Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

An experiment on experimental instructions

Maria Bigoni
Davide Dragone

Quaderni - Working Paper DSE N° 794

An experiment on experimental instructions.*

M. Bigoni[†], D. Dragone[‡]

October 26, 2011

Abstract

In this paper we treat instructions as an experimental variable. Using a public good game, we study how the instructions' format affects the participants' understanding of the experiment, their speed of play and their experimental behavior. We show that longer instructions do not significantly improve the subjects' understanding of the experiment; on-screen instructions shorten average decision times with respect to on-paper instructions, and requiring forced inputs reduces waiting times, in particular for the slowest subjects. Consistent with cognitive load theory, we find that short, on-screen instructions which require forced inputs improve on subjects' comprehension and familiarity with the experimental task, and they contribute to reduce both decision and waiting times without affecting the overall pattern of contributions.

Keywords: Cognitive load theory, Comprehension, Distraction, Experimental instructions.
Jel code: C72, C90, H41.

As we need to use clean test tubes in chemistry experiments,
so we need to get the laboratory conditions right
when testing economic theory. (Binmore, 1999)

Experimental instructions are like test tubes in hard science experiments: they are necessary to carry out an experiment, and they must be as neutral and "sterile" as possible. In this study we focus precisely on this issue and run an experiment where the treatment variable is the format used to present informationally-equivalent instructions. This allows improving control over the variables that may affect the outcome of an economic experiment and obtaining useful insights on the way people process the information they receive.

*We gratefully acknowledge the financial support from the Italian Ministry of Education, FIRB – Futuro in Ricerca – grant no. RBFR084L83, and from the Max Planck Institute for Economics of Jena (Strategic Interaction Group). We thank Andreas Matkze and Claudia Zellmann for invaluable research assistance and the participants at the 2011 Workshop on Behavioral and Experimental Economics in Florence, at the Max Plank Institute for Economics of Jena, and at the University of Bologna for their comments. The usual disclaimer applies.

[†]Dipartimento di Scienze Economiche, Università di Bologna, Piazza Scaravilli 2, 40126, Bologna, Italia; Phone: +39-051-209-8122; Fax: +39-051-0544522; e-mail: maria.bigoni@unibo.it.

[‡]Dipartimento di Scienze Economiche, Università di Bologna, Piazza Scaravilli 1, 40126, Bologna, Italia; and Strategic Interaction Group, Max Plank Institute of Economics, Jena, Germany; Phone: +39-051-209-8880; Fax: +39-051-209-8143; e-mail: davide.dragone@unibo.it.

Information not only yields benefits, but also entails some costs: "What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the over-abundance of information sources that might consume it" (Simon, 1971). In light of these considerations, we conjectured that not only the content of the instructions, but also their format may affect (i) the subjects' understanding of the experimental set up, (ii) their choices during the experiment, and (iii) their decision and waiting times. To our knowledge, this is the first economic experiment that explores this perspective.

The experimental treatments differ along three dimensions. The first one is the length of the instructions, which is manipulated by using redundant explanations. The use of repeated explanations, which is rooted in the latin motto *repetita iuvant* (repetitions are useful), is a common technique among experimenters. It is based on the assumption that the costs associated to repeating an explanation are negligible with respect to the potential benefits associated to a better understanding. Others experimentalists, however, prefer using concise instructions. This reduces the time spent in reading the instructions, but it may fail to ensure that subjects get a full understanding of the experimental scenario, specially if they get distracted during the instruction reading. Second, we study whether instructions written on paper and instructions presented on the screen produce significant differences. The latter case is relevant both for on-line experiments, where instructions are necessarily on-screen, and for experiments run through computerized interfaces (e.g. z-Tree), where adopting on-screen instructions may help the subjects to become familiar with the experimental environment. Finally, we consider the role of examples, comparing the case where the examples are simply illustrated to the subject, with the case where the examples require active participation (a forced input) by the participant.

The experiment is based on a standard public good game and is run with inexperienced subjects. Consistent with the rich existing literature, we find that contributions are positive and declining over time in all treatments, with a significant restart effect. Within this general pattern, interesting differences emerge among treatments, because as subjects gain experience contributions are lower when instructions are presented on screen rather than on paper. Using repetitions and redundancies does not significantly improve the subjects' understanding of the experiment: short instructions perform equally well when compared to long instructions, except for the case where they are presented on screen and require no active participation from the subjects. Concerning individual characteristics, we find that impulsive subjects have a lower understanding of the instructions, specially when instructions are short and on paper, or are on screen and require no active participation; in all treatments subjects with better logical abilities understand the instructions better than the others. Finally, subjects are the fastest in making choices during the experiment when instructions are presented on screen and require forced inputs. We conclude that short, on-screen instructions which require forced inputs improve subjects' comprehension and familiarity with the experimental task, and help having shorter experimental sessions.

The paper is structured as follows. In next Section we discuss the related literature and in Section 2 we present the experimental design. Results are presented in Section 3 and discussed in Section 4. Section 5 concludes.

1 Related Literature

As a recent science, experimental economics has relied on the rich literature in social sciences and in statistics, and on the corresponding guidelines aimed at writing good instructions, i.e. instructions which conduce to reliable data. When writing and designing surveys and questionnaires, the typical recommendations require making questions easily understandable, using words and questions that have the same meanings for all participants and avoiding ambiguities. Parsimony and clarity are often suggested because long questions and long surveys might confuse the subjects, increase the time to collect the answers and, ultimately, increase research costs. Sentences should be constructed in the simplest possible way, and avoid using passive form, double or triple negations; wording should be as consistent as possible with the cultural level of the participants (see Fowler, 1995, among others). Demand effects should be carefully avoided and, consistent with Tversky and Kahneman (1981)'s results on framing, special care is recommended when handling suggestive or emotionally loaded words and sentences.

Within the above mentioned guidelines, experimental instructions are often written according to the experimenter's sensitivity and experience. Scarce attention is paid to the results of a branch of experimental psychology called instructional design theory, whose goal is to study how instructions should be designed to help people learning, understanding and applying a predetermined set of principles, concepts and procedures. Within this research field, a major role is played by the cognitive load theory and by a number of empirical studies which test instructions involving different working memory loads (see van Merriënboer and Sweller, 2005, for a review). Among other results, the available evidence suggests that subjects' understanding of the instructions is reduced when information is split among multiple sources, or when seemingly useful but non-essential explanatory material (e.g. a commentary on a diagram) is provided (Chandler and Sweller, 1991). On the contrary, understanding is improved when people can both read and listen to the instructions, possibly because working memory has partially independent pathways for processing visual and auditory stimuli (Mousavi et al., 1995). Inexperienced subjects seem to benefit more from studying instructions made of worked examples than from instructions which require solving problems, but for experienced subjects worked examples are redundant and problem solving yields better performance (Kalyuga et al., 2001). The bottom line of this literature is that the effectiveness of instructions increases if they succeed in focusing the subjects' attention on the relevant information, if the cognitive load required to understand the instructions is low, and if prior skills or knowledge, or familiarity with the task, are exploited (van Merriënboer and Sweller, 2010).

The role of attention, cognitive demand and individual behavior has been investigated in some recent economic experiments. Considering a large sample of individuals, Burks et al. (2009), Oechssler et al. (2009), and Dohmen et al. (2010) find that cognitive skills are correlated with measured risk and time preferences. Dave et al. (2010) compare two alternative procedures for eliciting risk references, with different degrees of complexity. They find that the correlations between numeracy proficiency and preferences elicitation may be obscured by the difficulty of the elicitation procedure. In particular, the two procedures yield similar results for subjects with high mathematical skills, while for subjects with low mathematical ability the simple procedure generates less noise and has a similar predictive accuracy. Interestingly, when the task interface includes visual displays of the gambles and images of money to be more user-friendly, no relationship between math skills and risk attitudes is found (Eckel et al., 2007).

The above experiments investigate the relations between individual cognitive skills and individual behavior. By contrast, our experiment studies the relation between individual cognitive skills and the format used to present informationally-equivalent experimental instructions. It does not focus on framing, nor on the effect of experimental tasks with different degree of complexity, but it investigates the consequences of different instruction formats *given* the same framing and the same experimental task to be performed.

