho et al. (1998) (HCW). The HCW 1 pool consisted of 3-player groups playing the game for the first time. Subjects in HCW 2 also played in 3-player groups but had experience of one

previous game with a different p-value. Finally, HCW 3 was composed of inexperienced 7-subject groups players. The U.S. study of Kovalchik et al. (2005) (KCGPA) compare one-round choices by college students (KCGPA 1) with those of mentally healthy senior citizens with an average age of 82 (KCGPA 2). Our experimental settings of group size, p-value and repetition are the same as in HCW 1, which is most useful for a direct comparison.

SL first round choices are relatively high (especially compared to our benchmark HCW 1) but by no means outside the range of previous results. The first panel in figure 7 shows mean choices over 10 rounds among SL-subjects and Singaporean students (HCW 1). Our subjects did appear to converge towards the equilibrium at similar rates to the latter group. The frequency distribution of individual SL-choices over all ten rounds is displayed in figure 6, along with the corresponding data for HCW 1 reported in Ho et al. (1998)(p. 955, figure 2E). Both distributions are similar in that a greater proportion of choices are low in later rounds. The SL-data appear different mainly in the more equal distribution in early rounds. However, towards the end of the game, the distributions are more similar, reflected in the convergence of curves in figure 7.

GG data generally show great variety in first-round average choices. Part of the reason may be the role that players' common knowledge of rationality has in equilibrium reasoning. Lower choices are not merely associated with greater strategic sophistication among players, but also with greater expectations concerning the sophistication of others. Groups that are more sophisticated as well as more uniformly so, such as Caltech students, may therefore be expected to exhibit lower choices than comparatively heterogeneous groups such as SL where little is known about others that take part. Our first-round results may have not been much different had our pool consisted of anonymous and mutually unaware game theorists disguised by avatars. The fact that SL-subjects' learning resulted in similar final-round choices supports this possibility. The anonymity of SL, potentially subverting the common knowledge of rationality, may therefore partly explain any differences in round one choices in SL.

5.6 Universal Human Values

In order to assess whether an idiosyncratic cultural environment exists within SL, we administered the ESS human values survey. This survey is based on Schwartz's portrait value questionnaire, a well-tested instrument for ten universal value dimensions (listed in figure 8). An individual's scores are calculated on the basis of responses on a 6-point Likert scale indicating own similarity with 21 hypothetical value portraits. Subjects completed the survey on a webpage immediately after the decision task stage of the session. Upon completion, each subject was paid L$1000 (ca. U.S. $3.85) for the survey in addition to the pay-outs from the decision task.

