5-6-2022

On the Generalizability of Using Mobile Devices to Conduct Economic Experiments

Yiting Guo
Wuhan University, yitingguo@whu.edu.cn

Jason Shachat
Chapman University, jason.shachat@durham.ac.uk

Matthew J. Walker
Newcastle University, matt.walker@newcastle.ac.uk

Lijia Wei
Wuhan University, ljwei@whu.edu.cn

On the Generalizability of Using Mobile Devices to Conduct Economic Experiments

Comments

ESI Working Paper 22-05

On the generalizability of using mobile devices to conduct economic experiments*

Yiting Guo[1], Jason Shachat[2], Matthew J. Walker[3], and Lijia Wei[4]

May 06, 2022

Abstract

Recent technological advances enable the implementation of online, field and hybrid experiments using mobile devices. Mobile devices enable sampling of incentivized decisions in more representative samples, consequently increasing the generalizability of results. Generalizability might be compromised, however, if the device is a relevant behavioural confound. This paper reports on a battery of common economic games and decision-making tasks in which we systematically randomize the decision-making device (computer versus mobile phone) and the laboratory setup (physical versus online). The results offer broad support for conducting decision experiments using mobile devices. For six out of eight tasks, we find robust null results in terms of average treatment effects and variability. This should give researchers confidence to conduct studies out-of-laboratory via mobile phones. However, we find two caveats. First, with respect to decisions, subjects using a mobile phone are significantly more risk averse and offer less during bargaining. Second, decision response times and the time taken to read instructions are significantly shorter for the online-mobile treatment. These caveats suggest the importance of ensuring device consistency across treatments in the digital age of experimentation.

Keywords: mobile phone, digitization, methodology, experiment, generalizability

JEL Classification: C90, C93, C70

---

* All experiments in this study were approved by the Academy of Humanities and Social Sciences, Wuhan University.

[1] Economics & Management School, Wuhan University, Wuhan, China, HUB 430072 (yitingguo@whu.edu.cn).

[2] Durham University Business School, Millhill Ln, Durham DH1 3LB, UK (corresponding author: jason.shachat@durham.ac.uk) and Wuhan University.

[3] Newcastle University Business School, Newcastle Upon Tyne, NE1 4SE UK (matt.walker@newcastle.ac.uk).

[4] Economics & Management School, Wuhan University, Wuhan, China, HUB 430072 (ljwei@whu.edu.cn).

1 Introduction

A traditional methodological strength of laboratory economic experiments is the ability to apply strict control to factors such as communication, information and external stimuli. This control instils confidence in the internal validity of such results, a confidence reinforced by successful replication projects (e.g., Camerer et al. 2016). However, laboratory experiments also suffer from certain long-standing criticisms and limitations. Laboratory studies typically rely heavily on the participation of students, leading to commonly expressed concerns of generalizability. Physical laboratory studies also limit the type of decision tasks and research questions one can pursue. Time limits for tasks and interactions, cohort sizes limitations, and difficulties of establishing true anonymity are all constraints facing someone designing a physical laboratory study.

Recent advances in digital technologies and software based on web applications optimized for mobile compatibility facilitate the implementation of field, online and hybrid experiments using portable and low-cost mobile devices.[1] Mobile laboratories with remote setups enable the measurement of economic preferences in more representative samples where the field provides external validity and fills the gap between the physical laboratory and naturally occurring data (see Harrison and List 2004, Levitt and List 2009). Mobile phones and tablets are increasingly used to implement field experiments and large-scale surveys (Himelein 2021). They also afford rapid deployment to monitor individual attitudes and behaviours, and to collect digitized and geocoded data instantaneously in times of crisis, such as natural disasters (e.g., Beine et al. 2020) or public health emergencies (e.g., Lohmann et al. 2021).

An understudied yet fundamental aspect of experimentation in the digital age is the decision-making device. We believe this design aspect to be fundamental to the generalizability of economic experiments because – at least until human brains can communicate directly with computers – a physical decision interface will remain a necessary condition for the controlled measurement of economic preferences.[2] Generalizability would be compromised if the device used to implement

[1] Some prominent examples of these software platforms are oTree (Chen et al. 2016), z-Tree *unleashed* (Duch et al. 2020), LIONESS (Giamattei et al. 2020) and Qualtrics (Qualtrics, Provo, UT).

[2] Brain-computer interfaces are no longer the realm of science fiction (see Willett et al. 2021). Also relevant to this topic is the use of mechanical methods (e.g., rice) to conduct economic experiments in the field (see Tognetti et al. 2012).

the experiment is a relevant behavioural confound that is not adequately controlled for in the design process. In this paper, we investigate this claim by systematically varying the decision-making device used by subjects in a set of widely adopted experimental tasks across physical and remote setups. Participants in our experiment are randomly assigned *ex ante* to complete the experiment using either a computer or mobile phone device, in either a physical laboratory or remote setting. Compliance with the assigned device is verified by measuring the device resolution and screen dimensions *ex post*. We thereby circumvent endogeneity problems inherent in subjects choosing their preferred device. We further control for the role of device ownership by running variants with either public or privately owned computers.

We selected eight classic behavioural economics tasks to test in our experiment based on their frequent use not just as primary outcome variables, but also as extraneous measurements of underlying preferences. These tasks are the Dictator Game, Trust Game, Ultimatum Game, Beauty Contest game, Prisoner's Dilemma game, Stag Hunt game, a number-reporting task to infer dishonest behaviour, and a sequence of lottery choices to measure risk preferences in the gain domain. From the perspective of generalizability, our results are broadly positive: for six out of eight tasks, we find robust null results in terms of average treatment effects across decision-making devices both in- and out-of-laboratory. The effect sizes observed are small enough in magnitude that we would require very large sample sizes to detect a significant difference at conventional statistical levels. There is only weak evidence of greater noise in the mobile phone data, and this is limited to a subset of tasks. Overall, these findings should give researchers confidence to conduct studies out-of-laboratory and via mobile phones.

There are three important qualifications to our results. First, we find that subjects who are randomly assigned to complete the experiment using a mobile phone display significantly greater risk aversion, and offer significantly less during ultimatum bargaining, than those subjects who are randomly assigned to complete the experiment using a computer device. These findings are robust to addressing the threat of multiple hypothesis testing and to controlling for subjects' observed characteristics. These findings extend to both average treatment effects and first-order stochastic dominance of the mobile phone sample distribution of responses. Thus, researchers conducting economic experiments in which objective or strategic risk-taking are important for the decision analysis should pay attention to the device used to elicit measurements. For treatment comparisons, ensuring a consistent device across treatment arms is likely to be sufficient.

Second, we find using a within-subjects analysis that the pairwise correlation between different outcome measurements depends on the decision-making device. That is, the device may mediate behavioural spillover effects between tasks in a non-random manner.[3] Establishing control over the device is therefore particularly important when eliciting more than one preference measurement in a session, which is often the case due to budget constraints. In practice, when conducting experiments online and given a mix of observed devices in the underlying sample, controlling for the device in analyses at the individual-level is preferable.