2 Experimental design

The experiment is based on a standard public good game. Subjects play in fixed groups of four for three rounds of 12 periods each. Before periods 13 and 25 they are randomly re-matched, so that two subjects cannot belong to the same group for more than one round (absolute stranger matching). At the beginning of every period, each subject receives 20 points which she can contribute to a common project, in part o totally. Total contributions to the common project are doubled by the experimenter and equally distributed to the four group members. All points not contributed to the common project remain in the subject's private account. After 36 periods, subjects fill in a demographic questionnaire and a set of questions aimed at measuring cognitive and linguistic skills. Cognitive skills are assessed using two IQ-type questions and the Cognitive Reflection test proposed by Frederick (2005).[1] Linguistic skills are assessed via a German test routinely used at the Max Planck Institute of Economics of Jena to check whether subjects have sufficient knowledge of the German language.[2] After the linguistic test, subjects are paid in private and leave the lab.

The experiment includes 4 treatments, named *Baseline*, *Short-on-paper*, *Short-on-screen*, and *Short-active*. The sole difference across treatments lies in the instructions format (see Table 1), while the experimental setup and the actions available to the subjects are identical across treatments

The *Baseline* treatment instructions are based on Fehr and Gächter (2000). They are printed on paper and given to each participant, and consist of five parts. In the first part, subjects are welcomed. In the second part, the public good game is described. This part includes information on the endowment of 20 points each subject receives, how the 20 points can be allocated and how payoffs are obtained, depending on the choices of the group members. In the third part, subjects are explained step-by-step the two main stages they would go through during the experiment, which correspond to the two screens they will see, named "Your choice" and

[1] A translation of these questions is reported in Appendix A

[2] Subjects must fill in a (German) text choosing between a list of words. Overall, they have to answer 12 questions.

Treatment	Words (no.)	Medium	Example
Baseline	1091	on paper	read-only
Short-on-paper	621	on paper	read-only
Short-on-screen	621	on screen	read-only
Short-active	633	on screen	forced input

Notes: No. of words refers to the German version of the instructions.

Table 1: Summary of the treatments.

"Results", respectively. To provide visual support, screenshots are printed on the instructions. In the fourth part ("Additional information") subjects are given additional information concerning the number of rounds, the matching protocol and the conversion rate of the points earned during the game. In the fifth section, two illustrative examples are provided. After correctly answering the control questions, subjects start playing.

In the three remaining treatments (*Short-on-paper, Short-on-screen*, and *Short-active*) instructions are substantially shorter than in the *Baseline* treatment (1091 vs. 621 or 633 words, depending on the treatment). This is obtained by using the instructions of the *Baseline* treatment as a starting point, and then cutting redundancies, repetitions and minor details. The short version of the instructions is therefore an extract of the *Baseline* instructions, but it does not differ from the long version in the wording, nor in the information or in the explanations provided to the subjects.[3] The differences between the three treatments with short instructions consist in the instructions format: in the *Short-on-paper* treatment instructions are presented on paper; in the *Short-on-screen* treatment they are presented on screen; in the *Short-active* treatment they are presented on screen and the two examples are replaced by forced input choices.

It is important to observe that repetitions and redundancies are cut from the *Baseline* instructions in such a way that the short instructions meet the following criteria. First, the subjects are instructed on what they have to do (i.e. that they have to contribute between 0 and 20 points); then they are explained the consequences of their choices (i.e. how the payoff corresponding to theirs and their opponents' choices is determined). Additional information which is not strictly required to understand what a subject has to do and what are the consequences of her choices (e.g., where the number of period is shown on the screen, the conversion rate of the experimental currency, etc.) is provided only afterwards. Interestingly, providing information following the above steps is particularly natural in the treatments where instructions are presented on screen (*Short-on-screen* and *Short-active* treatments) because in each period a subject goes through two screens. In the first one she is required to make her choice, in the second one she receives information on the others' choices and on her payoffs. Hence, what the subject has to do is the relevant information contained the first screen, while information on the consequences of her choices appears in the second screen. Consistently, in the *Short-on-screen* and *Short-active* treatments instructions are presented according to the effective sequence of

[3]An English translation of the instructions is reported in Appendix B.

screens the subjects will see.[4]

To enhance control over possible confounding factors, instructions were recorded in advance; the audio file was then played aloud with speakers during the instruction phase in all treatments. Hence, during this phase, the participants could both listen and read the instructions (on paper or on screen, depending on the treatment). After the instruction phase, subjects had to answer some control questions to verify their comprehension of the experimental set up. The experiment did not start until all subjects had answered the control questions correctly. The software recorded the number of mistakes made by each subject, and the total time each subject needed to answer all questions correctly, which are the measures we use to evaluate the subject's comprehension of the instructions depending on the different formats.

The experiment was run at the Laboratory for Experimental Economics of the Max Plank Institute in Jena, in November and December 2010. The software was developed using z-Tree (Fischbacher, 2007). We ran 2 sessions per treatment, with 16 subjects per session.[5] Subjects were recruited through the ORSEE platform (Greiner, 2004), and were randomly assigned to the treatments. None of them had previously participated in other experiments.[6] Upon arrival, subjects were welcomed and randomly seated at visually separated computer terminals. After the instruction phase, and for the whole duration of the experiment, a printed copy of the instructions was given to all participants. Sessions lasted on average 55 minutes, including instructions and payment. Subjects earned on average 13 Euros, with a show-up fee of 2.5 Euros.

3 Results

In this section we first describe the pattern of contributions across treatments. Then we discuss how the different formats affect the subjects' understanding of the game. Finally, we analyze how the subjects' decision and waiting time vary across treatments.

3.1 Contributions

Figure 1 presents the trend of average contributions across treatments, and across rounds. The qualitative pattern is substantially in line with the previous experiments (see, for example, Andreoni, 1988; Croson, 1996).

[4] In addition, the on-screen instructions are such that the screen is divided in two parts: in the top part there is the text of the instructions, while in the remaining part the subjects can see the corresponding screenshot. When the instructions describe some relevant object in the screenshot (e.g., 'after you have made your choice, press the "ok" button'), arrows appear on the screen to help locating the object (e.g., the arrow points to the "ok" button, see Figure B.3). This technique is commonly used in software tutorials to teach a beginner where buttons and commands are located on the screen.

[5] In fact, we ran 9 sessions, but we had to drop data from the first one due to technical problems with the software.

[6] To control for possible asymmetries across treatments, we collected data about subjects' individual characteristics by means of a questionnaire which was administered at the end of each session. Aggregate results from the questionnaire are presented in Appendix A.

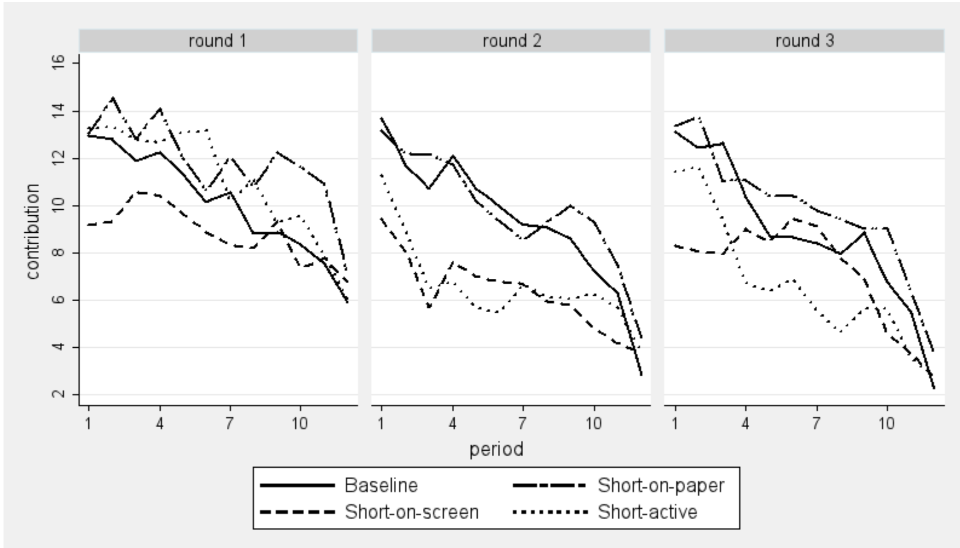


Figure 1: Contribution

Result 1 *In all treatments: the initial average contribution is between a half and two thirds of the endowment; average contributions decrease within each round; a significant restart effect is observed at the beginning of round 2 and round 3.*

Result 1 shows that the four formats used in the experiment do not produce "anomalous" responses in the participants' experimental choices, as the overall trend of contributions is in line with the existing literature on public good games for all formats of instructions. Yet, there are some differences across treatments both when considering the initial contribution (i.e. the average contribution in period 1 of round 1) and the average contribution for each round (Table 2).

treatment	period 1	round 1	round 2	round 3
Baseline	13.0	10.1	9.3	8.8
Short-on-paper	13.1	11.8	9.8	9.8
Short-on-screen	9.2	8.8	6.3	7.2
Short-active	13.3	11.0	6.6	6.7

Table 2: Average contribution by round.