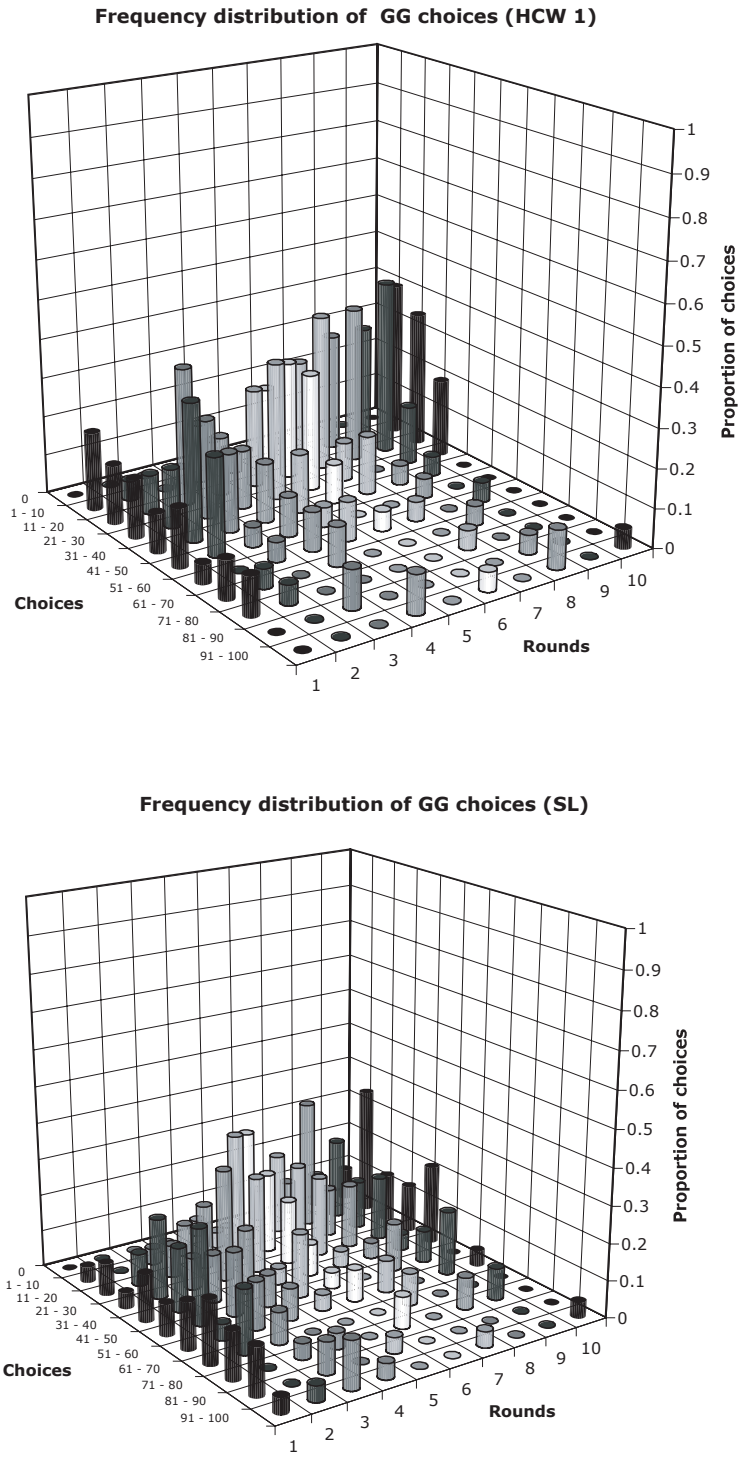


Figure 6: Subject choice frequency distributions over 10 rounds (group size 3, p=0.7) in Ho et al. (1998) (HCW 1) and SL.