Third, subjects' decision response times and the time spent on reading task instructions are impacted by both the environment and the device used. We find that decision response times are shorter for those subjects assigned to the mobile phone online treatment. We find that the time reading instructions is longer for those assigned to the mobile phone in the laboratory treatment, but shorter for those assigned to the mobile phone online treatment. Accordingly, researchers should consider adopting device consistent controls in complex experiments where concerns about response time and instruction attention are likely to be more salient behavioural drivers.

Two features of our experimental design are noteworthy. First, the necessity to conduct remote experiments throughout the Covid-19 pandemic has forced the hand of the experimental and behavioural economics community to loosen long held norms that the controls offered by physical laboratory experiments are a minimal acceptable research standard. As this community moves towards more prevalent use of mobile devices and experimentation online, they face a gap in knowing how the change in control affects behaviour (see also Buso et al. 2021 and Li et al. 2021).[4] To the extent that the Covid-19 pandemic caused transitory shifts in economic preferences over

[3] For a discussion of behavioural spillover effects in traditional lab experiments, see Bednar et al. (2012).

[4] The National Science Foundation recognised the potential of the online environment for behavioural research early on (see Bainbridge 2007). In terms of taxonomy, the online experiment sits in the region between the traditional controlled lab experiment and the more natural setting of the field experiment; Charness et al. (2013) refer to experiments in this region as "extra-laboratory" experiments.

time, a comparative strength of our experimental data is that both physical and online lab responses were collected *contemporaneously* and *before the Covid-19 pandemic*.[5]

Second, our protocols hold constant a larger number of elements than other physical versus remote participation studies. These elements include the subject pool, recruitment, randomization and matching protocols, experimenter communication channel, monetary stakes, and payment technology. That is, we tighten the *ceteris paribus* assumption relative to existing study designs (see Table A1 in the appendix for more details). Previous research that investigates the generalizability of incentivized experiments since the advent of the digital age has typically focused on the loss of control in online environments. Early studies tested for individual-level differences in consumption and savings decisions (Anderhub et al. 2001) lottery evaluations (Shavit et al. 2001) and trust (Charness et al. 2007), finding similar behaviour on average but larger variance, lower risk aversion and attenuated social preferences online. Other studies have successfully replicated experiments and behavioural anomalies on representative MTurk samples (Horton et al. 2011, Amir et al. 2012, Gupta et al. 2021, Snowberg and Yariv 2021) and even virtual world platforms (Chesney et al. 2009, Fiedler and Haruvy 2009).[6]

The above cited designs vary the subject pool, protocols and (often) monetary stakes between lab and online samples. This is necessary to establish the generalizability of lab findings with standard student subjects to non-standard subject pools and crowdsourcing labour platforms, but introduces a potential confound in assessing the validity of the remote laboratory environment. For interactive experiments, voluntary dropouts are also a significant challenge online and introduce concerns over differential attrition across treatments (see Arechar et al. 2018 for a discussion). To address some of these limitations, Hergueux and Jacquemet (2015) build an innovative experimental environment that maintains the same (student) subject pool, stakes and interface between lab and online samples. They find qualitatively similar social and risk preferences in the

[5] The mobile phone treatment data used in this study served as the baseline pre-pandemic sample in two related experimental papers examining how the pandemic shifted pro-social and risk-related preferences (Shachat et al. 2021a, Shachat et al. 2021b).

[6] The variable data quality of online labour platforms has been discussed extensively in the literature. We direct the interested reader to Peer et al. (2021) for an up-to-date account and references.

online elicitation. Dickinson and McEvoy (2021) observe an increase in dishonesty when moving from the physical lab to a remote setup using the same subject pool.

Online experiments are typically deployed through the web browser and so it is difficult to control the device used by the subject. Use of a smartphone device may increase measurement error due, for example, to smaller screens sizes, lower response times or a greater propensity to multi-task (Lugtig and Toepoel 2016). Human-computer interactions may also be influenced by the touch interface and ownership (Brasel and Gips 2014, Melumad and Pham 2020), whether via psychological channels (e.g., emotional benefits, sense of privacy) or functional mechanisms (e.g., touch interface, compactness of information).

2 Experimental Design

2.1 Decision-making tasks

The experiments reported in this study were conducted in the Spring of 2019, using the laboratory database of the Center for Behavioral and Experimental Research in Wuhan University, China. Each subject participated in seven incentivized economic games or preference elicitation tasks in sequence. These tasks were as follows: Dictator Game (DG); Beauty Contest (BC); Truth-Telling (TT); Stag Hunt (SH); Prisoner's Dilemma (PD); Risk Preference (RP); and Trust game (TG) or Ultimatum Game (UG).[7] Only one of the TG and UG was included in each session to mitigate behavioural spillover effects for second movers between these two tasks (the sample sizes for these two tasks are correspondingly smaller).

Below, we provide a brief description of the players, action sets, and payoffs in each task. The tasks were programmed using oTree software (Chen et al. 2016).

Task 1. *DG*. Subjects are randomly matched into pairs. Within a pair, subjects are randomly assigned to the role of either Player 1 or Player 2. Player 1 is allotted 5 RMB and decides how much of this endowment to send to Player 2. Player 2 has no decision to make. This task measures pure altruistic preferences.

[7] A limitation of our design is that we did not randomize the order of task presentation; nevertheless, this protocol is consistent across treatments.

Task 2. *BC*. Subjects are randomly divided into groups of four. Within a group, subjects choose an integer between 0 and 100 (inclusive) The subject whose guess is closest to one-half of the average value selected within the group wins 8 RMB (ties broken evenly); the remaining subjects earn zero payoff for the task. This task measures levels of rationality and strategic thinking.

Task 3. *TT*. Each subject chooses an integer between 0 and 9 (inclusive) and adds this to the final digit of their own student ID number (unobserved by the experimenter), keeping in mind the ones digit of the resulting sum. A random integer between 0 and 9 is then displayed on-screen. If this number matches the ones digit, then the subject earns 5 RMB; else zero payoff. This task measures preferences for truth-telling (inferred at the aggregate level).

Task 4. *SH*. Subjects are randomly matched into pairs. Each player within a pair simultaneous chooses either Option A or Option B. If both players choose A, then both players earn 3 RMB. If both players choose B, then both players earn 8 RMB. If one player chooses A and the other player chooses B, then the first player earns 3 RMB and the second player earns 0 RMB. This task measures preferences to coordinate on the risk-dominant (A) or efficient (B) equilibrium.

Task 5. *PD*. Subjects are randomly matched into pairs. Each player within a pair simultaneous chooses either Option C or Option D. If both players choose C, then both players earn 6 RMB. If both players choose D, then both players earn 3 RMB. If one player chooses C and the other player chooses D, then the first player earns 0 RMB and the second player earns 9 RMB. This task measures preferences to cooperate (C) or defect (D).

Task 6. *RP.* Each subject is presented with a series of nine pairwise choices between a lottery (option A) and a sure amount of money (option B). The lottery remains fixed across all choices: a 50% chance of receiving 9 RMB, and a 50% chance of receiving 3 RMB. The sure amount increases evenly with each choice from 3 RMB up to 9 RMB. After all choices have been made, the system randomly selects one of the nine pairs of options for payment. This task measures risk tolerance (a greater number of lottery choices indicates a greater willingness to take risks).[8]

Task 7. *TG*. Subjects are randomly matched into pairs. Within a pair, subjects are assigned to the role of either Player 1 or Player 2. Player 1 is allotted 8 RMB and decides how much of this endowment to send to Player 2. Any money sent is multiplied by a factor of three before reaching

[8] In our construction of the risk tolerance variable below, risk neutrality corresponds to a score of 5.5.