The initial contribution is significantly lower in the *Short-on-screen* treatment than in other treatments (Wilcoxon, $p < 0.05$ for all three comparisons), but no significant differences emerge among the *Baseline*, the *Short-on-paper* and the *Short-active* treatments. To study whether and how contributions differ across treatments as subjects gain experience, we run three linear regressions, one for each round. Among the independent variables we include dummies for the *Short-on-paper*, *Short-on-screen*, and *Short-active* treatments; the variable *Period* introduces a linear time trend and helps understanding the role of experience. In addition, we control for the

	Round 1		Round 2		Round 3	
	coeff.	(s.e.)	coeff.	(s.e.)	coeff.	(s.e.)
Short-on-paper	1.884	(1.478)	0.214	(1.421)	0.606	(1.381)
Short-on-screen	-1.252	(1.393)	-2.870*	(1.399)	-1.325	(1.249)
Short-active	1.234	(1.127)	-2.302***	(0.463)	-1.688*	(0.740)
Period	-0.475***	(0.100)	-0.537***	(0.065)	-0.677***	(0.065)
Constant	18.918***	(4.965)	10.653**	(4.378)	15.572***	(3.392)
Indiv. characteristics	Yes		Yes		Yes	
N.obs.	1536		1536		1536	
R-squared	0.133		0.180		0.196	

Notes: standard errors robust for clustering at the session level (in parentheses). In this and in the following regressions, the symbols *, **, and *** indicate that a coefficient is significant at the 10%, 5%, and 1% level, respectively.

Table 3: Linear regression on contributions.

set of individual characteristics listed in Appendix A.[7]

Regressions' results are reported in Table 3. Consistent with Result 1 and Figure 1, the estimated coefficient of the variable Period is negative and significant. In addition, we find that, in round 1, contributions are not significantly different across treatments.[8]

In the next rounds, however, average contributions are lower in the treatments where instructions are presented on screen, rather than on paper. More precisely, we find that the average contributions in the *Short-active* treatment are between 20% and 30% lower than those in the *Baseline* treatment and the regression results show that the difference is statistically significant ($p < 0.01$ in round 2 and $p < 0.10$ in round 3, see Table 3). A difference also emerges when comparing the average contributions in the *Short-on-screen* treatment with those in the *Baseline* treatment, but it is weaker ($p < 0.10$ in round 2 and $p > 0.10$ in round 3).

Result 2 *In rounds 2 and 3 average contributions are lower when instructions are presented on screen rather than on paper.*

3.2 Comprehension of the instructions

Result 2 suggests that average contributions are lower in the treatments in which instructions are presented on screen than in the treatments where instructions are presented on paper. A possible explanation for this finding is that on screen instructions help subjects to better understand the strategic environment of the game, which would justify the faster convergence to the Nash equilibrium of the static game. Support for this explanation requires the participants in the

[7]Of these characteristics, only risk aversion has a significant effect in all three rounds: subjects reporting a higher degree of risk aversion on average tend to contribute less.

[8]According to a Wald test on coefficients, there is a difference only between the *Short-on-screen* and the *Short-active* treatments, although it is weakly significant ($p < 0.10$).

on-screen treatments to perform well in the control questions. To test whether this is the case, we consider (i) the number of wrong answers given by each subject in the control questions, and (ii) the amount of time spent to complete the whole test correctly. Both measures provide valuable information about the subjects' level of comprehension of the instructions. We assume that a participant who has had problems in understanding the experimental set up may either answer quickly but make many mistakes, or answer correctly but be very slow (because she needs time to think and possibly to re-read the instructions). Our measure of comprehension of the instructions is built in order to take into account both possibilities. First, we make two rankings. The first one is based on the total number of wrong answers given by each participant in the control questions: the higher the rank attributed to a participant, the lower the number of mistakes she made. The second ranking is based on the total time spent to answer all the questions: the higher the rank, the fastest the participant in correctly answering all control questions. For every subject, we then take the minimum between these two ranks and we use it as a measure of her comprehension of the instructions. Hence, our measure is high for subjects that are both fast and correct in answering (which is consistent with a good understanding of the instructions), and it is low for subjects that were either very slow or that made too many mistakes, or both. Figure 2 reports the distribution of this measure of comprehension across treatments.[9]

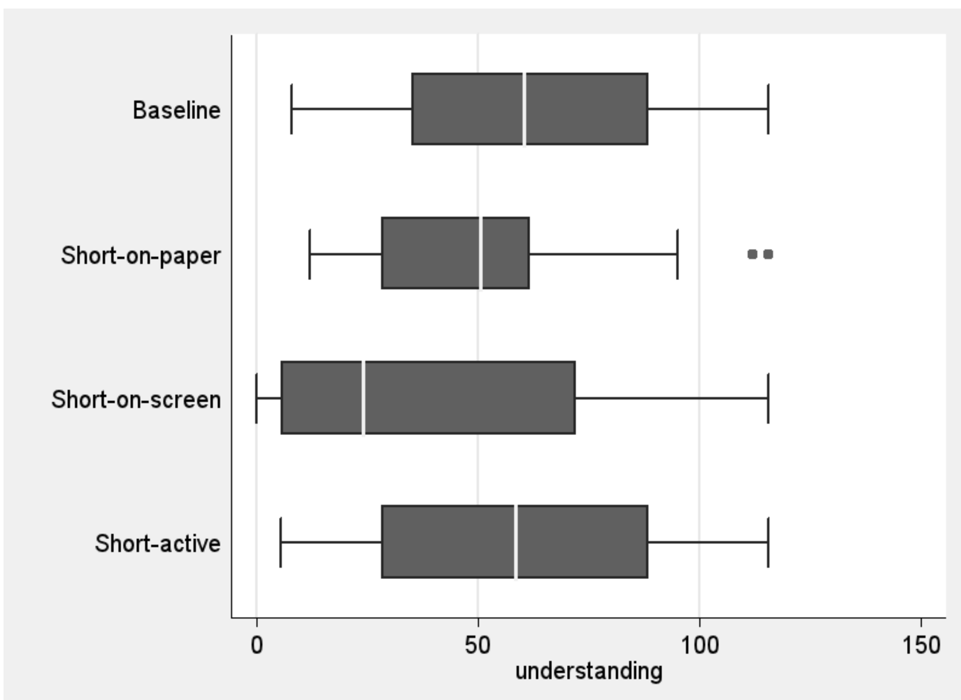


Figure 2: Understanding of instructions across treatments.

[9]The white vertical line amid the boxes indicates the median, the left and right sides of the boxes indicate the 25th and the 75th percentiles, respectively, and the extremes of the whiskers represent the upper and lower adjacent lines.

The median level of comprehension is much lower in the *Short-on-screen* treatment than the other treatments, and the difference is statistically significant according to Wilcoxon tests.[10] No significant differences emerge among the other three treatments.

Result 3 *Comprehension of the instructions is significantly lower in the Short-on-screen treatment, and similar in the Baseline, Short-on-paper and Short-active treatments.*

To dig deeper into this result, we run an additional regression to study whether our measure of comprehension is correlated with the subjects' logical abilities, their impulsiveness, their linguistic skills and their education. Neglecting for a moment the possible treatment effects, we pool together the data from all treatments. The results are reported in Table 4.

	coeff.	(s.e.)
graduated	-9.338	(5.934)
impulsiveness	-11.895***	(2.377)
IQ-classification	8.525	(8.303)
IQ-logic	12.974**	(5.121)
German test	-2.399	(3.078)
Constant	62.049***	(10.156)
N.obs.	128	
R-squared	0.266	

Notes: standard errors robust for clustering at the session level (in parentheses).

