Journal of Economic Behavior & Organization 72 (2009) 618–635

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Virtual world experimentation: An exploratory study

Thomas Chesney, Swee-Hoon Chuah, Robert Hoffmann*

International Centre for Behavioural Business Research, Nottingham University Business School, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB, United Kingdom

ARTICLE INFO

Article history:
Received 14 April 2008
Received in revised form 29 May 2009
Accepted 29 May 2009
Available online 12 June 2009

JEL classification:
C72
C88
C99
Z13

Keywords:
Virtual worlds
Experiments

ABSTRACT

We explore the scientific potential of virtual worlds for experimental economics in terms of the subject pools and experimental platforms they present. Our results offer tentative, qualified support for virtual world experimentation. Overall, the behaviour of virtual subjects recruited, incentivised and observed within *Second Life* across a range of five standard experimental games was not found to differ significantly from established standard results. In addition, we identify certain methodological opportunities and challenges which confront virtual world experimenters.

1. Introduction

Social scientists are becoming increasingly interested in virtual worlds, three-dimensional environments in which communities of networked individuals interact (Castranova, 2005; Bainbridge, 2007; Bloomfield, 2007). There are two reasons. First, the growing number of users and the scope and nature of socio-economic activity between them are seen as interesting phenomena that merit investigation in their own right (Castranova, 2005). Virtual worlds present evolving cultures with independent social institutions that are becoming more significant to society at large (Noveck, 2004). In economic terms, their evolution from specialised video game networks to general social platforms has generated a global industry of firms that leverage installed user bases for subscription fees, advertising opportunities or virtual support services (Cagnina and Poian, 2007). Many virtual worlds have evolving economies with fully convertible currencies as well as functioning financial, labour and product markets that are capable of producing a host of micro and macroeconomic phenomena (Guest, 2007).

Second, the computer technology underlying virtual worlds provides novel methods of conducting social science research (Bainbridge, 2007). To begin with, it facilitates the economical and large-scale recruitment of diverse subjects from different cultural-geographical and socio-economic groups for participation in interviews, focus groups, surveys or experiments. In addition, it affords control of the environment in which they decide and interact that can be used to manipulate decision conditions, observe behaviour and collect data. Conversely, however, both these features also present potential methodolog-

* Corresponding author. Tel.: +44 0115 846 66 02; fax: +44 0115 846 66 67.
E-mail addresses: thomas.chesney@nottingham.ac.uk (T. Chesney), swee-hoon.chuah@nottingham.ac.uk (S.-H. Chuah), robert.hoffmann@nottingham.ac.uk (R. Hoffmann).

0167-2681/$ – see front matter
doi:10.1016/j.jebo.2009.05.026

ical problems. As subjects, virtual world users may not reflect standard populations in terms of demographic or cultural characteristics and therefore may display different behaviours. The electronic interface that moderates communication and interaction between them precludes physiological signals and proximity that moderate economic behaviour in physical settings. Virtual world culture, social institutions and conventions that evolve as a result may shape economic interactions in ways that differ from traditional social settings. The anonymity of the interface may hamper quality control in the data collection process.

The current study is intended as a first, exploratory step towards the methodological issue. While virtual worlds may provide useful research tools for a number of social science disciplines, we concentrate here on their potential as platforms for designing and conducting economic experiments, an area which may be especially conducive to benefit from the new methods virtual worlds offer (Bainbridge, 2007; Castranova, 2006). Traditional experimental economics involves testing economic theories by observing the incentivised decisions of representative subjects under choice conditions systematically manipulated in laboratory settings. Virtual worlds may provide opportunities for methodological innovation here. The discipline has recently begun to broaden its scope by exploring new methods and applications outside the standard controlled laboratory environment commonly populated by Western student subjects. There are two related ways in which experimentalists are trying to improve the realism of the behaviour 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). Second, 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., 2004, e.g.). Virtual worlds 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. Second, virtual worlds 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. In this sense, virtual worlds may bridge the gap between laboratory experiments and field studies, allowing researchers to use representative subjects in more natural environments to study the relationship between the conditions of interaction and the evolution of social institutions in a controlled manner.

We assess to what extent virtual worlds can be used in this context. We approach the issue in two ways, by replication and by observation. First, virtual world experimentation can be a useful, alternative experimental tool to the extent that the results it generates for particular tasks and conditions are the same as those generated by traditional experimental methods. We assess this aspect by conducting virtual experiments with a range of standard tasks in standard conditions and comparing virtual subject behaviour with that of traditional pools reported in existing work. The suitability of virtual experimentation as an alternative would be supported to the extent that no differences are found. As the observed subject behaviour may be related to their underlying culture, demographics and values, we also used a survey instrument to collect data on these which can be compared to standard populations. The difference or similarity of virtual users to these provides additional insight into their suitability as experimental subjects representative of economic agents generally. This first part of our approach tests the scope virtual worlds hold for traditional economic experimentation, rather than for new avenues of experimental research they may promise. We conceive of it as measuring the 'output' of the virtual experimentation method. The second part of our approach is more qualitative and focuses on its 'input' side. This involves gathering informal insights about the practical feasibility of economic experimentation in virtual environments from the process of conducting experiments. We hoped to learn by observation to what extent virtual worlds can provide a suitable platform for experimental research generally, what the advantages and disadvantages are, and what modifications may be made to render virtual worlds more amenable to experimentation. This second part may also provide insights into what opportunities for new research approaches or methods virtual worlds hold.

The rest of the paper proceeds as follows. In the next section, we discuss the features of virtual worlds, their significance for experimentalists and our procedure of methodologically assessing them. The results we obtained are reported in Section 3. Section 4 discusses our general observations from the experiment in terms of the methodological issues we consider. The final section contains concluding remarks.

2. Virtual experimentation

2.1. Virtual worlds

While there is considerable variation between the many alternative virtual worlds that exist, they typically reproduce features of the physical world such as a three-dimensional topography containing virtual objects obeying simulated physical laws as well as the possibility of communication, social interaction and economic exchange between users virtually represented by *avatars*. We chose Second Life (SL, see Linden-Labs, 2008) as the virtual platform for our study. At the time of writing (November 2008), SL has over 15.7 million registered avatars.[1] Accounting for multiple and dormant registrations, there are an estimated one million regular users who spend over twenty million hours logged in per month. Between twenty and thirty thousand users are online at any one time. In terms of demographics, the majority of these are from populous

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

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

and industrialised countries including the USA, the UK, Germany, Brazil, France and Japan, with a median age of 36 and 57 percent 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 Fig. 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$) which can be bought from or sold to Linden Lab, the creators and owners of SL, at a relatively stable exchange rate of about 270 L$ per 1 U.S. dollar. A total of 5.3 billion L$ (U.S. $19.7 million) are currently 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.

2.2. *Experimental economics*

Virtual worlds 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)

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 who control the participating avatars. 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 virtual subjects generally. This may generate a sample bias that renders virtual experimentation inappropriate to test general economic theories. They 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.

2.3. *Experimental design*

The purpose of our study is to conduct experiments within SL to assess the overall feasibility of virtual experimentation. Our approach is to gauge to what extent the behaviour and values of virtual subjects conform to those of standard subjects. In

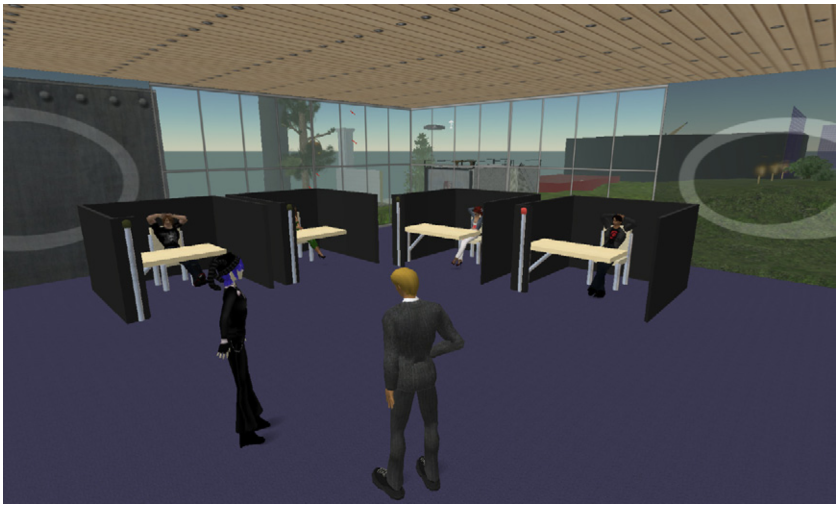


Fig. 2. A typical experimental session in progress. The experimenters' avatars are standing.