Player 2. Any money not sent is kept by Player 1. Player 2 observes the multiplied amount sent and decides how much of it to return to Player 1. Any money not returned is kept by Player 2. This task measures levels of trust and reciprocity.

Task 8. *UG*. Subjects are randomly matched into pairs. Within a pair, subjects are assigned to the role of either Player 1 or Player 2. Player 1 is allotted 8 RMB and decides how much of this endowment to send to Player 2. Player 2 can accept or reject the allocation. In case of rejection, both players receive zero payoff. This task measures fairness preferences.

2.2 Treatments and protocols

Our design randomizes two factors: first, the laboratory setup (physical versus online); second, the decision-making device (personal computer [laptop] versus mobile phone [smartphone]).[9] For the online experiments, we verified compliance with the randomly assigned device ex post by recording information about the system and screen dimensions. To check for any effect on decision-making of device ownership, we implemented an additional variant in the physical lab in which we supplied subjects with a laptop computer owned by the laboratory. Hence, there are five experimental treatments, using a between-subjects design (Table 1).

Table 1 – Treatment matrix (N=581)[a]

	Computer		Mobile Phone
	Personal	Public	Personal
Physical Lab	n =108 (3*28 + 1*24)	n =112 (4*28)	n = 112 (4*28)
Online	n =160 (4*28 + 2*24)		n = 112 (4*28)

Notes: Terms in parentheses are (number of sessions * number of subjects in the session).
[a] We exclude 23 subjects from our final dataset for using the wrong device to complete the experiment (4 in Lab/Computer, 13 in Online/Computer, 6 in Online/Mobile).

[9] In practice, all subjects in the computer treatments used a laptop rather than desktop computer to complete the experiment. Based on screen dimension data, we were able to verify this for both the physical lab and online.

Participants were students registered on a wide range of academic majors at Wuhan University (mean age = 20.6, 65% female see Table A2). Both in the physical lab and online, the recruitment of participants, randomization protocol and payment transfers were executed using the Ancademy platform for conducting social science experiments (https://www.ancademy.org/). Ancademy is based on the open interface of WeChat. Invitations were sent at random to members of this database (c. 9,000 members) to participate in a session at a scheduled time. The day before a session, all subjects received a confirmation message specifying the device needed for participation (computer or mobile phone) and details of how to sign in to the Ancademy platform at the scheduled time (whether remote or in the physical lab).

Each session followed the same procedure. After all subjects had signed in, a six-digit quick join code was distributed; subjects were informed that this code would enable them to rejoin the session quickly in case of disconnection. During the session, a private communication channel with the experimenter was available via WeChat for questions or clarifications. For the two-player games, matching was conducted simultaneously at the end of a session to ensure joint determination of payoffs. Upon conclusion of the session, earnings were transferred directly to subjects' WeChat wallets within 24 hours. Subjects were paid based on choices in all decision tasks and no feedback was provided until the completion of all tasks.

We conducted a total of 22 sessions across the five treatments: 12 sessions in the physical lab (4 sessions with mobile phone, 4 sessions with public computer, 4 sessions with personal computer); and 10 sessions online (4 sessions with mobile phone, 6 sessions with personal computer). At the end of a session, subjects completed a short questionnaire with questions about their age, gender, and family background. Average earnings were 40.6 RMB (approximately 6 US Dollars), including a show-up fee of 10 RMB. A session lasted approximately 30 minutes. We recruited 28 subjects for each session.[10] We exclude data from 23 subjects who failed to comply with the assigned device to complete the experiment. Thus, the final sample size is N=581.[11]

[10] Due to no-shows on the day, three sessions only had 24 subjects (see Table 1).

[11] For the RP task analysis only, we further exclude 10 "inconsistent" subjects who switch from the lottery to the safe option more than once in the list. We return to discuss these inconsistent responses below in the context of data variability among treatments.

Across the computer treatments, we find no significant differences in behaviour between the use of personal and public computers (see Table A3) and so we pool this data below.

In summary, across treatments we hold constant the subject pool, recruitment and protocols, availability of the experimenter communication channel, monetary stakes, and payment technology. Our design mitigates involuntary dropouts (there is no attrition) and permits verification of the randomly assigned device in both physical lab and remote settings.

3 Results

3.1 Generalizability of the device and remote setup

Table 2 presents the pooled (lab and online) statistics for sessions in which subjects were randomly assigned to use either a computer or a mobile phone to complete the decision-making tasks. Table 3 presents the pooled (computer and mobile device) statistics for sessions in which subjects were randomly assigned to participate either in the physical lab or online. Since no feedback is provided until after the completion of all tasks, we use the subject as the independent level of observation. To address the threat of multiple hypothesis testing and the possibility of false positives, we calculate False Discovery Rate (FDR) adjusted q-values across the ten outcome measurements, based on two-sided Wilcoxon rank-sum tests (Benjamini et al., 2006).

Result 1. *For six out of eight decision-making tasks, we find no significant average treatment effect of the device on outcomes.*

There are no significant differences in pure altruistic preferences, trust or cooperation between the computer and mobile phone treatments (all q-values > 0.320). Amounts sent by dictators in the DG are around 30% of the endowment independently of the assigned device. There are similar relative differences between amounts sent in the DG and amounts offered in the UG in both sets of treatments, although the mobile phone sample exhibits higher variance (see Table A4). Trustees send around 40% of the endowment in the TG on both devices, and this is a breakeven strategy on average based on the trustor's response. We also observe similar rates of cooperation in the PD game - around one-third of subjects choose to cooperate - and in the SH game - around 88% of subjects select the efficient outcome.

Table 2 – Decision-making using a computer versus mobile phone device.

	Computer			*q*-value	*Mobile phone*		
	n	Mean	SD		n	Mean	SD
DG Sent [0,5]	185	1.575	0.992	0.321	106	1.416	1.079
BC Guess [0,100]	363	28.725	19.823	0.868	218	27.592	17.192
TT Match {0,1}	363	0.733	0.443	0.264	218	0.670	0.471
SH Efficient {0,1}	363	0.879	0.327	0.945	218	0.881	0.325
PD Cooperate {0,1}	363	0.325	0.469	0.924	218	0.317	0.466
RP Tolerance {1, 2,…,10}	355	4.789	1.268	0.012*	212	4.443	1.111
TG Sent [0,8]	93	3.188	2.524	0.850	55	3.345	2.612
TG Return [0,3*Sent]	93	2.930	4.584	0.470	55	3.309	3.983
UG Offer [0,8]	92	3.448	0.951	0.025*	51	3.010	1.051
UG Accept {0,1}	92	0.902	0.299	0.044*	51	0.745	0.440

Notes: Multiple hypothesis testing adjusted False Discovery Rate (FDR) *q*-values (10 comparisons) based on two-sided Wilcoxon rank-sum tests, $*q < 0.05$.

Table 3 – Decision-making in the physical lab versus online.