Table 4: Linear regression on understanding.

Interestingly, the subjects who gave the correct answer to the IQ-logic question tend to perform significantly better in the control questions ($p < 0.05$), while impulsiveness is negatively correlated with the level of comprehension of the instructions ($p < 0.01$). To check whether the correlation between these individual characteristics and understanding is constant across treatments, we run two additional regressions in which we separately include *impulsiveness* and *IQ-logic* as regressors, together with their interactions with the dummies *Short-on-paper*, *Short-on-screen*, and *Short-active*. The corresponding results are presented in Tables 5 and 6.

Table 5 shows that impulsiveness is correlated with low comprehension in all treatments. The negative effect of impulsiveness is exacerbated in the *Short-on-paper* and *Short-on-screen* treatments ($p < 0.001$ for both treatments), while in the *Short-active* treatment is is not significantly different from the *Baseline* ($p > 0.10$).

Table 6 instead shows that the subjects who answered correctly to the IQ-logic question understand better the instructions in all treatments, with no significant difference among treatments ($p > 0.10$ for all three treatment dummies).

[10] $p < 0.05$ for *Short-on-screen* vs. *Baseline*, and for *Short-on-screen* vs. *Short-active*; $p < 0.10$ for *Short-on-screen* vs. *Short-on-paper*

	coeff.	(s.e.)
impulsiveness	-11.802***	(2.559)
Short-on-paper x impulsiveness	-5.347***	(0.559)
Short-on-screen x impulsiveness	-9.014***	(0.482)
Short-active x impulsiveness	0.511	(1.311)
Constant	78.728***	(6.179)
N.obs.	128	
R-squared	0.273	

Notes: standard errors robust for clustering at the session level (in parentheses).

Table 5: Linear regression: understanding and impulsiveness.

	coeff.	(s.e.)
IQ-logic	29.469***	(6.308)
Short-on-paper x IQ-logic	-6.275	(9.444)
Short-on-screen x IQ-logic	-21.690	(15.883)
Short-active x IQ-logic	-3.895	(11.106)
Constant	39.531***	(3.299)
N.obs.	128	
R-squared	0.126	

Notes: standard errors robust for clustering at the session level (in parentheses).

Table 6: Linear regression: understanding and logical abilities.

Result 4 *In all treatments, impulsiveness and comprehension are negatively correlated, while logical abilities and comprehension are positively correlated. In the Short-active treatment impulsive subjects' comprehension is as good as in the Baseline. In the Short-on-paper and in the Short-on-screen treatments the negative impact of impulsiveness on understanding is stronger.*

3.3 Decision and waiting time

The last part of our analysis focuses on the effects that the different formats of instructions have on the subjects' decision time and, consequently, on the time subjects spend just waiting for the other participants to make their choices.

Figure 3 presents box-plots summarizing the distribution of subjects' average decision time in period 1 of round 1, and across all periods of the other three rounds. The mean decision time in the first period is between 25% and 30% shorter in the *Short-active* treatment (11.2") than in the *Baseline* (15.8"), *Short-on-paper* (16.8"), and *Short-on-screen* (14.8") treatments. Results from Wilcoxon tests indicate that the difference is highly significant.[11]

To test whether the difference in decision time between the *Short-active* and the *Baseline* treatment remains significant as subjects gain experience, we run three additional regressions – one for each round – in which the dependent variable is the individual decision time in a period. Among the dependent variables we include dummies for the *Short-on-paper*, *Short-on-screen*, and *Short-active* treatments. In addition, we include variable *Period* to check for the presence of a linear time trend, and we control for the set of individual characteristics listed in Appendix A.[12] The regressions' results show that the difference persists until the second round, while in the third round the difference is no more significant.

A second finding that emerges from Figure 3 is that, in all three rounds, the distribution of decision times in the *Short-on-paper* and *Short-on-screen* treatments is more disperse that in the *Baseline* and in the *Short-active* treatments. Levene's tests on equality of variances confirm that in treatments *Short-on-paper* and *Short-on-screen* the variance in decision time in the first period is significantly higher than in the *Baseline* treatment ($p < 0.01$ for both treatments), while no significant difference emerges between the *Baseline* and the *Short-active* treatment. Levene's tests also allow to conclude that, in all three rounds, the variance in the subjects' average decision time is significantly smaller in the *Short-active* treatment than in the *Short-on-paper* and in the *Short-on-screen* treatments (the difference is significant at the 5% level for all rounds and for all bilateral comparisons). This information is relevant because in our experiment (and in many other economic experiments) a new period begins only when all participants have made their choices. Hence, if a participant is fast in making her choice, she must wait until the slowest participant has taken her choice. As a consequence, a large variance in decision times has a direct impact on the waiting times experienced by fast subjects, i.e. on the amount of idle time they have to wait before the other participants to the experiment make their choice. For each treatment, Table 7 presents the average waiting and decision time across all periods of all

[11] $p < 0.01$ for the comparisons with the *Baseline* and *Short-on-paper* treatments, and $p < 0.05$ for the comparison with the *Short-on-screen* treatment.

[12] The regressions results are reported in Table C.1 in Appendix C.

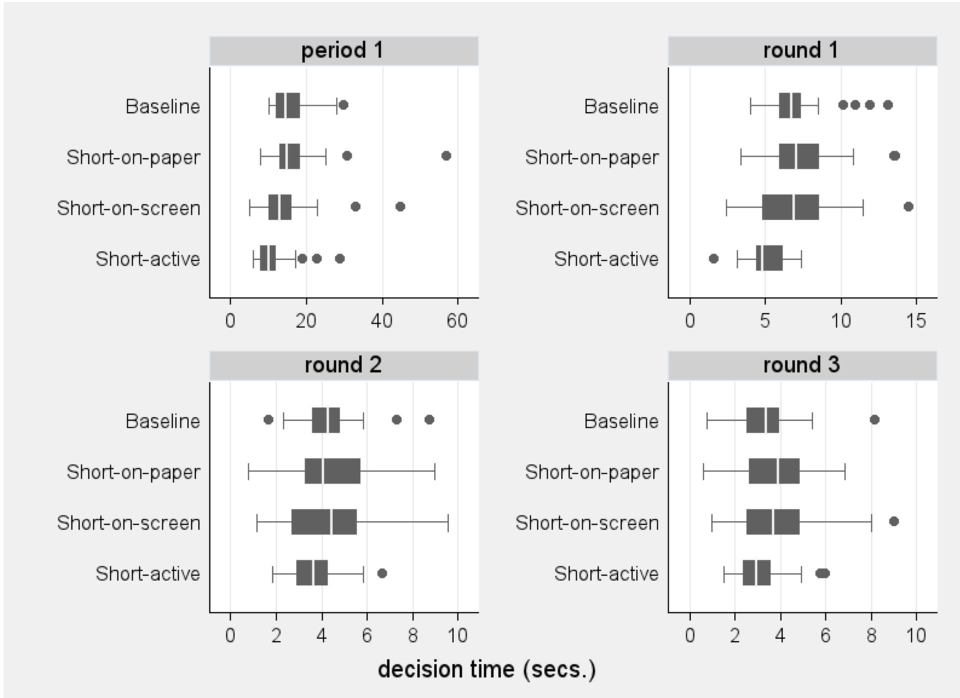


Figure 3: Decision time across rounds and treatments.

rounds. Notably, in the *Short-active* treatment the average waiting time is substantially shorter

treatment	decision time	waiting time
Baseline	10.4	9.5
Short-on-paper	11.0	16.7
Short-on-screen	10.0	15.8
Short-active	7.6	9.7

Table 7: Decision and waiting time across treatments.

than in the *Short-on-paper* and *Short-on-screen* treatments.