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. All our experimental sessions were conducted during standard GMT working hours between July and November 2007. Experimental instructions are available upon request.

In the recruitment stage, we solicited participation by approaching online users *in situ* immediately prior to a particular experimental session in the following manner. Half an hour before a scheduled experimental session, we used a search feature in the SL-inteface to identify the currently busiest locations in terms of number of avatars present (excluding locations with an adult thematic focus). Next, each of the three experimenters used their avatar to access one of these locations and to address groups of avatars gathered there using public IM with a standardised recruitment message. This message was in English and stated our institutional affiliation and general information about the nature of the task, its duration and incentivisation. Whenever interested users responded, we answered any additional questions and informed volunteers of the time and venue of the session. This process was repeated for a number of locations and avatar groups in each until the recruitment of the desired number of participants was complete. The thirty minute period was in almost all cases sufficient to recruit between four and seven subjects.

Participants were transported to our virtual experimental laboratory in a dedicated virtual building with controllable access rights and purpose-built laboratory furniture. In the briefing stage, subjects who have arrived (typically in groups between two and seven depending on the task) were given virtual documents containing general information on experimental etiquette, anonymity, confidentiality and incentivisation. The two to three experimenters present at all times communicated with subjects using either public or private (i.e. one-to-one) IM. Once they have finished reading the briefing documents, subjects were asked to occupy cubicles that were purpose-built to restrict their vision and communication in order to prevent collusion between them. They were then given virtual documents containing the experimental instructions and a comprehension quiz. The decision task stage commenced after all subjects completed the quiz successfully. Experimenters instructed individually when subjects were initially unable to do so. Subjects communicated their decisions to the lead experimenter and received feedback via private IM. Next, in the survey stage, subjects were sent the URL of a webform containing a values survey as well as some demographic questions which they had to fill out. 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 Fig. 2.

Table 1 provides some general information about the decision tasks of our experiments. Our choice of tasks was guided by our objective to assess whether a virtual 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 the ultimatum (UG), dictator (DG), public good (PGG), guessing (GG) and minimum effort (MEG) games.

Table 1
Summary statistics for experimental games and survey.

Task	UG	DG	GG	PGG	MEG	ESS
Subjects (N)	64	60	31	32	31	113
Subjects per session (n)	4–5	4–5	3–7	4	5–6	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 (r) or questions	1	1	10	10	10	21

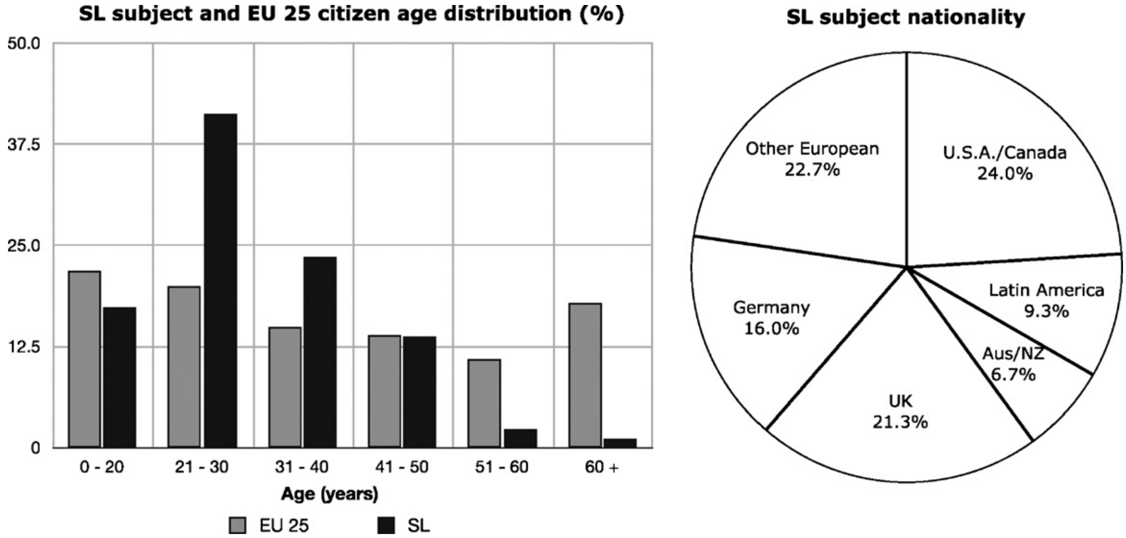


Fig. 3. Age and nationality distribution of SL-subjects.

Previous experimental results for all of these standard games abound for a variety of conditions as well as demographic and cultural groups and provide ready benchmarks for our own results. They also permit eliciting a broad spectrum of different types of strategic choice. In the following, we do not explain or analyse these standard games in detail, but report data from our and those previous studies most appropriate for comparison. We also report results from tests of differences in means, medians and overall distributions between them using t-tests, Mann–Whitney U (MW) and Kolmogorov–Smirnov Z (KS) tests, respectively. While means tests can indicate differences between the overall behavioural propensities in two pools of subjects, distribution tests can also reveal differences in the incidence of a variety of behaviours when average behaviour does not differ. For experimental tasks with multiple decision rounds, we also used regression analysis to test for differences with previous results. In particular, we pooled available data from our own and the previous study used as a comparator and estimated the following regression equation:

$$Y_i^t = \alpha + \beta Y_i^{t-1} + \gamma X_i + \delta n_i \tag{1}$$

where Greek letters represent constant and parameters, Y is observed behaviour, t the task round, n experimental group size and X a dummy variable for the comparator study. No differences between SL and comparator study behaviour exist to the extent that the coefficient for the latter variable is insignificant. The inclusion of the lagged variable on the right-hand side was intended to reduce omitted-variable bias in our model. In particular, it is well established that simple learning processes may explain some changes in behaviour over time in specific game and choice contexts (see, e.g., Camerer, 1987; Erev and Roth, 1998). As a result, we opted for a specification similar to a partial adjustment model, where the behaviour in the current period is adjusted to that in the previous one. These kinds of dynamic model have been previously applied to the three games for which we seek to estimate behaviour, i.e. the PGG (Healy, 2006), the GG (Kurz, 2008) as well as the MEG (Crawford, 1995).

It should be noted that our design makes no provision for establishing a control treatment by replicating our virtual experiments in a standard physical setting with otherwise identical experimental parameters. While this alternative has certain advantages, our approach was to rely instead on the replicability of existing studies and to design virtual experiments that mirror their task conditions such as to permit using their results as a comparator.

An additional 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.

3. Experimental results

3.1. Subject demographics

Subjects' basic demographical data are summarised in Fig. 3. The average age of respondents was 32, with the youngest at 18 and the oldest at 64. Compared with the general population of the European Union (EU), the age range 20–40 years was over represented, an expected result given the technological and cultural status of virtual worlds. In line with SL generally,

Table 2
Summary statistics of ultimatum game offers (in percent of the U.S. $ stake) and rejections for $N/2$ subject pairs in SL as well as in selected previous studies.

	SL	HMSS	RPOZ 1	RPOZ 2	CHJW
$N/2$	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	21.0	9.5
Rejections					
Percent of offers <20 percent	33.33	–	–	50	–
Percent of all offers	6.25	8.3	22	24	15

most subjects 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 bias sample selection in terms of time zone. In terms of gender, exactly half of our respondents were male.

3.2. Ultimatum game

Separate sessions with UG-proposers and responders were conducted on 6, 25 and 26 July 2007. 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.