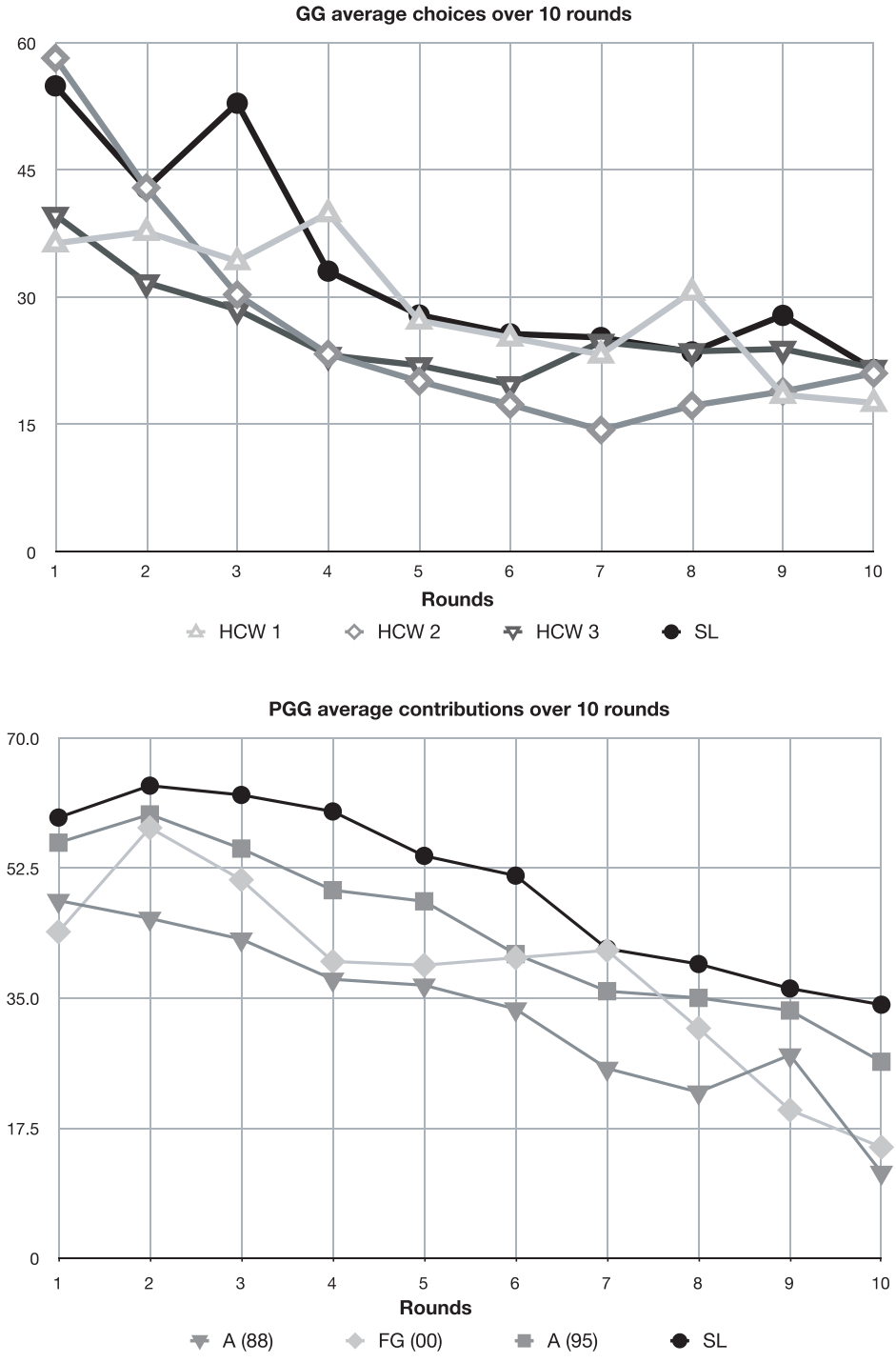


Figure 7: Average subject decisions in GG and PGG over ten rounds in SL and selected previous studies.

Again, a host of existing data for this survey generates scope for comparing SL-subjects with standard populations. Cultural and demographic factors may have an influence on economic behaviour as they shape an individual's social interaction and socialisation into particular values or attitudes. Values are therefore an important indicator of how representative particular subject pools are of the underlying population to which economic theory relates. We conducted the human values survey in order to ascertain to what extent SL-residents resemble standard experimental subjects culturally. Figure 8 shows the average value orientations of our subjects compared with those of respondents of the 2002-2003 ESS, as well as a standard sample of thirty-six UK university students (UKU) we also administered the questionnaire to. The ESS randomly samples more than 1500 adults from each participating nation's resident population. The students were UK nationals recruited and surveyed following traditional methods. For comparative purposes, we follow the ESS practice of presenting averages of ipsative scores, i.e. an individual's Likert-scale responses standardised in terms of his or her overall response average and variance. Ipsatised scores for different value dimensions have the advantage of being comparable in terms of relative strength.

Schwartz' ten human values have established empirical interrelationships that are commonly used to reduce them to two basic dimensions. The first, self-transcendence v. self-enhancement, encompasses values of hedonism, stimulation, self direction relative to tradition, conformity and security. The second, openness to change v. conservatism, weighs universalism and benevolence against power and achievement. Figure 9 plots nations and subject pools according to these two.

Our survey data indicate that while SL-users' value orientations differ from those of ESS-respondents, they do so to a lesser extent than those of the UK student subjects. The SL and student average value orientations correlate at 90% with each other, and respectively at about 70 and 64% with the averaged overall ESS-orientation. By comparison, individual national samples within ESS correlate with the average ESS-value profile at about 94%. The graph shows a relatively small distance between randomly-sampled individuals from European nations and larger ones to SL-users and students. The students place a greater value on the factors underlying self-enhancement, as can be verified in figure 8. This is consistent with age effects found in previous value surveys comparing students and teachers (Schwartz, 2001). Another reason for the difference may lie in a slightly higher relative socio-economic background and educational potential of students. However, caution has to be exercised due to our small sample size.