Result 5 *In the Short-active treatment, subject are the fastest in making choices, and the waiting time is as short as in the Baseline. In the other Short-on-paper and Short-on-screen treatments subjects are as slow as in the Baseline in making choices, but the waiting time is much longer.*

4 Discussion

Experimentalists often rely on the assumption that instructions are adequate, which implies that the subjects' decisions are robust with respect to instructions (Smith, 2002). In this paper we show that the format used to present instructions that are informationally equivalent may significantly affect the subjects' understanding of the experimental set up, their decision times, and their choices during the experiment.

While it is widely acknowledged that limited attention and cognitive load may significantly impact on people's understanding and behavior, very few economic experiments have focused on explicitly studying it. In our experiment we show that, for inexperienced subjects, redundancies and repetitions do not significantly improve understanding (*Baseline* vs. *Short-on-paper*). Notice that instructions in the *Baseline* treatment are very close to the ones adopted in the most recent and prominent experiments on public goods. They are the outcome of a long refinement process, are widely adopted, and generally recognized as good instructions. Yet, they contain repetitions and redundancies, and we show that they can be safely removed. With short instructions, reading time decreases by about 38%.[13] Although this effect might seem negligible, one has to bear in mind that ours is a single-task experiment, and that the required task is simple. In sessions with multiple and more complicated tasks, instructions may well absorb a consistent proportion of the total session time. In these cases, shortening instructions by 30-40% may produce relevant benefits, both for the experimenter and for the subjects. On the one hand, the experimenter saves money because one has to pay subjects according to the time spent in the lab, which implies that longer sessions are more costly. On the other hand, our own experience suggests that exposing subjects to excessively long instructions induces boredom and distraction, and possibly create undesirable noise in the data.

In tailoring the instructions for the on screen treatments, we followed the technique used to design tutorials, which are commonly used to familiarize a person with a new software. To help subjects getting to grips with the z-Tree interface, we presented the information step-by-step. As a result, subjects in the *Short-on-screen* treatment were indeed faster than in the *Short-on-paper* treatment. Yet, they did not pay enough attention to the content of the instructions and, in the control questions, they made significantly more errors than all others. This may be due to the fact that our sample of subjects is made of students, who are more used to study on paper books and notes, and that they paid more attention when reading on paper than when they were passively exposed to instructions written on the screen. We cannot conclude, however, that on screen instructions are bad for comprehension because we find a significant difference in understanding between the *Short-on-screen* and the *Short-active* treatment. In the latter case, the subjects had to exert active effort, instead of just being passive recipients of information. Interestingly, in our experiment being active simply meant being required to enter a forced input, rather than just being exposed to the corresponding worked example. Despite the fact that no particular cognitive effort is required to make this action, it seems that this minimum requirement has been useful in maintaining the subjects' focus on the instructions and in helping them to understand the experimental set up.

[13] Instruction reading time, was 8 minutes in the *Baseline* treatment, 6 minutes in the *Short-on-paper*, 6 minutes and 30 seconds in the *Short-on-screen*, and 8 minutes in the *Short-active* treatment. These figures include time to answer questions from the participants and, only in the Short-active treatment, time for entering the forced inputs.

5 Conclusion

Writing good experimental instructions is not a trivial task. The completeness and accurateness of instructions is critical to ensure that the subjects understand the experimental scenario, and that they know what they are required to do and the consequences of their actions. In most experiments particular effort is devoted to ensure that subjects understand the experiment, including aspects of the experimental design that might lead the theoretical predictions in different directions (e.g. the matching procedure and the information available to other participants). Moreover, experimenters are well aware of the fact that minor details in instructions may have large effects on the subjects' behavior, and thus try to write good instructions in order to minimize such distortions.

Interestingly, other important features of the instructions are often left to the experimenter's experience, sensitivity and personal taste. This paper is a methodological contribution to the literature on experimental economics, and it is aimed at improving control over the variables that can affect an economic experiment. We focus on three specific aspects. First, we consider the length of the instructions. Some experimenters use repetitions in their instructions, with the idea that this may improve subjects' understanding of the experiment without harm. Others prefer more concise instructions, which shortens the time spent in reading the instructions, but may fail to provide full understanding of the scenario. To understand whether repetitions and redundancies can be safely avoided, we compare treatments where instructions contain the same information, but differ in the use of repeated explanations. Second, we study whether there is a difference between instructions written on paper and instructions presented on the screen. The latter possibility is relevant because an increasing number of experiments adopts a computerized interface, and presenting instructions on the screen might help the subjects getting to grips with the computerized environment through which the experiment takes place. Finally, we compare a treatment where examples are simply illustrated to the subjects, with a treatment where the examples require active participation (a forced input).

We study the impact of these differences on inexperienced subjects playing a standard public good game, one of the most widely studied games in experimental economics. The overall results on the subjects' contributions are compatible with the existing literature, but we find significant differences in the subjects' understanding of the experimental set up, the decision times, and the contributions made during the experiment, depending on the experimental treatment. Contributions are lower when the instructions are presented on screen rather than on paper. When considering comprehension, we find no significant difference across treatments, except when instructions are presented on screen and require no active participation to the subjects. Individual differences matter: in all treatments impulsive subjects have more problems in understanding the experimental set up, while subjects with good logical skills have a better comprehension. These findings are consistent with the cognitive load theory and the idea that bounded rationality may determine a trade-off between the amount of attention devoted to the instructions and the cognitive effort required to the subjects.

Ours is the first experiment on this issue. To single out in the neatest possible way the effects of different instruction formats, we invited only inexperienced subjects, and we focused on a simple and extremely popular game. It is still an open question whether the results we get extend to different subject-pools, and different games. This is left for future research.

Appendix A Individual Characteristics

Individual characteristics are roughly homogeneous across treatments. The variable *Impulsiveness* measures the number of mistakes made in the Cognitive Reflections test. Similarly, the variable *German test* measures the number of wrong answers given in the test for linguistic skills.

characteristic	Baseline		Short-on paper		Short-on screen		Short-active
male	0.500	$\sim$	0.375	$\sim$	0.469	$>^*$	0.250
age	21.781	$\sim$	22.406	$\sim$	22.188	$>^*$	20.969
risk aversion	5.906	$\sim$	6.063	$\sim$	5.594	$\sim$	5.469
IQ-classification	0.781	$\sim$	0.906	$\sim$	0.844	$\sim$	0.781
IQ-logic	0.625	$\sim$	0.625	$\sim$	0.656	$\sim$	0.594
impulsiveness	1.688	$\sim$	1.375	$<^{**}$	1.938	$\sim$	1.750
German test	1.063	$\sim$	0.750	$\sim$	1.125	$\sim$	0.969
graduated	0.125	$<^*$	0.313	$\sim$	0.313	$>^{**}$	0.094

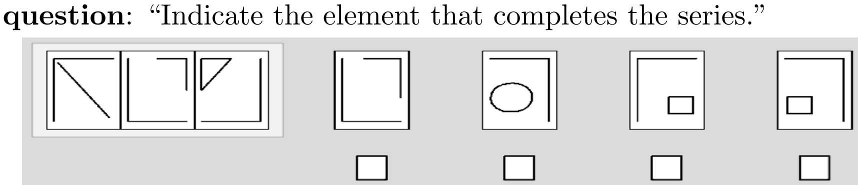
Notes: The symbol $\sim$ indicates that the difference is not significant at the 10% level according to a Wilcoxon-Mann Withney test; the symbols *, **, and *** indicate that the test returns a p-value smaller than 0.1, 0.05 and 0.01, respectively.

Table A.1: Individual characteristics: differences across treatments

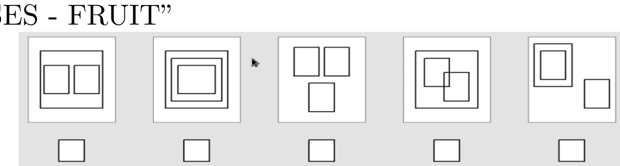
The Cognitive Reflection test (Frederick, 2005)

- A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?
- If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

"IQ-logic" question: "Indicate the element that completes the series."



"IQ-classification" question: "Which, among these diagrams, represents the relation among: ORANGES - CITRUSES - FRUIT"



Appendix B Instructions

We report the instructions for the *Baseline* treatment, translated from German. Sentences in **Bold** are as in the original version.
The sentences in italics were cut from the Baseline treatment to produce the short version of the instructions (Short-on-paper, Short-on-screen, Short-active)
The adjustments needed for the on-screen treatments (Short-on-screen, Short-active) are [within brackets].