Although there is little evidence for stake size effects in the UG (see Camerer, 2003), we aimed for comparability by using a stake in the U.S. $10–15 interval used in many previous studies, as well as for easy mental divisibility. Theory predicts that, because instrumentally rational responders should accept any share of the stake, rational proposers should offer the minimum. However, proposers in previous studies offer in the region of 42–48 percent (see Table 2.2 in Camerer, 2003), reflecting a mixture of altruistic and strategic thinking on their part (Forsythe et al., 1994). In standard task conditions and subject pools recruited in industrialised nations, UG-results are relatively robust. Roth et al. (1991) (RPOZ) found little difference between offers made by urban subjects recruited in the U.S. (RPOZ 1), Tokyo (RPOZ 2), Yugoslavia and Israel. However, alternative cultural and demographic characteristics can generate differences (Camerer, 2003; Oosterbeek et al., 2004). Buchan et al. (1997) and Chuah et al. (2007) (CHJW) identified slightly but significantly higher offers of South-East Asian subjects potentially linked to their collectivist values. Henrich et al. (2004) found a much wider range of offers (between 25 and 57 percent) in a series of experiments with traditional, small-scale societies across the developing world.

Table 2 reports summary statistics of UG bargaining by SL-subjects compared with behaviour reported by RPOZ (1 and 2), by Hoffman et al. (1994) for U.S. subjects (HMSS) and by CHJW for UK subjects. The SL mean offer is 45.73 percent of the stake with a modal offer of half. These central tendencies in the proposals are very similar to those reported for comparable samples. Fig. 4 shows the distributions of offers 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. Statistical tests bear these observations out. As the UK formed the largest national group among our subjects (see Section 3.7), we used UK subject data from CHJW as a comparator for our findings. No differences in the mean ($t = 0.216$, $p = 0.829$), median ($U = 2706.5$, $p = 0.422$) or distribution ($Z = 0.595$, $p = 0.870$) of offers were found between their and SL subjects.

3.3. Dictator game

DGs were also conducted in separate sessions for proposers and responders, except that responders were not given the opportunity to accept or reject offers. The sessions were conducted on 27 and 31 July 2007. As a number of previous studies employed stakes divisible by 10, and since stake size effects are not noticeable between studies with significantly different stakes (see Table 3), we opted for a stake size of 1000 $L (U.S. $3.90). The DG was originally conceived as a way of separating altruistic and strategic motives in UG-offer behaviour (Forsythe et al., 1994). While instrumentally rational players should keep all of the stake, experimental subjects offer in the region of 20–35 percent 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.

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). Fig. 4 displays the distributions 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 percent 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 percent. Of particular interest to us is the study by CBV, who identified marked differences in DG offer levels based on age and experimental location (bottom panel of Fig. 4).

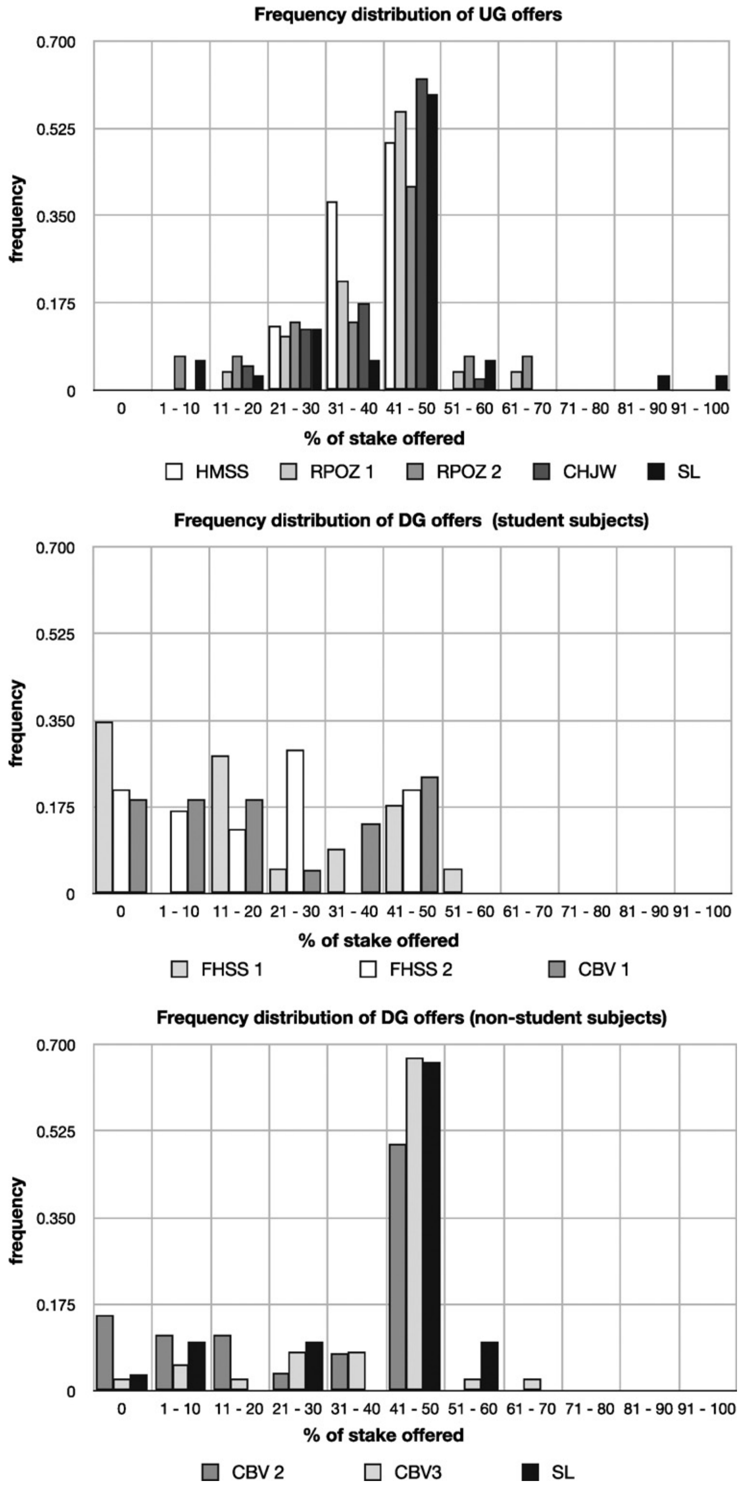


Fig. 4. Distribution of UG and DG offers in SL as well as in selected previous studies.

In their study, they compare offers made by students (average age: 19 years) in standard college settings (CBV 1), by older community college students (27, CBV 2) and by workers in a warehouse setting (37, 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 reflect the greater average age of our subjects (see Section 3.7) and the fact that DG-offers are sensitive to age (Harbaugh et al., 2003). Previous and current DG-results pertaining to older subjects

Table 3
Summary statistics of dictator game offers (in percent of the U.S. $ stake) for $N/2$ subject pairs in SL as well as reported in selected previous studies.

	FHSS 1	FHSS 2	CBV 1	CBV 3	SL	CBV 2
$N/2$	24	45	21	26	30	37
Stake	10	5	100	100	3.90	100
Offers						
Mean	24	24	25	33	43	45
Mode	30	0	50	50	50	50
Median	25	20	20	45	50	50
St. dev.	17.68	20.44	19	20	16.17	12

Table 4
Summary statistics of public good game contributions (in percent of the U.S. $ stake averaged over r rounds) for N subjects playing in groups of n in SL as well as reported in selected previous studies. Stakes are given as U.S.$-values of tokens subjects were asked to allocate per round.

	A (88)	A (95)	FG	SL
N	30	40	24	32
n	5	5	4	4
r	10	10	10	10
Stake	0.50	0.60	0.86	1.50
α	0.5	0.5	0.4	0.4
Contributions				
Mean	33.20	44.09	37.94	50.34
Median	32.00	42.50	40.25	45.63
St. dev.	21.65	27.47	16.89	22.54

are shown in the bottom panel in Fig. 4. 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.