The demographical data our survey produced are summarised in figure 10. The average age of respondents was 31.7 with a minimum of 18 and a maximum of 64. Compared with the general European population, the age range 20-40 years was over-represented, an expected result given the tech-

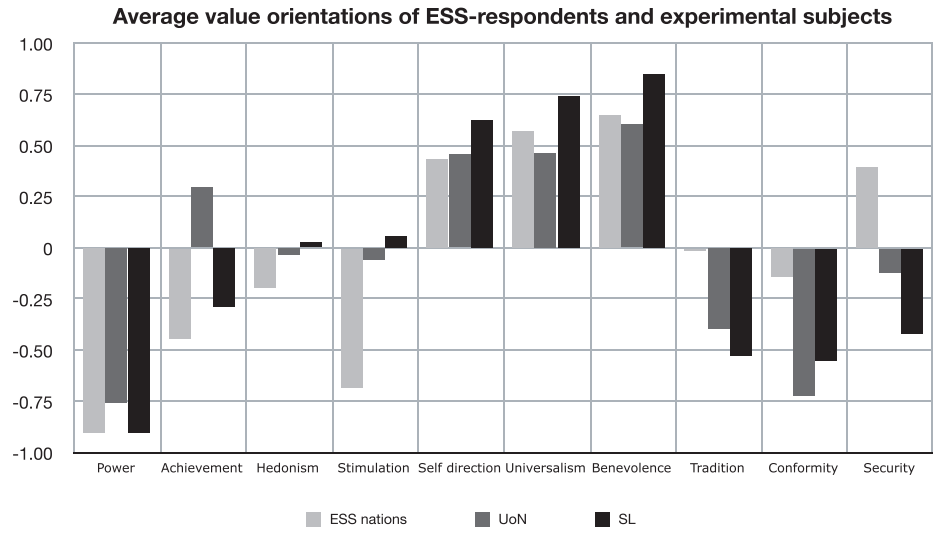


Figure 8: Average orientations of ESS-respondents (ESS), SL and UK student subjects according to Schwartz' ten value dimensions.

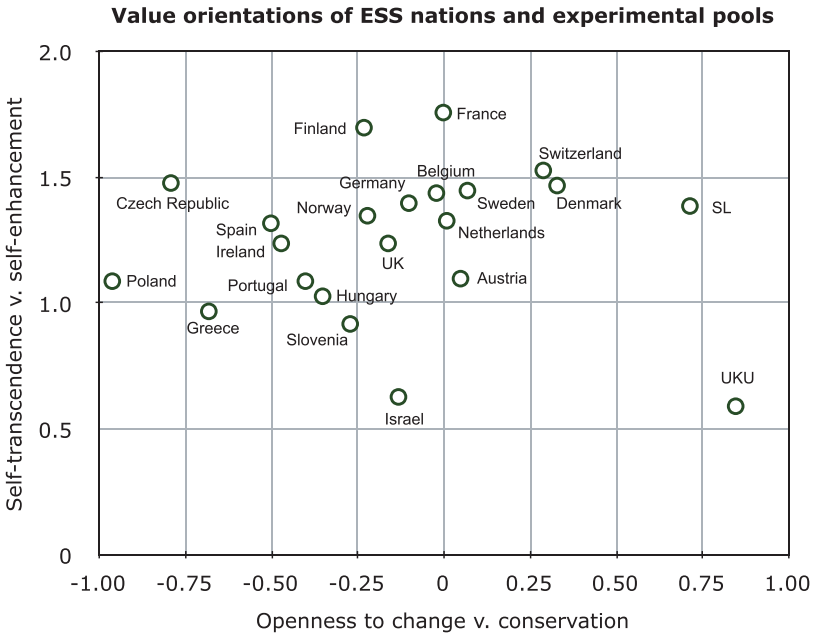


Figure 9: Average orientations of ESS-respondents by nationality, SL and UK student subjects according to Schwartz' two composite value dimensions.