Welcome!

You are now taking part in an economic experiment financed by the Max Plank Institute of Economics. *This institution supports and finances economic research to advance knowledge on how people make decisions.*

If you read the following instructions carefully, you can earn a considerable amount of money, *depending on your decisions. It is therefore very important that you read these instructions with care.*

The amount of money you will earn with each decision you will make will be added up and paid to you in cash at the end of the session.

Every participant receives the same information and is reading the same instructions.

INSTRUCTIONS

Before starting let us inform you that communication is not allowed. ***Please, do not communicate with the other participants during the experiment!*** If you violate this rule, we shall have to exclude you from the experiment and from all payments. *In such a case you will not receive the money you earned during the experiment, nor the fixed fee of 2.50 euros that we pay to all participants who showed up for the experiment on time.*

Should you have any questions raise your hand and we will answer in private. *Neither your question, nor the relative answer will be announced aloud to the other participants in this room.*

You will learn how the experiment will be conducted later.

We first introduce you to the basic decision situation. We will ask you to answer a few control questions at the end of these instructions, to help you to understand the decision situation.

The Decisional Situation

You will be a member of a group consisting of 4 people. Each group member has to decide on the allocation of 20 points. You can put these 20 points into your private account or you can invest them fully or partially into a project. Each point you do not invest into the project, will automatically remain in your private account.

Your income from the private account:
You will earn one point for each point you put into your private account. No one except you earns anything from your private account.

Your income from the project:
Each group member will profit equally from the amount you invest into the project. On the other hand, you will also get a payoff from the other group members' investments. The income for each group member will be determined as follows:

$$\textit{Income from the project} = \textit{sum of all contributions} \times 0.5$$

Total Income
Your total income is the sum of your income from your private account and your income from the project:

$$\begin{gathered}\textit{Total Income} = \\ \textit{Income from your private account (= 20 - contribution to the project)} + \\ \textit{Income from the project} = (0.5 \times \textit{sum of all contributions to the project})\end{gathered}$$

Now *that you know the basic decision situation* we will describe step by step the screens you will go through.

Your Choice

The experiment includes the decision situation just described to you. You will be paid at the end of the experiment based on the decisions you make in this experiment.

As you know, you will be a member of a group consisting of 4 people. Each group member receives 20 points, and *must decide how to use them.* You *and the other group members* have to decide how many points you contribute to a project. Each point you do not contribute to the project, will automatically remain in your private account (see Fig. B.1).

To indicate your contribution, use the "+" and "-" buttons. To confirm your choice, you have to click the "OK" button.

Results

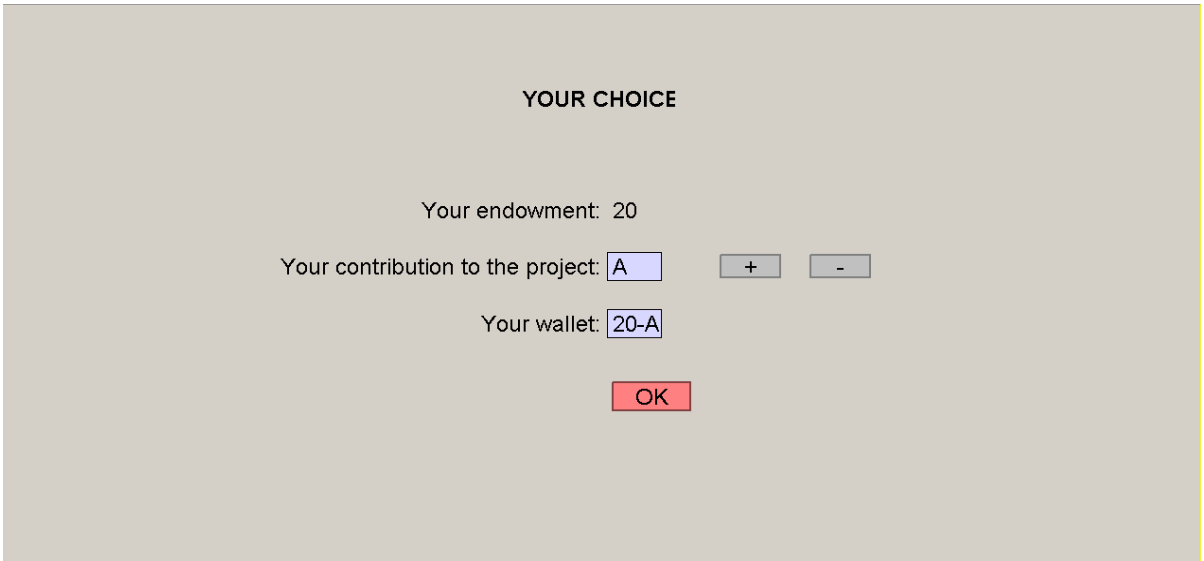


Figure B.1: Your Choice

When everybody has made her choice, you will see the RESULTS screen. The screen will show (see Fig. B.2):

- your contribution to the project
- the contribution to the project by the other 3 group members
- the total contribution of your group

Your earnings from the project is computed by the pc multiplying the total contributions by 2 and dividing equally among the 4 group members:

$$YOUR\ EARNINGS\ FROM\ THE\ PROJECT = \frac{2\times\ total\ contributions}{4}$$

This formula is equivalent to the one you have seen before, when we introduced the decision situation.

[In other words: Income from the project=sum of all contributions $\times$ 0.5]
Your earnings also depend on the points you have in your private account. Hence, for each choice you make your earning are equal to the sum of your earning from the project and the earnings in your private account.

Additional information

The experiment consists of **36 periods**. The number of the current period is always displayed in the top-left corner of the screen. **Your task is the same** in all 36 periods. *In every period, you have to decide how many of the 20 points you want to contribute to the project, and how many points you want to put into your private account.*

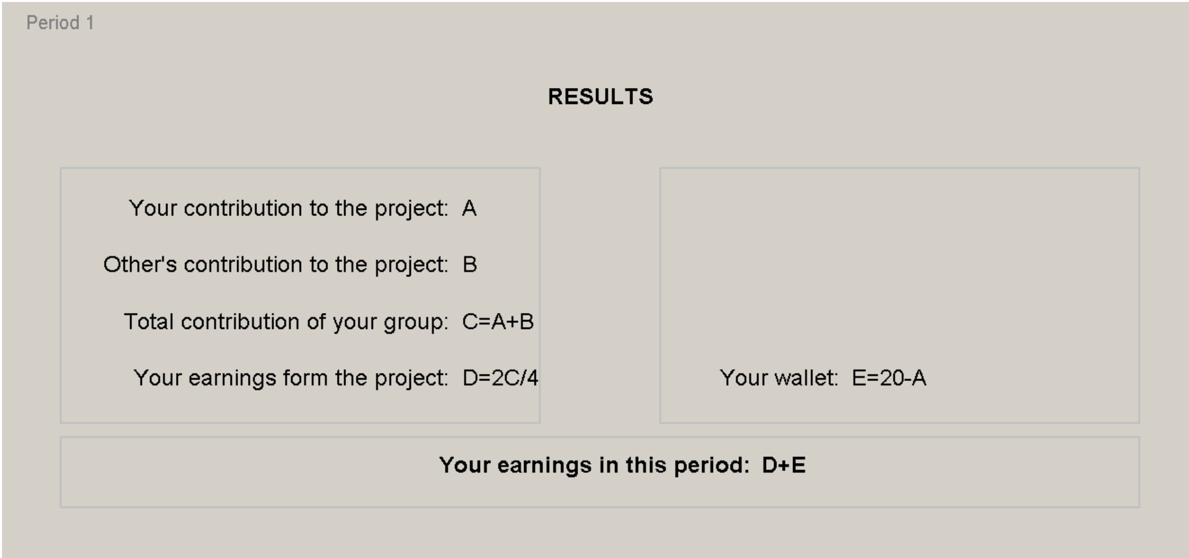


Figure B.2: Your Choice

The composition of the groups changes randomly every 12 periods. So your group consists of **the same four people** for 12 periods, after which you will be randomly grouped with other 3 people. People that are in your group once can never be re-assigned to your group in the future. *This means that you will be part of 3 groups and that you will never interact with the same participants for more than 12 periods.*