The age similarity between warehouse workers in CBV 3 to our own SL-subjects provides us with an appropriate benchmark for the comparison of DG-behaviour. No statistically significant differences were found between the means ($t = -0.700$, $p = 0.485$) medians ($U = 981.5$, $p = 0.823$) and distributions ($Z = 0.383$, $p = 0.999$) of DG offer data in these two pools.

3.4. Public good game

The PGG sessions were conducted on 25 October and 2 November 2007. In them, subjects in groups of $n = 4$ were asked to divide a stake of 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 was repeated $r = 10$ times. The parameter values for n, r and α were chosen with comparability with other studies in mind (see Table 4). The PGG 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 private fund. In repeated PGGs, players decisions may be guided both by strategic considerations of reciprocation and purely altruistic motives. 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 as the game is repeated. The studies reporting 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, 1995) (A (88) and A (95)) as well as Fehr and Gächter (2000) (FG), who used values for parameter a of 0.5, 0.5 and 0.4, respectively. Table 4 reports summary statistics of SL-PGG behaviour compared to subjects in these three.

The top panel in Fig. 5 shows the average contribution to the group fund subjects made in SL and in the three previous studies over ten rounds. 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. (2004) report on PGGs played with traditional society subjects in many continents and find mean contribution rates to vary between 22 and 65 percent. The SL subjects differ from standard college students in a number of ways, age being one. Our result may also be due to the apparent greater altruism of SL-subjects compared with students we observed in the DG.

For our statistical tests of PGG behaviour, we chose A (95)'s Western student subject data as a benchmark. It should be borne in mind that this experiment differs from our study in two ways; the differences in experimental platform we are assessing, and the differences in subject demographics. We performed mean, median and distribution tests between the offers for each of the ten rounds played by A (95) and SL subjects (see Table 5). Only one of the resulting thirty test statistics was significant ($Z_{n=10} = 1.370$, $p = 0.047$). As the repeated testing procedure amplifies the probability of Type I errors, we also estimated Eq. (1) to compare the two data sets. The factor n could not be entered due its insufficient variation in the data

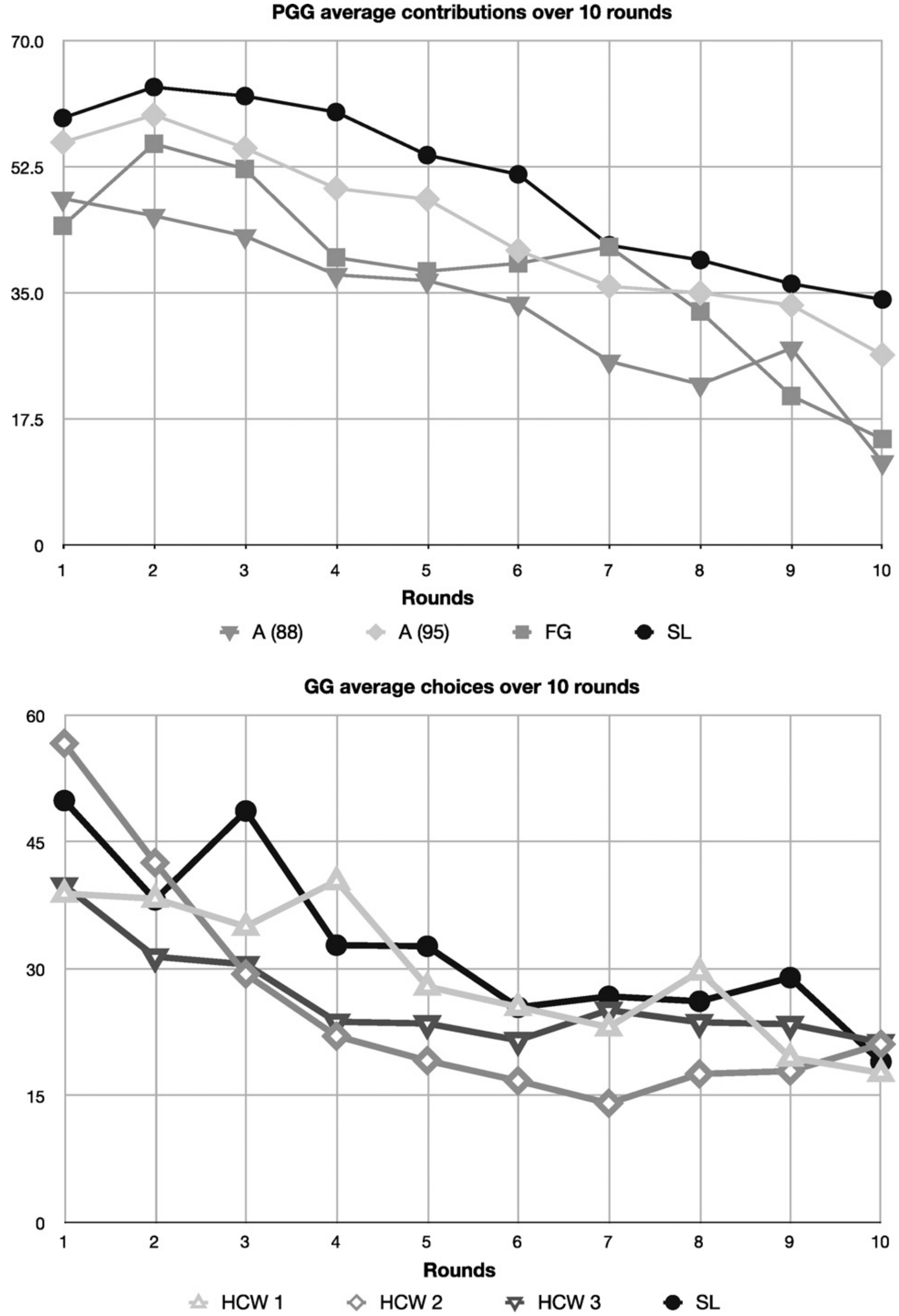


Fig. 5. Average subject decisions in GG and PGG over $r = 10$ rounds in SL and selected previous studies.

set. The regression results are given in Table 6 and show an insignificant coefficient for *X*, leading us to conclude that no behavioural differences are in evidence.

3.5. Minimum effort game

The MEG sessions were conducted between 16 and 21 November 2007. In them, groups of $n = 5$ to 6 subjects were asked to choose an integer in the interval [1,7] and informed that payoffs would be determined by the smallest number chosen within the group according to the payoff matrix adapted from Van Huyck et al. (1990) (VBB) and shown in Table 7.

Table 5
Test statistics for differences in mean (t), median (U) and distribution (Z) of behaviour between SL subjects and those in selected previous studies for $r = 10$ rounds. Corresponding p-values are given in parentheses. The symbols *, ** and *** denote significance at the 10, 5 and 1 percent levels, respectively.