	Lab			*q*-value	*Online*		
	n	Mean	SD		n	Mean	SD
DG Sent [0,5]	166	1.517	0.993	0.895	125	1.516	1.070
BC Guess [0,100]	328	28.271	19.019	0.895	253	28.336	18.716
TT Match {0,1}	328	0.704	0.457	0.895	253	0.715	0.452
SH Efficient {0,1}	328	0.881	0.324	0.895	253	0.877	0.329
PD Cooperate {0,1}	328	0.317	0.466	0.895	253	0.328	0.470
RP Tolerance {1,2,…,10}	321	4.717	1.226	0.386	246	4.585	1.215
TG Sent [0,8]	82	3.305	2.663	0.895	66	3.174	2.419
TG Return [0,3*Sent]	82	3.104	4.594	0.895	66	3.030	4.086
UG Offer [0,8]	84	3.283	1.029	0.895	59	3.305	0.983
UG Accept {0,1}	84	0.833	0.375	0.895	59	0.864	0.345

Notes: Multiple hypothesis testing adjusted False Discovery Rate (FDR) *q*-values (10 comparisons) based on two-sided Wilcoxon rank-sum tests.

At the aggregate level, a preference for truth-telling appears to be slightly higher in the mobile phone data, where 67% report matches, than in the computer data, where 73.3% report matches, but this difference is not significant (q-value = 0.264). That is, most participants in both samples choose to lie to get a higher payoff in the TT task. Neither is there any significant difference in strategic reasoning between samples using different devices (q-value = 0.868). Average guesses in the BC game (rounded to the nearest integer) are 29 in the computer sample and 28 in the mobile phone sample.

Result 2. *Subjects on average display greater risk aversion and offer less during ultimatum bargaining when randomly assigned to complete the experiment using a mobile device.*

We observe significantly lower ultimatum offers (q-value = 0.025) and lower acceptance rates (q-value = 0.044) when subjects bargain with a mobile phone. Average UG offers are 43.1% of the endowment when using a computer versus 37.6% of the endowment when using a mobile phone; the acceptance rates are 90.2% and 74.5%, respectively. Subjects in the mobile phone treatments also display significantly greater risk aversion than those in the computer treatments. Although subjects in both samples exhibit risk aversion in the RP task overall (score < 5.5), the degree of risk aversion is larger for those subjects using a mobile phone (q-value = 0.012).

Result 3. *We find no significant average treatment effect of the remote setup on behaviour.*

There are no significant differences in average behaviour between the physical lab and online setups for any of the eight decision-making tasks (all q-values > 0.385). This result supports previous studies (e.g., Hergueux and Jacquemet 2015) as to the high internal validity of the online laboratory after holding a range of other design aspects constant.

3.2 Statistical power and robustness of the null treatment effects

In Table 4, we calculate – separately for each of the behavioural measurements – the required sample size for our study to detect a significant effect at the 5% statistical level and with 80% power, alongside the actual effect size and the minimal detectable effect size given the number of subjects in our four between-subjects samples. The purpose of this analysis is to demonstrate that the null results observed above are robust and not likely to be due to a lack of statistical power.

The results of this analysis suggest that, in the six decision-making tasks for which we found null average treatment effects (Result 1), the effect sizes observed in our experiment would require very large sample sizes for a well-powered study to detect a significant effect. The required sample sizes range from 687 to 8,025,766 subjects. For all 10 online versus physical lab statistical comparisons, and for 5 out of 7 null computer versus mobile phone comparisons, the required sample size is 4-digits or more.

Table 4 – Required sample size and minimal detectable effects.

	Online vs. Physical Lab			Computer vs. Mobile phone		
	Required sample size	Cohen's d effect size	Min. detectable effect	Required sample size	Cohen's d effect size	Min. detectable effect
DG Sent [0,5]	8,025,766	0.001	0.330	687	0.155	0.358
BC Guess [0,100]	1,404,109	0.003	4.456	4,568	0.060	4.372
TT Match {0,1}	27,382	0.025	0.107	853	0.139	0.111
SH Efficient {0,1}	132,855	0.011	0.076	461,543	0.006	0.078
PD Cooperate {0,1}	29,806	0.023	0.109	49,190	0.018	0.112
RP Tolerance {1,2,…,10}	1,427	0.107	0.291	204	0.285*	0.285
TG Sent [0,8]	6,301	0.051	1.233	4,346	0.062	1.229
TG Return [0,3*Sent]	58,675	0.017	2.127	2,187	0.087	2.008
UG Offer [0,8]	33,587	0.022	0.489	85	0.444*	0.497
UG Accept {0,1}	33,587	0.022	0.178	85	0.444*	0.193

Notes: Required sample size in each group is calculated by G*Power with 80% power and 0.05 significance level (two tails). Cohen's d effect size: calculated by R package effsize. Min. detectable effect with 80% power and 0.05 significance level, * FDR q-value < 0.05 (see Table 2).

3.3 Covariates and interaction effects

The aggregate statistics reported so far neither control for observed subject characteristics, nor consider interaction effects. To address these concerns, we regress each of the decision-making outcomes on the full interaction between the device (computer versus mobile phone) and experiment setup (lab versus online), controlling for age, gender, monthly expenditure, and academic major. Estimation is using OLS for the continuous outcomes (Table 5) and logistic regression for the binary outcomes (Table 6). We report heteroskedasticity-consistent standard errors. We again calculate FDR adjusted q-values for the mobile phone treatment dummy and these are reported alongside conventional p-values in the regression output tables.

Table 5 – Linear regression analysis of decision-making in the experiment.

	Dependent Variable					
	DG Sent	BC Guess	RP Tolerance	UG Offer	TG Sent	TG Return
	(1)	(2)	(3)	(4)	(5)	(6)
Online	-0.003	-0.62	-0.24*	-0.07	-0.20	-0.31
	(0.15)	(2.08)	(0.13)	(0.22)	(0.59)	(0.58)
	[0.985]	[0.958]	[0.711]	[0.958]	[0.958]	[0.958]
Mobile	-0.14	-1.12	-0.48***	-0.57***	0.36	-0.25
	(0.16)	(2.07)	(0.13)	(0.22)	(0.73)	(0.72)
	[0.782]	[0.887]	[0.003]	[0.0498]	[0.887]	[0.909]
Online * Mobile	-0.10	-1.1	0.23	0.30	-0.35	0.56
	(0.25)	(3.23)	(0.21)	(0.34)	(0.94)	(0.76)
TG Sent						1.42***
						(0.09)
Constant	3.69***	37.82***	5.76***	4.13***	5.58***	-1.15
	(0.71)	(11.90)	(0.67)	(0.86)	(2.76)	(2.79)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	291	581	567	143	148	148

Notes: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors are shown in the parentheses, calculated using the Huber/White sandwich estimator of variance. Multiple hypothesis testing adjusted False Discovery Rate (FDR) q-values in square brackets (10 comparisons) for online and mobile treatment dummies. Control variables include subject age, gender, monthly expenditure, and academic major.

Table 6 – Logistic regression analysis of decision-making in the experiment.