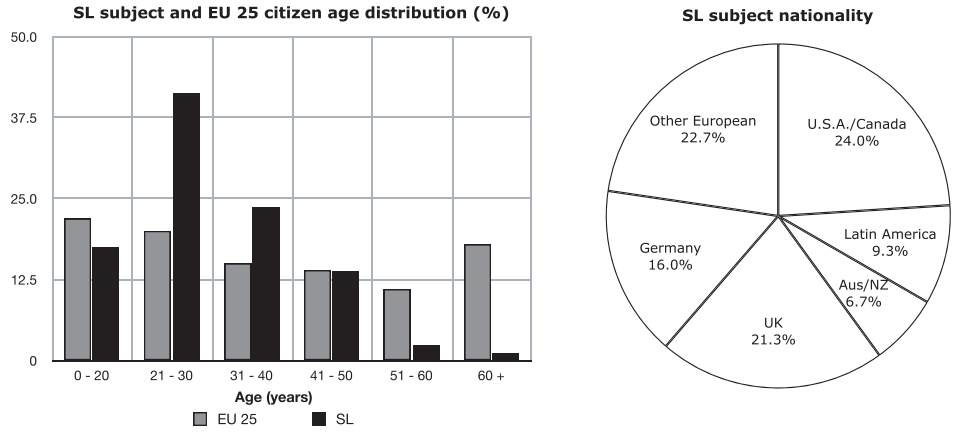


Figure 10: Age and nationality distribution of SL-subjects.

nological and cultural status of VWs. In line with SL generally, most were from populous Western nations, although UK and European countries were somewhat over-represented in our sample. The reason may lie in using the English language and our institutional affiliation in recruitment. Recruiting during GMT daytime hours further selects in favour of subjects from countries in similar time zones. Exactly half of our respondents declared female gender, and the remainder male.

6 Methodological Findings

In addition to our formal results, our experiment generated a host of observations concerning the suitability of SL as an experimental platform. We now discuss these as they relate to certain dimensions of experimental method (see, for instance, Davis and Holt (1993); Roth (1995); Harrison and List (2004)). The methodological literature has identified a number of desiderata for each, by which we measure the potential of the virtual platform.

6.1 Recruitment

The purpose of recruitment is to select a sufficiently large number of subjects from an appropriate pool representative of the general population of economic agents to which a particular theory relates. In addition, subjects should be able and willing participants unencumbered by expert knowledge of the task or high opportunity cost. To what extent can virtual experiment recruitment meet these conditions? We identified two obvious ways to recruit subjects in-world. Experiments can be advertised on SL-related Internet fora to solicit expressions of interest in scheduled sessions via e-mail. Alternatively, individual avatars and groups can be approached di-

rectly within SL for on-the-spot participation. We opted for the latter route to capitalise on the uninhibited social culture and conventions of SL. As we wanted to recruit a representative SL-subject pool, we selected locations on the basis of traffic alone. In terms of quantity, this method proved fruitful and required little preparation. In this manner, we were able to recruit two to five subjects within about fifteen minutes.

This procedure has both advantages and disadvantages. First, recruitment locations have an impact on sample characteristics. SL provides a wide range of sectors with themes based on activity (such as money making, adult entertainment or role playing) or first-world locations (nations or cities) that enable the experimenter to target particular subject types based on interest, demographics or nationality. Conversely, representativeness may be compromised to the extent that SL generally and its busiest locations in particular are unrepresentative. In addition, there may be response bias in soliciting volunteers, in the recruitment language, time and institutional affiliation, as discussed earlier. As a result, our sample may not be wholly representative of SL generally.

We also encountered a number of recruitment problems resulting from the anonymity of users and the lack of physical presence that moderates behaviour in a real laboratory. The first was the possibility of repeat participation by the same individual using another avatar. Within the SL-community, this is known as *farming* and is practised by some users to exploit lucrative paid surveys. The problem cannot be avoided altogether, but may be mitigated to some extent by disqualifying unsolicited volunteers, ones referred by previous subjects or avatars registered after the first experimental session. Another problem concerns the physical state or identity of the individuals controlling subject avatars. It is in practice difficult to assess whether recruits are unfit (tired or intoxicated) or unsuitable (such as children) for participation. In isolated incidents, subjects appeared unwilling to participate earnestly and were excluded. When a participant is intent on capitalising on the incentives the experiment provides, it is difficult in practice to prevent these types of problem. There may be scope for pre-experimental suitability and sobriety tests.