Task	r	t	MW U	KS Z
PGG	1	0.431 (0.667)	619.0 (0.808)	1.054 (0.216)
	2	0.499 (0.619)	611.5 (0.743)	0.949 (0.329)
	3	0.864 (0.391)	572.0 (0.436)	0.764 (0.603)
	4	1.231 (0.223)	536.5 (0.235)	1.001 (0.269)
	5	0.697 (0.488)	567.0 (0.403)	0.817 (0.517)
	6	1.231 (0.222)	534.5 (0.227)	0.870 (0.436)
	7	0.673 (0.503)	573.5 (0.446)	0.738 (0.648)
	8	0.568 (0.572)	544.5 (0.274)	1.370 (0.047**)
	9	0.372 (0.711)	567.5 (0.405)	0.817 (0.517)
	10	0.926 (0.358)	539.5 (0.240)	1.133 (0.153)
MEG	1	1.482 (0.141)	982.5 (0.139)	0.895 (0.452)
	2	1.218 (0.226)	1023.0 (0.236)	0.833 (0.491)
	3	1.927 (0.057)	931.0 (0.070*)	1.109 (0.171)
	4	2.660 (0.009***)	822.5 (0.011**)	1.353 (0.051*)
	5	1.449 (0.150)	986.0 (0.153)	0.713 (0.690)
	6	1.382 (0.170)	990.0 (0.162)	0.983 (0.289)
	7	1.571 (0.119)	955.5 (0.102)	0.888 (0.409)
	8	0.785 (0.435)	1059.0 (0.351)	0.951 (0.326)
	9	0.518 (0.606)	1073.5 (0.406)	1.042 (0.228)
	10	2.364 (0.020**)	841.0 (0.014**)	1.347 (0.053*)
GG	1	1.798 (0.078*)	219.0 (0.079*)	0.928 (0.355)
	2	0.091 (0.928)	305.0 (0.923)	0.478 (0.976)
	3	2.195 (0.033**)	212.5 (0.060*)	1.226 (0.099*)
	4	−1.090 (0.281)	268.5 (0.423)	0.821 (0.510)
	5	1.003 (0.321)	289.5 (0.692)	0.664 (0.771)
	6	0.032 (0.974)	280.5 (0.569)	0.703 (0.706)
	7	0.538 (0.593)	283.0 (0.602)	0.664 (0.771)
	8	−0.552 (0.583)	278.5 (0.543)	0.652 (0.788)
	9	2.107 (0.041**)	250.5 (0.250)	1.277 (0.077*)
	10	0.279 (0.781)	292.5 (0.735)	1.063 (0.209)

Each group played ten rounds of this game. Again, these parameter values are standard to the extent that they have been adopted by the majority of previous studies. 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' reasoning. 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 also be late. After a number of meetings, such expectations may become increasingly self fulfilling as general punctuality disintegrates. Previous experimental evidence shows this kind of convergence on payoff-dominated

Table 6
Regression results for experimental behaviour across three tasks in SL and one comparator study, respectively. The symbols *, ** and *** denote significance at the 10, 5 and 1 percent levels, respectively.

PGG	Estimate	t-Value	p-Value
Constant	14.24	6.39	0.000***
Y_{t-1}	0.67	22.63	0.000***
X	−2.47	−1.15	0.252
	R^2 (adj.) = 0.45	F = 260.69	p = 0.000***
MEG	**Estimate**	**t-Value**	**p-Value**
Constant	1.22	1.36	0.17
Y_{t-1}	0.61	24.03	0.000***
n	0.06	0.43	0.664
X	−0.34	−2.12	0.034**
	R^2 (adj.) = 0.39	F = 204.55	p = 0.000***
GG	**Estimate**	**t-Value**	**p-Value**
Constant	26.84	11.38	0.000***
Y_{t-1}	0.25	8.09	0.000***
n	−1.11	−2.82	0.005**
X	−1.36	−0.81	0.416
	R^2 (adj.) = 0.09	F = 30.42	p = 0.000***

Table 7
MEG payoff matrix (in L$). The first column represents player choices which, combined with the smallest choice in the group, determines payoffs. Dashes denote logically impossible outcomes.

	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 8
Summary statistics of minimum effort game choices over *r* rounds for *N* subjects playing in groups of *n* in SL as well as reported in selected previous studies. Stakes are given as U.S.$-value of payoff associated with unique Pareto-efficient outcome.

	VBB	KC	BGN	DT	SL
N	107	30	42	77	31
n	14–16	6	7	7	5–6
r	10	5	10	14	10
Stake	1.30	1.30	1.30	1.82	1.46
Choices					
Mean	2.72	2.87	3.65	3.75	4.44
Median	2.50	2.80	2.40	3.60	4.60
St. dev.	1.30	1.07	1.34	1.57	1.51

outcomes to be dependent on the size of the group, the size of payoffs and information players receive about the choices of others.

Fig. 6 shows the round-to-round changes in the choices and minimum choices averaged over experimental groups in SL and comparable previous studies of Knez and Camerer (1994) (KC), Bornstein et al. (2002) (BGN), Devetag (2005) (DT) and VBB. Table 8 reports summary statistics of SL-PGG behaviour compared to subjects in these studies. All these studies used VBB's payoff matrix and had groups between 5 and 7 subjects except VBB, which had groups of 14–16. The figure shows similar declines in choices in all these studies. On the other hand, there appears to be greater variability in the overall level of average choices, with SL-averages appearing higher than those in other studies.

We used the data reported by DT for the comparison with SL-observations. In terms of means, medians and distributions for $r = 10$ rounds, round four and ten behaviours were different in terms of all three at the 10 percent-level of significance (see Table 5). With one exception ($U_{n=3} = 931.0$, $p = 0.070$), the other twenty-four tests were negative, suggesting no differences exist in the rounds concerned. Again, we regressed Eq. (1) for the combined data set (Table 6). The results show that at the 95 percent significance level, our data are different to those of DT as the coefficient for X is significant ($p = 0.034$). It should be noted that the same model also yields differences between the data of DT and BGN ($p = 0.084$) as well as between SL and BGN ($p = 0.002$). As a result, for the MEG, these findings do not provide firm conclusions about the ability of virtual world experimentation to replicate laboratory results. The two comparator experiments differ from ours in an additional, demographical dimension and also differ from each other in terms of results. The reason may lie in greater general variability in MEG-behaviour due to the presence of multiple equilibria.

3.6. Guessing game

The GG sessions were conducted on 8 and 15 November 2007. In them, $n = 3$ to 7 subjects were asked to choose integers in the interval [0,100] and informed that the subject with a response closest to $g = 0.7$ times the average of all choices would receive L$200 (U.S. $0.75). Ties were resolved by dividing this sum among the winners. Each group of subjects played $r = 10$ rounds of this game.