Total earnings from the experiment

After the last period you will see a screen summarizing your total earnings from the experiment. You will read the total amount of points earned during the experiment. Points will be converted into Euros at the rate:

1 point= 2.5 cent

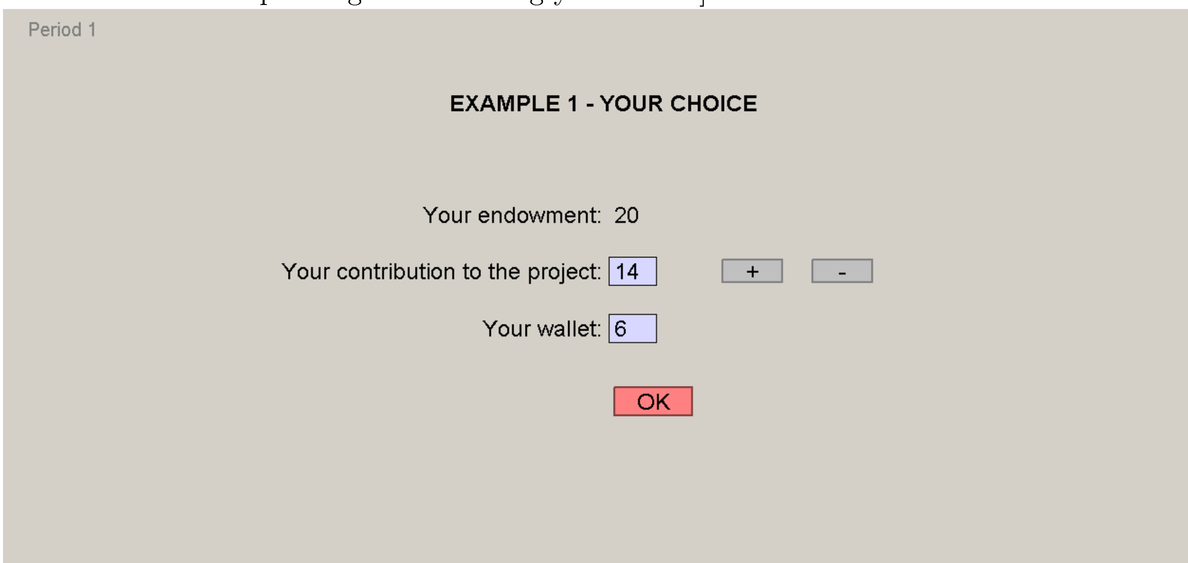
In addition, you will receive a show-up fee of 2.5 Euros for your participation.

At the end of the experiment, *please remain seated and quiet until we call you for the payment.* You will be paid in cash, and in private. No other participant will be able to know the payment you receive.

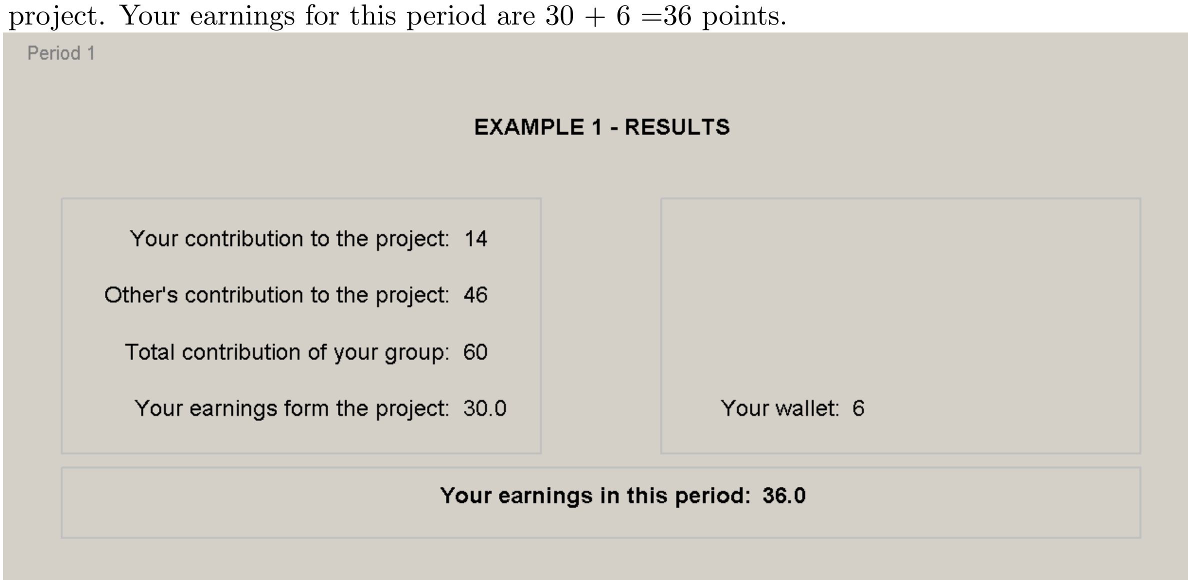
As illustrative examples, consider the following cases.

Example 1

Suppose you contribute 14 points to the project (and you keep 6 points in your private account). [Please select the corresponding amount using your mouse][14].



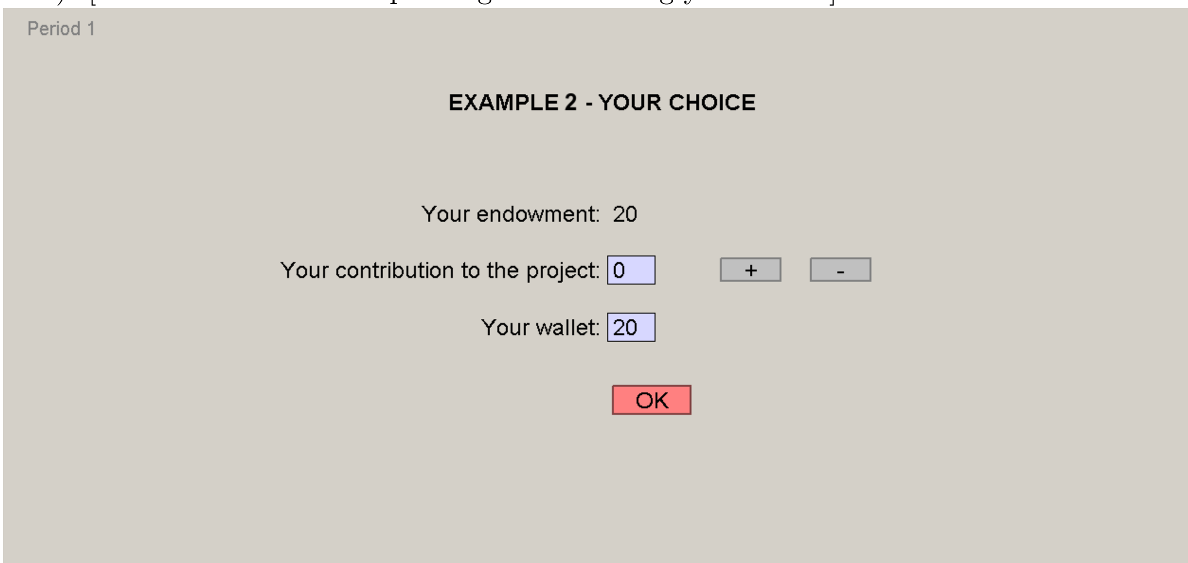
Suppose the sum of the contributions of the others is 46 points. The value of the project is (14+46)x2=120. You and the other members of the group each earn 120/4=30 points out of the project. Your earnings for this period are 30 + 6 =36 points.