	Dependent Variable			
	TT Match	PD Cooperate	SH Efficient	UG Accept
	(7)	(8)	(9)	(10)
Online	0.15	0.02	-0.15	-0.55
	(0.25)	(0.24)	(0.33)	(0.94)
	[0.958]	[0.985]	[0.958]	[0.958]
Mobile	-0.32	-0.03	0.02	-1.6
	(0.26)	(0.26)	(0.37)	(1.18)
	[0.563]	[0.962]	[0.962]	[0.563]
Online * Mobile	-0.01	0.18	-0.13	1.28
	(0.39)	(0.38)	(0.55)	(1.32)
UG Offer				2.49***
				(0.56)
Constant	-0.33	-3.22**	2.03	9.71**
	(1.41)	(1.59)	(2.06)	(4.09)
Control variables	Yes	Yes	Yes	Yes
Observations	581	581	581	143

Notes: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors are shown in the parentheses, calculated using the Huber/White sandwich estimator of variance. Multiple hypothesis testing adjusted False Discovery Rate (FDR) q-values in square brackets (10 comparisons) for online and mobile treatment dummies. Control variables include subject age, gender, monthly expenditure, and academic major.

Consistent with the aggregate findings, there is greater risk aversion in the RP task both online and when using a mobile phone and this effect is only statistically robust on the mobile device (q-value = 0.003). UG offers remain significantly lower with a mobile device after controlling for covariates (q-value = 0.0498); however, UG acceptance rates no longer differ significantly between devices after accounting for the difference in offers. As expected, UG acceptance rates are significantly increasing in offers and TG returns are significantly increasing in TG amounts sent across all treatments. We find no significant interaction effect between the mobile device and online setup in any of the decision-making tasks.

3.4 Distribution of responses

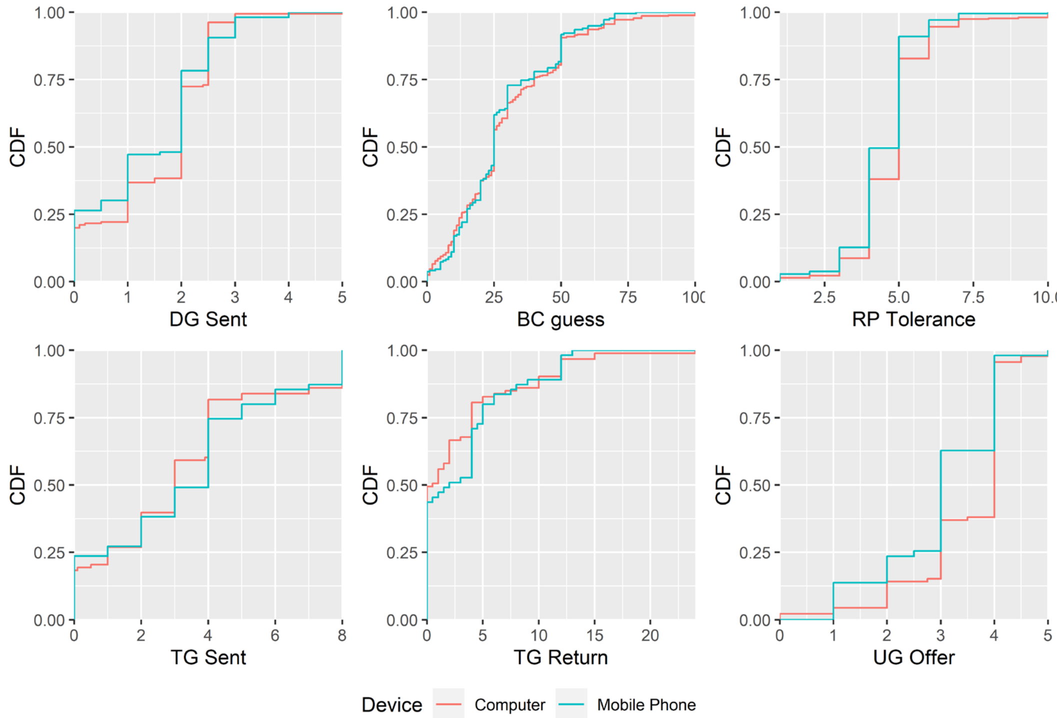
In Panel (a) of Figure 1, we present the empirical distributions of our continuous behavioural measurements for the two independent samples using different devices. There is a clear pattern: the computer sample exhibits a first-order stochastic dominance relationship with the mobile phone sample for the RP Tolerance and UG Offer measurements. The mobile phone effects serve as a lower-bound on the distribution of risk tolerance and ultimatum offers in our experiment. This implies that Result 2 is not driven by a few extreme outlier subjects and may be better interpreted as population shifts. In Panel (b) of Figure 1 we plot the corresponding cumulative distributions for the physical lab and online samples. Consistent with Result 3, there are no discernible differences in behaviour across the distribution.

3.5 Variance-covariance of responses and similarity of correlations

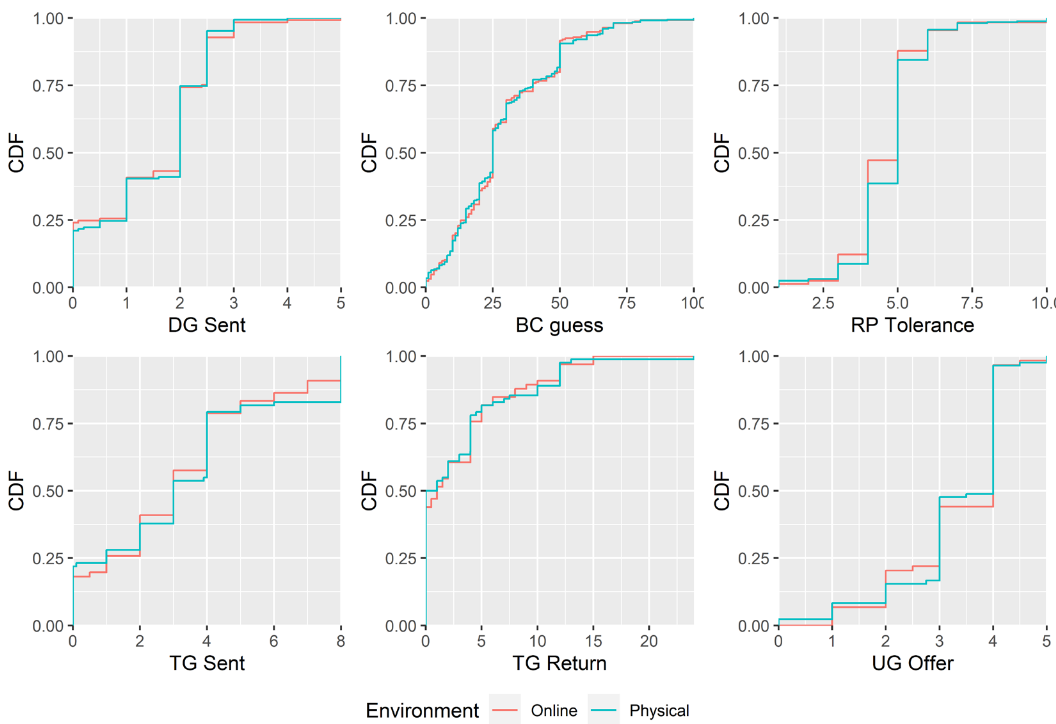
To examine whether decision-making is systematically noisier among subjects randomly assigned to participate using a mobile device or remote setup, which we might expect if these factors increase measurement error, we first consider the coefficients of variation (see Tables A6 and A7).[12] Based on Feltz and Miller tests (Feltz and Miller 1996), we find that the UG acceptance rate is more variable in the mobile phone sample than in the computer sample (p-value < 0.001), and that there is weak evidence of greater variation on the mobile device for UG offers (p-value = 0.072) and reported TT matches (p-value = 0.060). There are, however, no differences in variability for the remaining seven behavioural measurements.