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

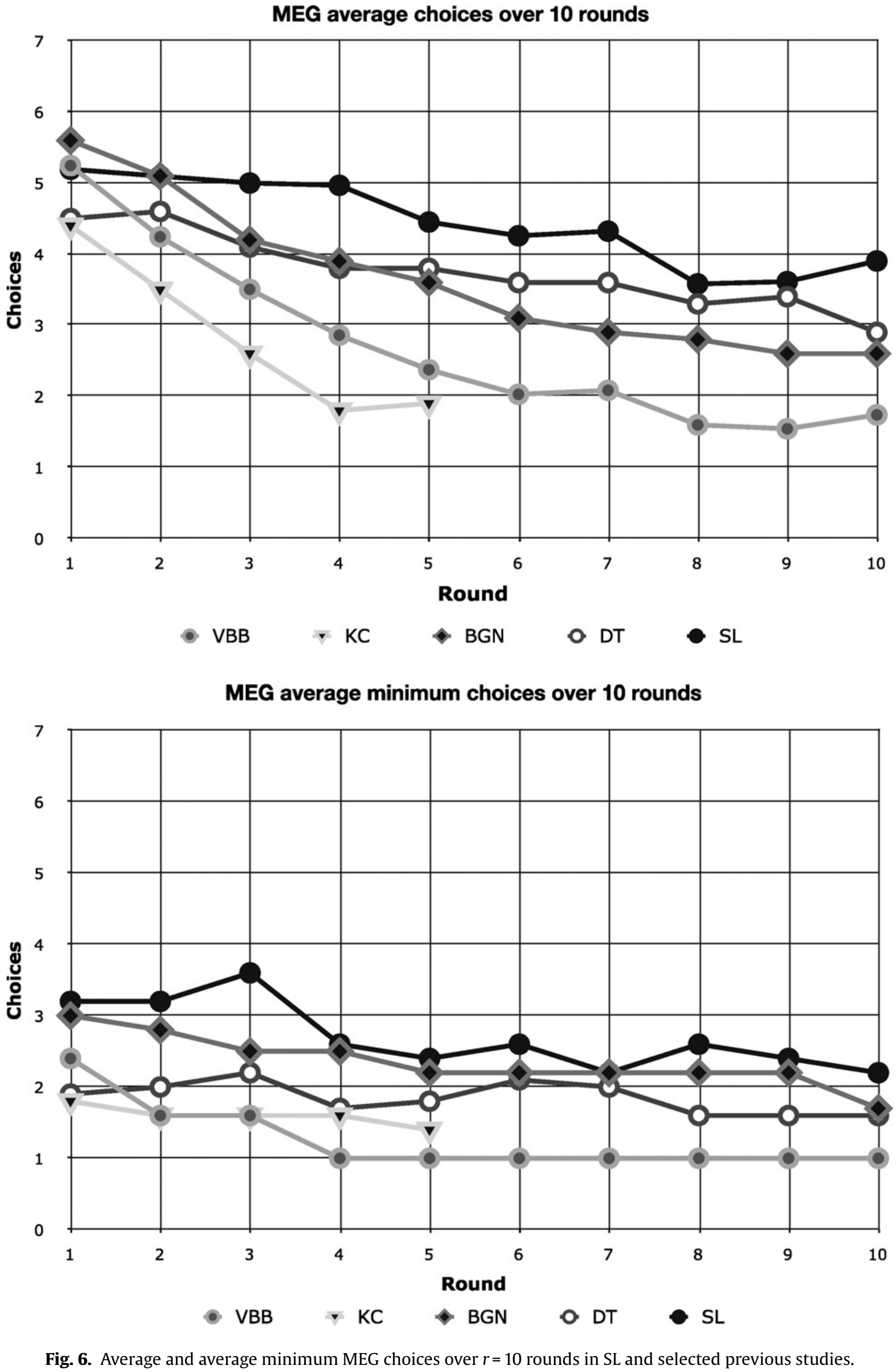


Fig. 6. Average and average minimum MEG choices over $r = 10$ rounds in SL and selected previous studies.

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 9 shows statistics concerning subjects' choices in single or first rounds of repeated games played in groups of different sizes with a parameter $g = 0.7$. 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 g-value. Finally, HCW 3 was composed of inexperienced 7-subject group players. In all HCW-treatments, the winning subject received 50 Singapore cents (ca. U.S. \$0.34). The U.S. study of Kovalchik et al. (2005) (KCGPA) compares one-round choices by college students (KCGPA 1) with those of mentally healthy senior citizens with an average age of 82 (KCGPA 2). Our

Table 9
Summary statistics for round 1 GG choices in n-subject pools in SL as well as reported by Camerer (2003), Camerer (1997) and Kovalchik et al. (2005). The percentage of subjects choosing 0 is given by %0.

Subjects	Mean	Median	St. dev.	%0	N
Caltech students	21.88	23.00	10.35	0.07	27
Portfolio managers	24.31	24.35	16.15	0.08	26
Economics PhDs	27.44	30.00	18.69	0.13	16
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
Singaporean students (HCW 1)	36.45	35.00	24.28	0.00	21
German students	36.73	33.00	20.21	0.03	67
Senior citizens (KCGPA 2)	37.00	33.00	17.46	0.00	50
University CEOs	37.81	36.50	18.92	0.03	73
Wharton students	37.92	35.00	18.84	0.00	35
Singaporean students (HCW 3)	39.78	35.00	25.46	0.02	49
SL	50.00	56.00	27.10	0.00	31
Singaporean students (HCW 2)	58.27	50.00	26.98	0.05	21

experimental settings of group size, g-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 bottom panel in Fig. 5 shows mean choices over ten rounds among SL-subjects and Singaporean students (HCW). Table 9 reports summary statistics of SL-GG behaviour compared to subjects in this study. Our subjects did appear to converge towards the equilibrium at similar rates to the latter. The frequency distribution of individual SL-choices over all ten rounds is displayed in Fig. 7, along with the corresponding data for HCW 1 reported in Ho et al. (1998) (p. 955, Fig. 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 Fig. 5.

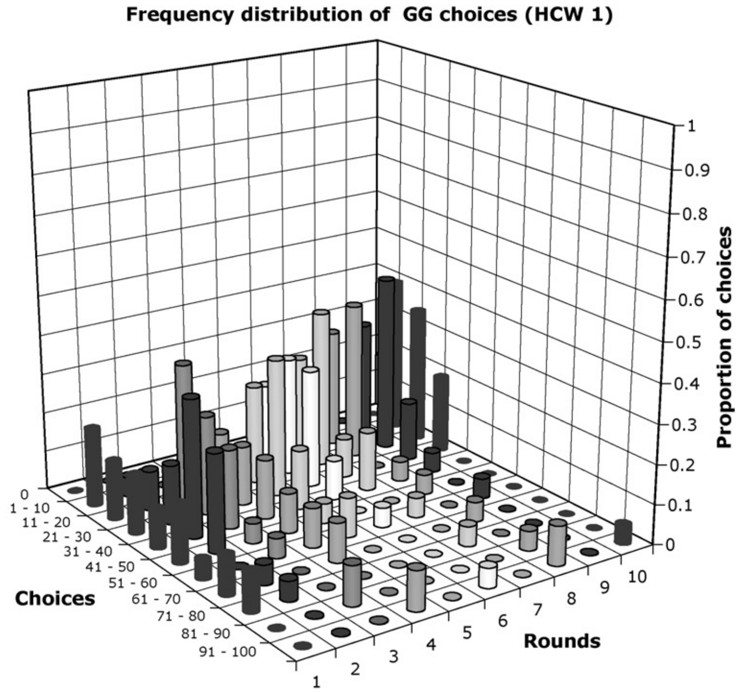
GG data generally show divergence in first-round average choices. Part of the reason may be the role that players' common knowledge of rationality has in equilibrium reasoning. Lower choices are not merely associated with greater strategic sophistication among players, but also with greater expectations 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 who 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.

We compared the means, medians and distributions of SL choices with HCW 1 over $r = 10$ rounds (see Table 5). Rounds 1, 3 and 9 show differences in all three dimensions. In total, seven of the thirty tests were positive, most only at the 10 percent-significance level. Table 6 shows the regression results for Eq. (1) pooling SL data with HCW 1 and 3. The latter study was not used for the tests as its larger subject group size rendered it inappropriate for a direct comparison; however, we were able to control for that difference using variable n in the regression. The results show an insignificant coefficient for X ($p = 0.416$). We conclude differences are not in evidence between the data sets.

3.7. 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 values questionnaire, a well-tested instrument for identifying ten universal value dimensions (listed in Fig. 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 webform 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.

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. 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. Fig. 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 invited randomly by automated email from the experimental subject database maintained by the *Centre for Decision Research and Experimental Economics*. For comparative purposes, we follow the ESS practice of presenting averages of ipsative scores, i.e. an individual's



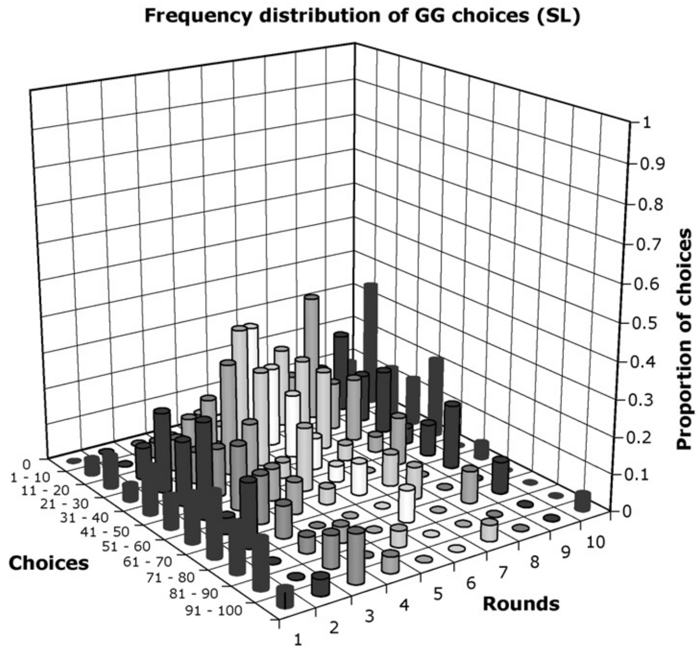


Fig. 7. Subject choice frequency distributions over $r = 10$ rounds (group size 3, $p = 0.7$) in HCW 1 and SL.