6.2 Environment

The experimental environment should conform to a number of conditions conducive for data generation and collection. Communication should be effective, no disturbing factors should be present and subjects should be prevented from colluding. Experimental behaviour should be observed, recorded and stored accurately. The computerised interface of SL has several advantages in these respects. In a virtual environment, it is simple and economical to maintain appropriate laboratory facilities to which subjects have easy access. SL-experimentation allows the equivalent both of pen-and-

paper and computerised experiments. Electronic documents can be easily distributed. Alternatively, subjects can be given addresses of online documents or computerised experimental interfaces. Communication using IM permits addressing subjects as a group as well as privately. Data collection using webforms and IM transcripts is straightforward and provides detailed records about all communication with subjects and the time needed to perform tasks. Screenshots can be used to record visual information.

Again, there were also disadvantages associated with the interface. In general, the process of communicating with subjects using IM made it difficult to deal with more than a handful per session. In addition, private IM prevents the detection of collusive behaviour or conferring amongst subjects. We tried to minimise the effects of this by never pairing subjects present against each other. In order to reduce the possibility of subject collusion, we commissioned the creation of customised lab furniture that restricts subjects' vision and alerts the experimenter of private IM traffic. While this furniture can prevent most SL-subjects from communicating with each other, experienced users may be able to circumvent the mechanism. Another problematic issue is establishing subject trust in the experimenters. Because of the nature of VWs, it is difficult to convince subjects of the genuine nature of the experiment and incentivisation. A further problem, generic to all VWs, is *griefing*. This involves individuals' attempts to to disrupt the activities of others using computer programs. We resolved this issue by restricting access to the virtual laboratory to invited subjects.

6.3 Control

This parameter refers to the extent to which the experimenter can systematically manipulate the subjects' decision task in order to assess behaviour under alternative task conditions or treatments. The virtual laboratory can be used when control parameters can be explicitly communicated in the experimental instructions. However, VW experimentation is clearly not appropriate when physical phenomena are part of the experimental treatment, such as when the effect of face-to-face interactions is tested. For instance, physical presence moderates behaviour through involuntary non-verbal communication such as body language, vocal pitch and intonation, blushing, sweating and pupil dilation and so forth which reveal emotional states. In addition, the potential for anonymity means that the social consequences of virtual behaviour are different to those in physical laboratories, where the possibility of physical harm, social sanctions and economic consequence may have a stronger effect. The absence of physical signals and presence in VWs may create confounding differences between virtual and physical experimental conditions.

In designing decision tasks, there is a trade-off between the extent of control and the degree of realism. Bainbridge (2007) speculates that VWs

may provide a way to bridge the gap between both of these desirable features. In particular, it may be possible to observe avatar behaviour in typical VW social activities that involve risk or scope for conflict and cooperation with others. There are many such opportunities in SL that we are exploring in current work, such as the collective building of virtual artefacts.

6.4 Incentivisation

Experimenters aim for subject incentivisation that is sufficient in terms of subject opportunity cost, delivered promptly, obeys non-satiation, dominates all other concerns, relates to subject decisions, and to the decision task in a naturalistic way. The relatively sophisticated SL-economy provides some scope for appropriate incentive mechanisms. In particular, SL has developed informal labour and product markets which generate incentivisable subjects as well as money or in-kind rewards that can be used. Many residents participate in paid activities such as *camping*, i.e. receiving payment for avatar presence in commercial locations to advertise products or complete online surveys for commercial organisations. The hourly return for these activities is low compared with the average payoffs in standard economic experiments. Relatively generous incentivisation does however contribute to the problem of farming discussed earlier. Payment delivery is unproblematic as account-to-account transfers of L $are immediate and direct. In addition to virtual money, a range of alternative, potentially more naturalistic incentives are available in a well-defined range of virtual objects which most residents demand and can be bought in virtual shopping malls (Bainbridge, 2007).

7 Discussion and conclusions

The research question that motivated our work was to what extent SL can be an appropriate research tool for experimental economists. This question has two parts; the first is to what extent SL-residents constitute an appropriate subject pool, and the second to what extent the SL-interface generates a suitable experimental platform. Our formal results and general findings provide some answers to these two questions respectively.