[12] Li et al. (2021) find that online experiment data is less noisy when using a webcam-on protocol.

Figure 1 – Distributions of responses by device and laboratory environment.



Panel (a): Device.



Panel (b): Laboratory environment.

Comparisons of coefficients of variation for the physical lab versus online samples yield qualitatively similar conclusions.[13] Thus, on this metric, there is at best weak evidence to support the initial conjecture.

A further indicator of reduced data quality across samples is provided by the number of "inconsistent" responses in the RP task, i.e., those subjects who switch from the lottery to the safe option more than once in the list. Frequencies of inconsistent responses are similar on the two devices (1.6% for the computer treatment and 1.8% for the mobile phone treatment). The rate of inconsistent responses in the RP task is also only marginally higher online (2.0%) than in the physical lab (1.5%).

Finally, as Snowberg and Yariv (2021) point out, experimental economists often care about correlations between different behaviours and attributes. Since experimentalists are often subject to budget constraints and wish to maximize the usability of their datasets, they often elicit more than one preference measurement in a session. Consistency of pairwise correlations among these different measurements are thus of interest. We use Snowberg and Yariv's similarity of correlations approach to consider, respectively, how different behavioural measurements relate to one another between decision-making devices and laboratory environments.

The findings are presented in Figure 2 and Table 7. We record the sign and significance (at the 10 percent level) of pairwise correlations between all measurements. Panel (a) in Figure 2 relates to the device and Panel (b) in the figure relates to the environment. We classify a cell in which the pairwise correlations are qualitatively and statistically the same between samples as "Complete agreement". We classify a cell in which one sign is significant positive/negative and the other sign is not significantly different from zero as "Partial disagreement". If one sign in the cell is significant positive and the other sign is significant negative, then this is classified as "Complete disagreement".

[13] There are no significant differences in variation at the 5% level based on the Feltz and Miller test. One comparison (TT match) is significant at the 10% level, but this measurement is more variable in the physical lab than online.

Figure 2 – Within-subjects correlations across decision-making device and lab environment.

Mobile Phone vs. Computer	DG Sent	BC Guess	TT Match	SH Efficient	PD Cooperate	RP Tolerance
DG Sent						
BC Guess	+					
TT Match	0-	0				
SH Efficient	0	0	0			
PD Cooperate	0+	+0	-0	+		
RP Tolerance	-0	-0	0	0	-	
TG Sent	+	0	-	0	0	0
TG Return		0	0-	0	0	0+
UG Offer	+	+0	0	0	-+	0
UG Accept		0	0	0	0	0

Panel (a): Decision-making device. Mobile Phone vs. Computer.

Online vs. Physical Lab	DG Sent	BC Guess	TT Match	SH Efficient	PD Cooperate	RP Tolerance
DG Sent						
BC Guess	+0					
TT Match	-	0				
SH Efficient	0	0	0			
PD Cooperate	0	+	-0	+		
RP Tolerance	0	0	0	0	-	
TG Sent	+	+0	-	0	0	0
TG Return		0+	0-	0	0	0+
UG Offer	+	0+	0	0	0	-0
UG Accept		0	0	0	+0	0

Panel (b): Laboratory environment. Online vs. Physical Lab.

	Complete agreement			Partial disagreement		Complete disagreement

Notes: A "+" denotes a significant positive correlation, a "-" donates a statistically significant negative correlation and a "0" donates an insignificant correlation. We use a single symbol if the two signs in the same cell agree. "Complete agreement": the two signs in a cell are the same. "Partial disagreement": one sign in the cell is significant positive/negative, the other sign is insignificant. "Complete disagreement": one sign in the cell is significant positive, the other sign is significant negative. Only one of the TG and UG was included in each session to mitigate behavioural spillover effects for second movers between these two tasks.

Table 7 – Summary: similarity of correlations analysis.

	Phone vs. Computer	Online vs. Physical Lab
Complete agreement	27 (72.97%)	28 (75.68%)
Partial disagreement	9 (24.32%)	9 (24.32%)
Complete disagreement	1 (2.70%)	0 (0%)

Notes. In total 37 cells in each panel, the number (percent) of cells that are complete agreement / partial disagreement / complete disagreement (for definitions of terms, see the notes to Figure 2).

We observe greater inconsistency between the mobile phone and computer samples than between the physical lab and online samples. Out of 37 pairwise correlations, 1 indicates complete disagreement, with a significant negative correlation between PD Cooperate and UG Offer in the phone sample and a significant positive correlation in the computer sample. A further 9 cases (24.32 percent) show partial disagreement, while the remaining correlations feature a complete agreement in the sign and significance (72.97 percent). No pairwise correlation displays complete disagreement between the online and physical lab samples.

4 The Mediating Effect of Response and Instruction Times

What might explain our finding that subjects display greater risk aversion and offer less during ultimatum bargaining when randomly assigned to complete the experiment using a mobile device (Result 2)? One plausible mechanism is related to decision response time. If subjects are prone to decide more quickly on a mobile phone device than on a computer, perhaps due to perceived differences in time pressure, naturalness or decision heuristics, then this might manifest itself via changes in risky behaviour. There is some experimental evidence to suggest that time pressure may increase risk aversion in gains and risk seeking in losses (e.g., Cahlíková and Cingl 2016, Kocher et al. 2013, Kirchler et al. 2017). Both the RP and UG tasks involve risk, objective or strategic.

To explore this possibility, we first conduct separate OLS regressions of decision response time and instructions reading time on our treatment dummies and individual-level covariates, pooled across all eight tasks (Table 8). Both variables have long tails to the right and so we take the logarithmic transformations. The results of this analysis suggest that overall decision response time is significantly lower among those subjects assigned to complete the experiment using a mobile phone device online, relative to the traditional physical lab setting with computers (p-value = 0.030).

Relative to this benchmark, we also find that subjects using a mobile phone device online spend significantly less time reading the instructions (p-value = 0.012); the opposite is true for subjects assigned to complete the experiment in the physical lab using a mobile device (p-value = 0.006). There is, however, substantial heterogeneity among tasks. Suggestively, subjects on average take significantly less time to respond to the risk elicitation on a mobile device (44 seconds) than on a computer (49 seconds) after adjusting for multiple hypothesis testing (q-value = 0.019); there is no significant difference in instructions reading time for this task (see Tables A8 to A11).

Table 8 – Regression analysis of decision response time and instructions reading time (in seconds).