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 are shown along the horizontal axis of Fig. 8. They have established empirical interrelationships that are commonly used to reduce them to two basic dimensions shown along the two respective axes in Fig. 9. The first dimension, *self-transcendence v. self-enhancement*, encompasses six values: hedonism, stimulation and self direction relative to tradition, conformity and security. The former three values express underlying motivations such as pleasure, sensuous gratification, excitement, novelty and independence, while the latter express respect and acceptance of norms, self-restraint and harmony. The remaining four universal values are contained in the the second dimension, *openness to change v. conservatism*. It weighs the values of universalism and benevolence against those of power and achievement. The

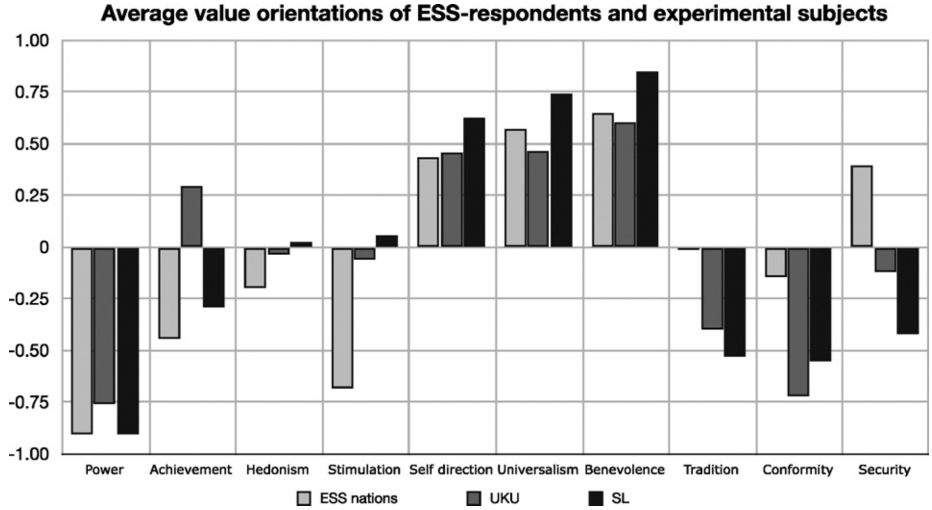


Fig. 8. Average orientations of ESS-respondents (ESS), SL and UK student subjects (UKU) according to Schwartz' ten value dimensions.

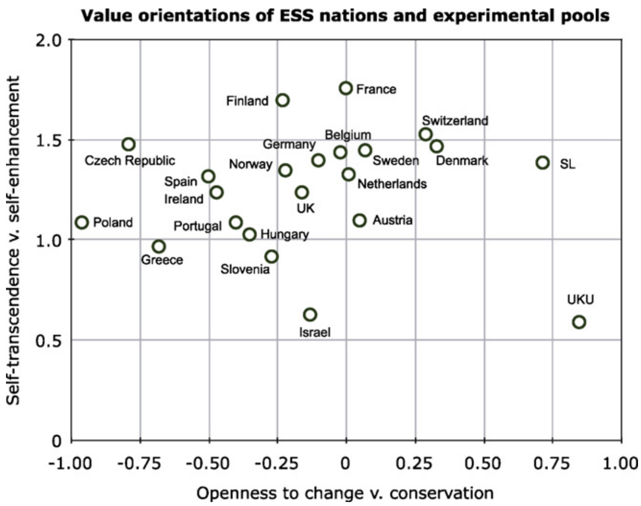


Fig. 9. Average orientations of ESS-respondents by nationality, SL and UK student subjects (UKU) according to Schwartz' two composite value dimensions.

former two values express motivations including tolerance and care for the welfare of others, while the latter two encompass social status, personal success and dominance over others. Fig. 9 plots nations and subject pools according to the two overall dimensions.

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 percent with each other, and respectively at about 70 and 64 percent with the averaged overall ESS-orientation of EU respondents. By comparison, individual national samples within ESS correlate with the average EU-values profile at about 94 percent. The graph shows a relatively small distance between randomly sampled individuals from European nations to SL-users and UK students. The students place a greater importance on the factors underlying self-enhancement, as can be verified in Fig. 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.

4. Methodological discussion

Our experience of conducting experiments in SL suggests a number of advantages and disadvantages of virtual experimentation generally as well as practical steps to adapt the platform for experimental purposes.

It was possible, with little organisation and preparation, to recruit subjects *in situ* in the numbers we could manage within the SL-interface. SL's features make it simple to create and maintain a database of subjects for future use. On the other hand, this procedure is prone to biased sample selection on the basis of choosing busy recruitment locations, of solicitation, in the recruitment language, time and institutional affiliation we used. In addition, the relative anonymity that avatars confer on subjects makes it difficult in practice to prevent financially motivated repeat participation or the recruitment of unfit (tired or intoxicated) or non-eligible or non-targeted subjects. While these issues may not be completely resolvable, we attempted to mitigate both repeat and unsuitable participation by disqualifying avatars using the following criteria. First, to avoid repeats, we excluded avatars who participated previously, who were created after the first experimental session or who made unsolicited approaches to us. To avoid unsuitable participants, we also excluded avatars less than a month old and potentially insufficiently familiar with the SL-environment, avatars referred by previous subjects who may have prior knowledge of the task, and avatars representing users who appeared to be in an unfit state. An additional identity issue both in our and in other virtual world studies concerns the potential for a disparity between user and avatar characteristics. For many users, the attraction of SL consists of the potential for using an avatar to assume a new and different identity. While our study was designed to elicit the behaviour and values of users and not avatars, we cannot be certain to what extent this was practised by subjects responding through their avatars.

Our demographical and values survey shows that virtual worlds provide opportunities for recruiting subjects who are demographically more representative than university students. In addition, targeting particular types of individuals is possible within those groups represented in virtual worlds, such as particular nations. Clearly, some groups are currently not sufficiently represented in virtual worlds, including individuals from smaller and traditional societies. However, 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 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 delivered easily. Many users regularly participate in paid online activities for returns which are modest compared with those of standard economic experiments. In addition, the developed markets for virtual objects provide alternative in-kind incentives.

While the computerised interface of SL provides an economical experimental environment that is well suited for data generation, collection and storage, it also has certain disadvantages. Communicating with subjects using IM makes it difficult to deal with more than a handful per session. In addition, private IM makes it hard to detect collusive behaviour or conferring amongst subjects. While is it not possible to override the communication mechanisms of SL, we developed virtual laboratory furniture that alerts the experimenter to the potential for clandestine communication between subjects (visible in Fig. 2). In particular, upon entering the virtual laboratory, subjects were asked to sit in cubicles and to enter *mouselook*, a SL-mode under which avatars are restricted to frontal vision and where private IM is suspended, in line with standard experimental conditions. Once activated, the furniture indicates whenever a subject suspends the mouselook mode and is therefore able to use private IM. While this furniture assured discipline in practice, it is in theory possible for experts to circumvent such mechanisms. On the other hand, this requires not only significant expertise on the part of a subject, but matching skills of and prior collusion with another subject present in order to establish a clandestine communication channel. Another problematic issue is establishing subject trust in the experimenters. Because of the nature of virtual worlds, it is difficult to convince subjects of the genuine nature of the experiment and incentivisation. A further problem involves the potential for disruption of experimental sessions by other users. This, however, may be controlled by restricting access to the virtual laboratory.

The absence of physical signals and presence in virtual worlds creates clear differences between virtual and physical experimental conditions. Virtual experiments preclude physical presence that may influence behaviour through involuntary non-verbal communication that reveals emotional states. In addition, the potential for anonymity means that the social consequences of virtual behaviour are different to those in physical laboratories. These factors may limit the comparability of virtual and physical experimental results in many cases. Virtual 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.

5. Conclusion

Despite the non-standard nature of the SL-subject pool and certain imperfections of the experimental environment that it provides, we were unable to detect significant and systematic overall differences between their behaviours and those observed in traditional settings. In particular, given SL-users' demographics in terms of age and cultural background, behaviour closely matched expectations based on a host of existing experimental evidence for a range of five important games. These results suggest tentatively that virtual world economic phenomena are based on similar behavioural regularities observed in standard economic settings and can be tested experimentally within the virtual environment.