To begin with, our findings indicate that there are no obvious disadvantages in recruiting subjects in SL compared with university environments. The behaviour we observed in a range of standard experimental games is typical in view of previous findings. When differences were found such as in the DG, demographical differences (such as age) rather than cultural idiosyncrasies or procedural difficulties were seen to be responsible. In addition, there is a slightly lesser cultural and age bias within SL than at the average university campus. Users' values are more in line with those of general populations of economic agents. There was little evidence of users'

niche interests or motivations generating an unsuitable subject pool. Our work therefore supports Yee (2006), whose study of VW-demographics dispels the popular notion that VWs are predominantly the domain of a male, adolescent sub-culture with niche interests. His data indicate that usage and appeal are equally strong over gender and age groups as well as based more on general social motivations (such as relationship building) than escapism. In addition, SL's greater subject diversity provides opportunities for targeting particular types of individuals. Clearly, VWs provide no scope for experimenting with individuals of demographic and cultural groups that are currently not represented, including traditional groups. Alternatively, it is possible to exploit the SL-interface while relying on traditionally-recruited subjects who can access the virtual environment from physical experimental laboratories. It should also be borne in mind that the bias of SL towards industrialised nations is likely to change as economic development provides greater access to the Internet to more people worldwide.

The second issue concerns the suitability of a virtual experimental platform. We identified a number of positive and negative issues which suggest that this issue will depend much on the type of experiment planned. While many standard decision tasks can be easily recreated virtually, the increasing number of studies that consider the effects of physical signals will find little value in VW experimentation. Overall, our experience suggests that the virtual environment can simulate most of the crucial features of a physical laboratory at much lesser cost. The future development of VW-technology will further increase the sophistication of the virtual experimental platform.

Acknowledgments

The authors area grateful to Nottingham University Business School for supporting this project financially. Our thanks for helpful comments go to Colin Camerer, Simon Gächter, Martin Sefton, Chris Starmer and Jonathan Tan.

References

Anderhub, V., R. Müller, and C. Schmidt, "Design and evaluation of an economic experiment via the Internet," *Journal of Economic Behavior and Organization*, 2001, *46* (2), 227–247.

Andreoni, J., "Why Free Ride? Strategies and Learning in Public Goods Experiments," *Journal of Public Economics*, 1988, *37*, 291–304.

_ , "Cooperation in Public-Goods Experiments: Kindness or Confusion?," *American Economic Review*, 1998, *85* (4), 891–904.

Bainbridge, W. S., "The Scientific Research Potential of Virtual Worlds," *Science*, 2007, *317*, 472–476.

Bornstein, G., U. Gneezy, and R. Nagel, "The Effect of Intergroup Competition on Group Coordination: An Experimental Study," *Games and Economic Behavior*, 2002, *41*, 1–25.

Buchan, N. R., Johnson. E., and R. T. A. Croson, "Culture, Power and Legitimacy: Contrasting influences on fairness beliefs and Negotiation Behavior in Japan and the United States," 1997. University of Wisconsin Department of Marketing Working Paper.

Camerer, C. F., "Progress in Behavioral Game Theory," *Journal of Economic Perspectives*, 1997, *11* (4), 167–188.

_ , *Behavioral Game Theory*, Princeton, NJ: Princeton University Press, 2003.

Carpenter, J. P., S. Burks, and E. Verhoogen, "Comparing Students to Workers: The Effect of Social Framing on Behavior in Distribution Games," in J. P. Carpenter, ed., *Field Experiments in Economics*, Vol. 10, Elsevier, 2005, chapter 7, pp. 261–290.

Cason, T. N. and V. L. Mui, "Social Influence and the Strategy Method in the Sequential Dictator Game," *Journal of Mathematical Psychology*, 1998, *42*, 248–265.

Castranova, E., *Synthetic Worlds: The Business and Culture of Online Games*, Chicago, IL: University of Chicago Press, 2005.

Chuah, S. H., R. Hoffmann, M. K. Jones, and G. A. Williams, "An Economic Anatomy of Culture: Attitudes and Behaviour in Inter- and Intra-National Ultimatum Game Experiments," 2006. Occasional Papers 13, Industrial Economics Division, Nottingham University Business School.

_ , _ , _ , **and** _ , "Do Cultures Clash? Evidence from Cross-National Ultimatum Game Experiments," *Journal of Economic Behavior and Organization*, 2007, *64*, 35–48.

Davis, D. D. and C. A. Holt, *Experimental Economics*, Princeton, NJ: Princeton University Press, 1993.