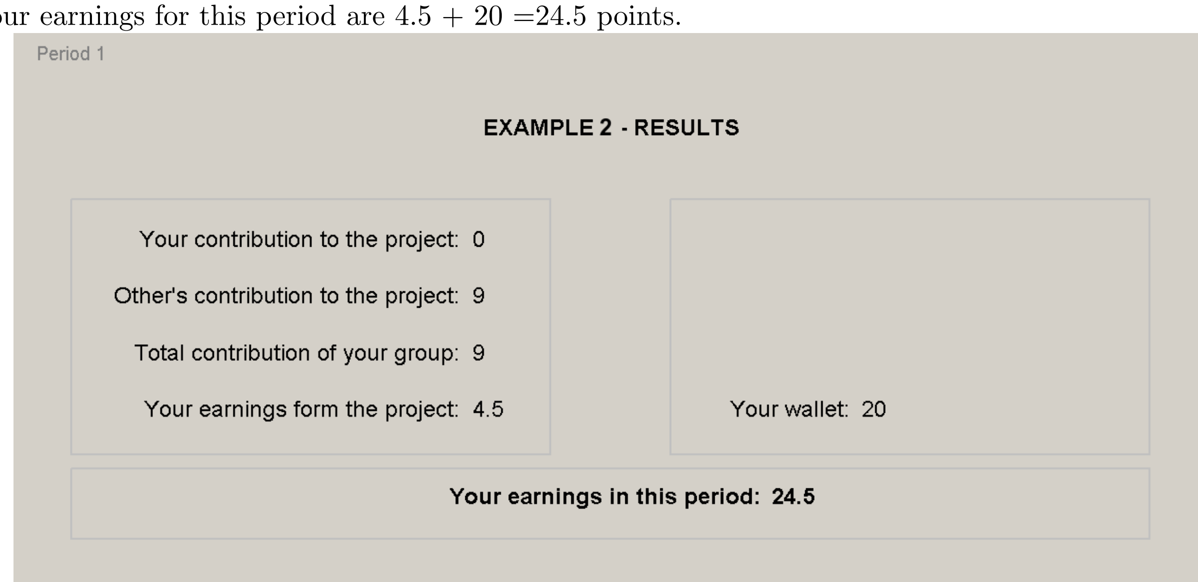
[14]This sentence was present only in the *Short-active* treatment

Example 2

Suppose now you contribute 0 points to the project (and you keep 20 points in your private account). [Please select the corresponding amount using your mouse][15]



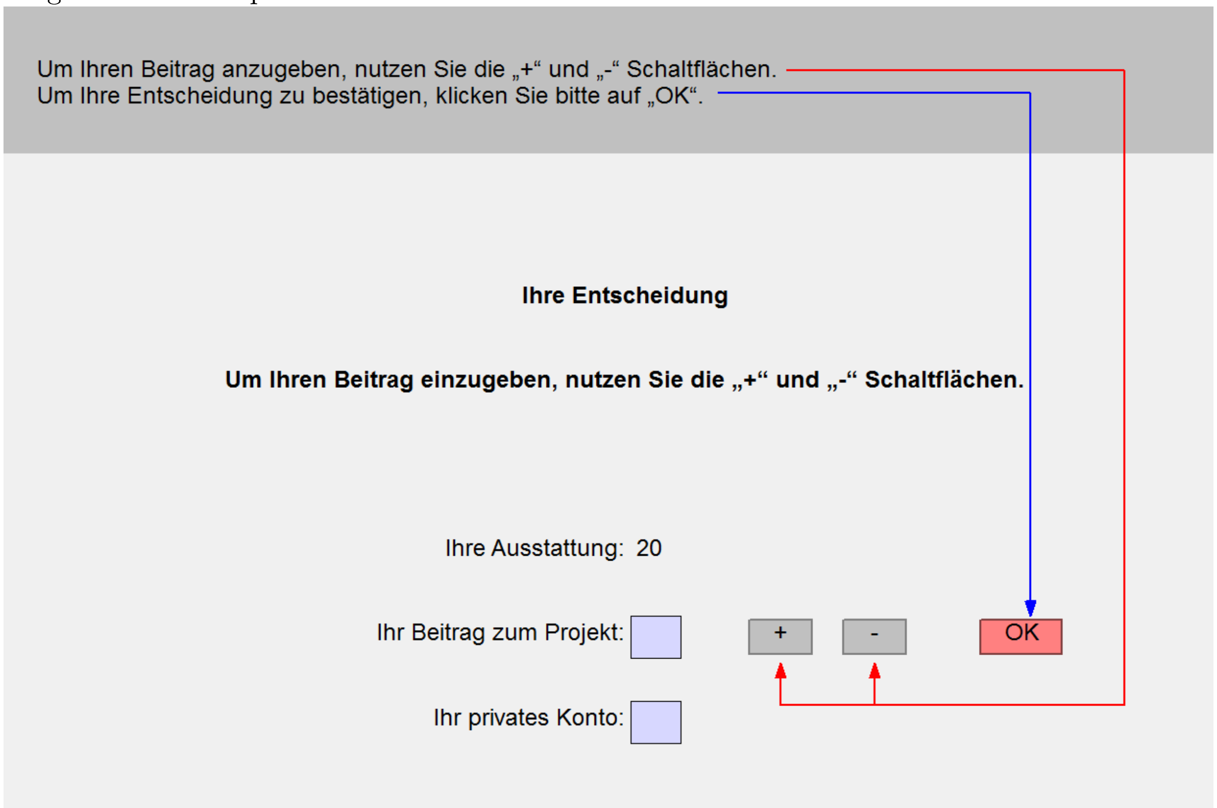
Suppose the sum of the contributions of the others is 9 points.
The value of the project is (0+9)x2=18.
You and each member of the group earn 18/4=4.5 points out of the project.
Your earnings for this period are 4.5 + 20 =24.5 points.



To test your understanding, please answer the following questions using the mouse. They have no consequences on your final earnings.

[15]This sentence was present only in the Short-active treatment.

Figure B.3: Example of instructions for the *Short-on-screen* and the *Short-active* treatment.



Notes: The text reads as "To indicate your contribution, use the '+' and '-' buttons. To confirm your choice, you have to click the 'OK' button."

Appendix C Additional regressions

	Round 1		Round 2		Round 3	
	coeff.	(s.e.)	coeff.	(s.e.)	coeff.	(s.e.)
Short-on-paper	0.440	(0.724)	0.022	(0.307)	0.633	(0.399)
Short-on-screen	-0.416	(0.523)	0.070	(0.167)	0.673**	(0.215)
Short-active	-2.094***	(0.578)	-0.735**	(0.221)	-0.440	(0.355)
Period	-0.630***	(0.040)	-0.130***	(0.018)	-0.101***	(0.011)
Constant	13.343***	(1.344)	6.946***	(1.283)	4.616***	(1.283)
Indiv. characteristics	Yes		Yes		Yes	
N.obs.	1536		1536		1536	
R-squared	0.267		0.065		0.075	

Notes: standard errors robust for clustering at the session level (in parentheses).

Table C.1: Linear regression on decision times.

References

Binmore, K. (1999). Why experiment in economics? *The Economic Journal 109*(453), 16–24.

Burks, S., J. Carpenter, L. Goette, and A. Rustichini (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences 106*(19), 7745.

Chandler, P. and J. Sweller (1991). Cognitive load theory and the format of instruction. *Cognition and instruction 8*(4), 293–332.

Dave, C., C. Eckel, C. Johnson, and C. Rojas (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty 41*, 219–243. 10.1007/s11166-010-9103-z.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review 100*(3), 1238–1260.

Eckel, C., P. Grossman, C. Johnson, A. Milano, C. Rojas, and R. Wilson (2007). Explaining Risk Preferences in High School Students: A Preliminary Look. Working paper, University of Texas at Dallas.

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

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics 10*(2), 171–178.

Fowler, F. (1995). *Improving survey questions: Design and evaluation.* Sage Publications, Inc.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives 19*(4), 25–42.

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

Kalyuga, S., P. Chandler, J. Tuovinen, and J. Sweller (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology 93*(3), 579.

Mousavi, S., R. Low, and J. Sweller (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology 87*(2), 319–334.

Oechssler, J., A. Roider, and P. Schmitz (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior and Organization 72*(1), 147–152.

Simon, H. A. (1971). *Computers, Communication, and the Public Interest,*, Chapter Designing Organizations for an Information-Rich World, pp. 40–41. The Johns Hopkins Press,.

Smith, V. (2002). Method in experiment: rhetoric and reality. *Experimental economics 5*(2), 91–110.

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. *Science 221*, 453.

van Merriënboer, J. and J. Sweller (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review 17*(2), 147–177.

van Merriënboer, J. and J. Sweller (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical education 44*(1), 85–93.

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it