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 virtual world demographics dispels the popular notion that they 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.

It should be noted that our study was not designed to provide support for or against virtual world experimentation as a method in absolute terms. Instead, we adopted a less ambitious research question regarding its ability to reproduce the results of traditional experimentation in physical laboratories with standard subjects. As a result, the absence of observed behavioural differences between the two environments does not necessarily make a case for virtual experiments *per se*, but rather suggests they may be a valid alternative to traditional method, subject to similar methodological advantages and limitations. Conversely, the presence of such differences would not necessarily invalidate virtual experimentation to the extent that the standard physical laboratory method is not without imperfections. As a result, these methodological issues remain and may benefit from renewed debate in the context of virtual experimentation.

While the above suggests that virtual world experimentation has potential as an economical and practical alternative to standard laboratory experiments, there are certain disadvantages associated with virtual worlds as experimental platforms which suggest that their suitability depends on the type of experiment planned. For instance, studies that consider the effects of physical signals or depend on recruiting specific types of subjects will find little value in virtual experimentation. On the other hand, suitably adapting experimental procedures to the virtual world environment makes it possible to effectively and cheaply recreate many standard decision tasks. In addition, virtual world users appear to constitute suitable subject pools to the extent that they display many of the economic behaviours associated with standard subjects. The future development of this technology will further increase the sophistication of the virtual experimental platform.

Acknowledgements

We are grateful to Simon Gächter, Sourafel Girma, Martin Sefton and Chris Starmer for a number of helpful suggestions.

References

Anderhub, V., Müller, R., Schmidt, C., 2001. Design and evaluation of an economic experiment via the internet. Journal of Economic Behavior and Organization 46 (2), 227–247.
Andreoni, J., 1988. Why free ride? Strategies and learning in public goods experiments. Journal of Public Economics 37, 291–304.
Andreoni, J., 1995. Cooperation in public-goods experiments: kindness or confusion? American Economic Review 85 (4), 891–904.
Bainbridge, W.S., 2007. The scientific research potential of virtual worlds. Science 317, 472–476.
Bloomfield, R.J., 2007. Worlds for study: invitation—virtual worlds for studying real-world business (and law, and politics, and sociology, and. . .). SSRN eLibrary.
Bornstein, G., Gneezy, U., Nagel, R., 2002. The effect of intergroup competition on group coordination: an experimental study. Games and Economic Behavior 41, 1–25.
Buchan, N.R., Johnson, E., Croson, R.T.A., 1997. Culture, power and legitimacy: contrasting influences on fairness beliefs and negotiation behavior in Japan and the United States, University of Wisconsin Department of Marketing Working Paper.
Cagnina, M.R., Poian, M., 2007. How to compete in the metaverse: the business models in second life. Tech. rep. URL http://ssrn.com/abstract=1088779.
Camerer, C.F., 1987. Do biases in probability judgment matter in markets? American Economic Review 77 (5), 981–997.
Camerer, C.F., 1997. Progress in behavioral game theory. Journal of Economic Perspectives 11 (4), 167–188.
Camerer, C.F., 2003. Behavioral Game Theory. Princeton University Press, Princeton, NJ.
Carpenter, J.P., Burks, S., Verhoogen, E., 2005. Comparing students to workers: the effect of social framing on behavior in distribution games. In: Carpenter, J.P. (Ed.), Field Experiments in Economics, vol. 10. Elsevier, pp. 261–290, (Ch. 7).
Cason, T.N., Mui, V.L., 1998. Social influence and the strategy method in the sequential dictator game. Journal of Mathematical Psychology 42, 248–265.
Castranova, E., 2005. Synthetic Worlds: The Business and Culture of Online Games. University of Chicago Press, Chicago, IL.
Castranova, E., 2006. On the research value of large games: natural expriments in Norrath and Camelot. Games and Culture 1, 163–186.
Chuah, S.H., Hoffmann, R., Jones, M.K., Williams, G.A., 2006. An economic anatomy of culture: attitudes and behaviour in inter- and intra-national ultimatum game experiments, Occasional Papers 13, Industrial Economics Division, Nottingham University Business School.
Chuah, S.H., Hoffmann, R., Jones, M.K., Williams, G.A., 2007. Do cultures clash? Evidence from cross-national ultimatum game experiments. Journal of Economic Behavior and Organization 64, 35–48.
Crawford, V.P., 1995. Adaptive dynamics in coordination games. Econometrica 63 (1), 103–143.
Devetag, G., 2005. Precedent transfer in coordination games: an experiment. Economics Letters 89, 227–232.
Duffy, J., Nagel, R., 1997. On the robustness of behaviour in experimental 'beauty contest' games. Economic Journal 107, 1684–1700.
Erev, I., Roth, A.E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. American Economic Review 88 (4), 848–881.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. American Economic Review 90 (4), 980–994.
Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M., 1994. Fairness in simple bargaining experiments. Games and Economic Behavior 6, 347–369.
Guest, T., 2007. Second Lives. Random House, London, UK.
Harbaugh, W.T., Krause, K., Liday, S.G., 2003. Bargaining by children, working paper.
Harrison, G.W., List, J.A., 2004. Field experiments. Journal of Economic Literature 42 (4), 1009–1055.
Healy, P.J., 2006. Learning dynamics for mechanism design: an experimental comparison of public goods mechanisms. Journal of Economic Theory 129 (1), 114–149.
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. (Eds.), 2004. Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies. Oxford University Press, Oxford, UK.
Ho, T., Camerer, C., Weigelt, K., 1998. Iterated dominance and iterated best response in experimental 'p-beauty contests'. American Economic Review 88 (4), 947–969.
Hoffman, E., McCabe, K., Shachat, K., Smith, V.L., 1994. Preferences, property rights and anonymity in bargaining games. Games and Economic Behavior 7, 346–380.
Hoffman, E., McCabe, K., Smith, V.L., 1996. Social distance and other-regarding behavior in dictator games. American Economic Review 86 (3), 653–660.
Knez, M., Camerer, C., 1994. Creating expectational assets in the laboratory: coordination in 'weakest-link' games. Strategic Management Journal 15, 101–119.
Kovalchik, S., Camerer, C.F., Grether, D.M., Plott, C.R., Allman, J.M., 2005. Age and decision making: a comparison between neurologically healthy elderly and young individuals. Journal of Economic Behavior and Organization 58, 79–94.
Kurz, M., 2008. Beauty contests under private information and diverse beliefs: how different? Journal of Mathematical Economics 44 (7–8), 762–784.
Linden-Labs, 2008. Second life website. URL http://secondlife.com/.
Nagel, R., 1995. Unraveling in guessing games: an experimental study. American Economic Review 85 (5), 1313–1326.

Noveck, B.S., 2004. Introduction: the state of play. New York Law School Law Review 49 (1), 1–18.
Oosterbeek, H., Sloof, R., van de Kuilen, G., 2004. Cultural differences in ultimatum game experiments: evidence from a meta-analysis. Experimental Economics 7, 171–188.
Rokeach, M., 1973. The Nature of Human Values. Free Press, New York, NY.
Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: an experimental study. American Economic Review 81 (5), 1068–1095.
Schotter, A., Weiss, A., Zapater, I., 1996. Fairness and survival in ultimatum and dictatorship games. Journal of Economic Behavior and Organization 31, 37–56.
Schwartz, S.H., 2001. Value hierarchies across cultures: taking a similarities perspective. Journal of Cross-Cultural Psychology 32 (3), 268–290.
Schwartz, S.H., 2002. A proposal for measuring value orientations across nations. In: European Social Survey Questionnaire Development Report. European Social Survey, London, UK.
Van Huyck, J., Battalio, R., Beil, R., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. American Economic Review 80, 234–248.
Yee, N., 2006. The demographics, motivations and derived experiences of users of massively-multiuser online graphical environments. Presence: Teleoperators and Virtual Environments 15, 309–329.