Devetag, G., "Precedent Transfer in Coordination Games: An Experiment," *Economics Letters*, 2005, *89*, 227–232.

Duffy, J. and R. Nagel, "On the Robustness of Behaviour in Experimental 'Beauty Contest' Games," *Economic Journal*, 1997, *107*, 1684–1700.

Fehr, E. and S. Gächter, "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 2000, *90* (4), 980–994.

Forsythe, R., J. L. Horowitz, N. E. Savin, and M. Sefton, "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, 1994, *6*, 347–369.

Harbaugh, W. T., K. Krause, and S. G. Liday, "Bargaining by Children," 2003. working paper.

Harrison, G. W. and J. A. List, "Field Experiments," *Journal of Economic Literature*, 2004, *42* (4), 1009–1055.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, and H. Gintis, eds, *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from fifteen Small-Scale Societies*, Oxford, UK: Oxford University Press, 2004.

Ho, T., C. Camerer, and K. Weigelt, "Iterated Dominance and Iterated Best Response in Experimental "p-Beauty Contests"," *American Economic Review*, 1998, *88* (4), 947–969.

Hoffman, E., K. McCabe, and V. L. Smith, "Social Distance and Other-Regarding Behavior in Dictator Games," *American Economic Review*, 1996, *86* (3), 653–660.

Hoffmann, R. and J. Y. Tee, "Adolescent-Adult Interactions and Culture in the Ultimatum Game," *Journal of Economic Psychology*, 2006, *27* (1), 98–116.

Huyck, J. Van, R. Battalio, and R. Beil, "Tacit coordination games, strategic uncertainty, and coordination failure," *American Economic Review*, 1990, *80*, 234–248.

Knez, M. and C. Camerer, "Creating Expectational Assets in the Laboratory: Coordination in 'Weakest-Link' Games," *Strategic Management Journal*, 1994, *15*, 101–119.

Kovalchik, S., C. F. Camerer, D. M. Grether, C. R. Plott, and J. M. Allman, "Age and Decision Making: A Comparison between Neurologically Healthy Elderly and Young Individuals," *Journal of Economic Behavior and Organization*, 2005, *58*, 79–94.

Nagel, R., "Unraveling in Guessing Games: an Experimental Study," *American Economic Review*, 1995, *85* (5), 1313–1326.

Noveck, B. S., "Introduction: The State of Play," *New York Law School Law Review*, 2004, *49* (1), 1–18.

Oosterbeek, H., R. Sloof, and G. van de Kuilen, "Cultural differences in ultimatum game experiments: evidence from a meta-analysis," *Experimental Economics*, 2004, *7*, 171–188.

Riechmann, T. and J. Weimann, "Competition as a Coordination Device: Experimental Evidence from a Minimum-Effort Coordination Game," *European Journal of Political Economy*, 2008, *forthcoming.*

Rokeach, M., *The Nature of Human Values*, New York, NY: Free Press, 1973.

Roth, A. E., "Introduction to Experimental Economics," in J. H. Kagel and A. E. Roth, eds., *The Handbook of Experimental Economics*, Princeton University Press Princeton, NJ 1995, pp. 2–109.

_ , M. Prasnikar V. Okuno-Fujiwara, and S. Zamir, "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An Experimental Study," *American Economic Review*, 1991, *81* (5), 1068–1095.

Schotter, A., A. Weiss, and I. Zapater, "Fairness and Survival in Ultimatum and Dictatorship Games," *Journal of Economic Behavior and Organization*, 1996, *31*, 37–56.

Schwartz, S. H., "Value Hierarchies Across Cultures: Taking a Similarities Perspective," *Journal of Cross-Cultural Psychology*, 2001, *32* (3), 268–290.

_ , "A proposal for measuring value orientations across nations proposal for measuring value orientations across nations," in "European Social Survey Questionnaire Development Report" European Social Survey London, UK 2002.

Solnick, S. J., "Gender Differences in the Ultimatum Game," *Economic Inquiry*, 2001, *39* (2), 189–200.

Yee, N., "The Demographics, Motivations and Derived Experiences of Users of Massively-Multiuser Online Graphical Environments," *PRESENCE: Teleoperators and Virtual Environments*, 2006, *15*, 309–329.