Dependent variable (ln):	Decision response time	Instructions reading time
	(1)	(2)
Computer & Online	0.004	0.05
	(0.04)	(0.04)
Mobile & Physical Lab	-0.03	0.11***
	(0.05)	(0.04)
Mobile & Online	-0.15**	-0.16**
	(0.07)	(0.06)
Age	0.02***	0.01
	(0.01)	(0.01)
Constant	2.63***	3.41***
	(0.27)	(0.20)
Control variables	Yes	Yes
Observations (clusters)	3,777[a] (581)	4,067[b] (581)

Notes: *p<0.1; **p<0.05; ***p<0.01. This table reports coefficient estimates from OLS regressions. The dependent variable in (1) is how much time subjects spend on making their decision in a given task. The dependent variable in (2) is how much time subjects spend on reading the instructions in a given task. Both variables are in natural logarithmic form. Standard errors clustered at the individual level are reported in parentheses. Control variables include subject age (reported), gender, monthly expenditure, and academic major. The constant term represents the Computer & Physical Lab treatment cell.

[a] (7 tasks x 581 individuals) – (290 recipients in the Dictator Game) = 3,777.

[b] 7 instruction sets x 581 individuals = 4,067.

To test whether differential decision response or instructions time can plausibly explain variation in behaviours in our experiment, we re-run the 10 regressions in Tables 5 and 6 controlling for decision response and instructions reading time as covariates (see Tables A12 and A13). The significant negative effect of the mobile treatment dummy on risk tolerance remains a robust finding (q-value = 0.007). The corresponding negative effect on UG offers no longer survives multiple hypothesis testing (q-value = 0.094). We find no economically or statistically significant effect of variability in individual-level decision response or instructions reading time on risk aversion, or of instruction reading time on UG offers. There is, however, a negative association between UG offers and decision response time, which is significant at the 5% level. This suggests that faster decision times on the mobile device may mediate the lower observed offers.

We note, tangentially, that for our sample there is a strong inverse statistical relationship between decision response time and the level of (naïve) equilibrium behaviour in the BC game – naïve to the extent that lower guesses do not necessarily increase the probability of winning. This is consistent with earlier evidence that deciding more slowly produces faster convergence to equilibrium behaviour (Kocher and Sutter 2006). Subjects who decide more quickly are also significantly more likely to cooperate in the PD game and SH game (see Rand et al. 2012, on the intuitiveness of cooperation).[14]

5 Concluding Remarks

Rapid developments in digital technology have led to a paradigm shift in opportunities for the conduct of incentivized decision-making experiments in settings where traditional laboratory experiments are not feasible. With a shift towards more remote learning and working likely to persist, and a potential reduction in the time cost of experiment management, implementation of experiments using low-cost mobile phone devices will remain an attractive option in the experimentalist's toolkit. To investigate whether generalizability might be compromised by the nature of the decision-making device, we presented evidence from a battery of economic games and decision-making tasks in which we randomly assigned the device (computer versus mobile phone) and the laboratory setup (physical versus online), holding constant the subject pool, experimental protocols, communication channel, monetary stakes, and payment technology.

As noted in the introduction, the mobile phone treatment data collected for this study served as the baseline sample in separate work that we conducted investigating the behavioural consequences of the Covid-19 pandemic (Shachat et al. 2021a, Shachat et al. 2021b). The ability to deploy incentivized decision-making experiments via mobile devices in Wuhan at the onset of the pandemic was crucial in enabling us to monitor individual attitudes and behaviours in real time. We used the mobile phone treatment data, rather than the computer treatment data, as the baseline

[14] The strength of evidence on the intuitiveness of cooperation is disputed (see, e.g., Tinghög et al. 2013). There is no evidence in our sample to suggest that subjects who take more time to decide are less likely to report a TT match (see Shalvi et al. 2012, for evidence that honesty requires time).

sample in that work to ensure consistency of the decision-making device between pre- and post-pandemic samples. That is, we were concerned that generalizability might be compromised if the device used to implement the experiment is a relevant behavioural confound that is not adequately controlled for in the design process.

While the findings of the present study offer support for conducting decision experiments using mobile devices across a class of common behavioural economics instruments designed to measure pro-sociality, cooperation, and strategic reasoning, they also suggest that we were right to be cautious. In our study, subjects who are randomly assigned to complete the experiment using a mobile phone are significantly more risk averse and offer significantly less during bargaining than those who are randomly assigned to use a computer. These findings survive multiple hypothesis testing, are robust to controlling for observed individual characteristics and extend to first-order stochastic dominance of the outcome distributions. Within-subjects correlational analyses also indicate systematic behavioural differences that are influenced by the decision-making device.

We identify response and instruction time as a potential behavioural mechanism for divergent behaviour across devices. However, we admit this is likely not a complete explanation as there remains residual differences in assessed risk aversion. One behavioural explanation, which our study design doesn't permit evaluation of, is that individuals experience a greater endowment effect and aversion to the risk of loss when making decisions using their mobile phone (Kahneman et al. 1991). Hein et al. (2011) observe that, compared with traditional laptop and desktop computers, tablet and smartphone devices may induce a greater association with an individual's extended self. Wang and Nelson (2014) argue that even if this is seen as a relationship role instead of an extension of self, the bond with touch devices is closer than the bond with other kinds of devices. Since this channel focuses on the "self" rather than the "other", it would also not contradict our null finding between devices in our measures of "pure" social preferences (altruism, trust and cooperation).

In the hierarchy of List (2021), our study constitutes a "Wave 2" study that delves deeper into the boundary conditions required for the generalizability of decision-making experiments in settings where physical lab experiments are nether the feasible nor natural choice. There are of course many background factors that may influence human behaviour in a particular setting. Some factors are likely to be more important than others. The traditional methodological strength of laboratory economic experiments is the ability to apply as much control over those factors as possible. Future work will continue to explore threats to the generalizability of online, field and hybrid experiments, in which control is necessarily reduced. As List (p. 5) surmises, it is through the discovery of such factors that science progresses.

References

Amir, Ofra; Rand, David G. and Gal, Ya'akov Kobi. "Economic Games On the Internet: The Effect of $1 Stakes." *PLoS One*, 2012, *7*(2), pp. e31461.

Anderhub, Vital; Müller, Rudolf and Schmidt, Carsten. "Design and Evaluation of an Economic Experiment Via the Internet." *Journal of Economic Behavior & Organization*, 2001, *46*(2), pp. 227-47.

Arechar, Antonio A.; Gächter, Simon and Molleman, Lucas. "Conducting Interactive Experiments Online." *Experimental Economics*, 2018, *21*(1), pp. 99-131.

Bainbridge, William Sims. "The Scientific Research Potential of Virtual Worlds." *Science*, 2007, *317*(5837), pp. 472-76.

Bednar, Jenna; Chen, Yan; Liu, Tracy Xiao; and Page, Scott. "Behavioral spillovers and cognitive load in multiple games: An experimental study." *Games and Economic Behavior*, 2012, *74*(1), pp. 12-31.

Beine, Michel A.R.; Charness, Gary; Dupuy, Arnaud; and Joxhe, Majlinda. "Shaking things up: on the stability of risk and time preferences." Working Paper, 2020.

Benjamini, Yoav; Krieger, Abba M. and Yekutieli, Daniel. "Adaptive linear step-up procedures that control the false discovery rate." *Biometrika*, 2006, *93*(3), pp. 491-507.

Brasel, S. Adam and Gips, James. "Tablets, Touchscreens, and Touchpads: How Varying Touch Interfaces Trigger Psychological Ownership and Endowment." *Journal of Consumer Psychology*, 2014, *24*(2), pp. 226-33.

Buso, Irene Maria; Di Cagno, Daniela; Ferrari, Lorenzo; Larocca, Vittorio; Lorè, Luisa; Marazzi, Francesca; Panaccione, Luca and Spadoni, Lorenzo. "Lab-Like Findings From Online Experiments." *Journal of the Economic Science Association*, 2021, *7*(2), pp. 184-93.

Cahlíková, Jana; and Cingl, Lubomir. "Risk preferences under acute stress." *Experimental Economics*, 2017, *20*(1), pp. 209-236.

Camerer, Colin F.; Dreber, Anna; Forsell, Eskill; Ho, Teck-Hua; Huber, Jürgen; Johannesson, Magnus, et al. "Evaluating replicability of laboratory experiments in economics. " *Science,* 2016, *351*(6280), pp. 1433-1436.

Charness, Gary; Gneezy, Uri and Kuhn, Michael A. "Experimental Methods: Extra-Laboratory Experiments-Extending the Reach of Experimental Economics." *Journal of Economic Behavior & Organization*, 2013, *91*, pp. 93-100.

Charness, Gary; Haruvy, Ernan and Sonsino, Doron. "Social Distance and Reciprocity: An Internet Experiment." *Journal of Economic Behavior & Organization*, 2007, *63*(1), pp. 88-103.

Chen, Daniel L.; Schonger, Martin and Wickens, Chris. "oTree—An open-source platform for laboratory, online, and field experiments." *Journal of Behavioral and Experimental Finance*, 2016, *9*, pp. 88-97.

Chesney, Thomas; Chuah, Swee-Hoon and Hoffmann, Robert. "Virtual World Experimentation: An Exploratory Study." *Journal of Economic Behavior & Organization*, 2009, *72*(1), pp. 618-35.

Dickinson, David L. and McEvoy, David M. "Further From the Truth: The Impact of Moving From In-Person to Online Settings On Dishonest Behavior." *Journal of Behavioral and Experimental Economics*, 2021, *90*, pp. 101649.

Duch, Matthias L.; Grossman, Max R.P. and Lauer, Thomas. "z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet." *Journal of Behavioral and Experimental Finance*, 2020, *28*, pp. 100400.

Feltz, Carol J. and Miller, G. Edward. "An Asymptotic Test for the Equality of Coefficients of Variation From K Populations." *Statistics in Medicine*, 1996, *15*(6), pp. 647-58.

Fiedler, Marina and Haruvy, Ernan. "The Lab Versus the Virtual Lab and Virtual Field—an Experimental Investigation of Trust Games with Communication." *Journal of Economic Behavior & Organization*, 2009, *72*(2), pp. 716-24.

Giamattei, Marcus; Yahosseini, Kyanoush S.; Gächter, S and Molleman, L. "LIONESS Lab: a free web-based platform for conducting interactive experiments online." *Journal of the Economic Science Association,* 2020, *6*(1), pp. 95-111.

Gupta, Neeraja; Rigotti, Luca and Wilson, Alistair. "The Experimenters' Dilemma: Inferential Preferences Over Populations", *ArXiv Preprint*, 2021, arXiv:2107.05064.

Harrison, Glenn W.; and List, John A. "Field experiments." *Journal of Economic literature*, 2004, *42*(4), pp. 1009-1055.

Hein, Wendy; O'Donohoe, Stephanie and Ryan, Annmarie. "Mobile Phones as an Extension of the Participant Observer's Self: Reflections On the Emergent Role of an Emergent Technology." *Qualitative Market Research*, 2011, *14*(3), pp. 258-273.

Hergueux, Jérôme and Jacquemet, Nicolas. "Social Preferences in the Online Laboratory: A Randomized Experiment." *Experimental Economics*, 2015, *18*(2), pp. 251-83.

Himelein, K. (2021, January 19). "Improved targeting for mobile phone surveys: A public-private data collaboration." Retrieved from https://blogs.worldbank.org/impactevaluations/improved-targeting-mobile-phone-surveys-public-private-data-collaboration-guest.

Horton, John J.; Rand, David G. and Zeckhauser, Richard J. "The Online Laboratory: Conducting Experiments in a Real Labor Market." *Experimental Economics*, 2011, *14*(3), pp. 399-425.

Kahneman, D.; Knetsch, Jack L. and Thaler, Richard H. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives*, 1991, *5*(1), pp. 193-206.

Kirchler, Michael; Andersson, David; Bonn, Caroline; Johannesson, Magnus; Sørensen, Erik Ø.; Stefan, Matthias; Tinghög, Gustav; and Västfjäll, Daniel. "The effect of fast and slow decisions on risk taking." *Journal of Risk and Uncertainty*, 2017, *54*(1), pp. 37-59.

Kocher, Martin G.; Pahlke, Julius; and Trautmann, Stefan T. "Tempus fugit: time pressure in risky decisions." *Management Science*, 2013, *59*(10), pp. 2380-2391.

Kocher, Martin G.; and Sutter, Matthias. "Time is money—Time pressure, incentives, and the quality of decision-making." *Journal of Economic Behavior & Organization*, 2006, *61*(3), pp. 375-392.

Levitt, Steven D. and List, John A. "Field experiments in economics: The past, the present, and the future." *European Economic Review*, 2009, *53*(1), pp. 1-18.

Li, Jiawei; Leider, Stephen; Beil, Damian and Duenyas, Izak. "Running Online Experiments Using Web-Conferencing Software." *Journal of the Economic Science Association*, 2021, *7*(2), pp. 167-83.

List, John A. "Non Est Disputandum de Generalizability? A Glimpse into the External Validity Trial." Working Paper, 2021.

Lohmann, Paul; Gsottbauer, Elisabeth; You, Jing; and Kontoleon, Andreas. "Social Preferences and Economic Decision-Making in the Wake of COVID-19: Experimental Evidence from China." Working Paper, 2021.

Lugtig, Peter, and Toepoel, Vera. "The use of PCs, smartphones, and tablets in a probability-based panel survey: Effects on survey measurement error." *Social Science Computer Review*, 2016, *34*(1), pp. 78-94.

Melumad, Shiri and Pham, Michel Tuan. "The Smartphone as a Pacifying Technology." *Journal of Consumer Research*, 2020, *47*(2), pp. 237-55.

Peer, Eyal; Rothschild, David; Gordon, Andrew; Evernden, Zak and Damer, Ekaterina. "Data Quality of Platforms and Panels for Online Behavioral Research." *Behavior Research Methods*, 2021, pp. 1-20.

Rand, David G.; Greene, Joshua D.; and Nowak, Martin A. "Spontaneous giving and calculated greed." *Nature*, 2012, *489*, pp. 427-430.

Shachat, Jason; Walker, Matthew J. and Wei, Lijia. "How the onset of the Covid-19 pandemic impacted pro-social behaviour and individual preferences: Experimental evidence from China." *Journal of Economic Behavior & Organization*, 2021a, *190*, pp. 80-494.

Shachat, Jason; Walker, Matthew J. and Wei, Lijia. "The Impact of an Epidemic: Experimental Evidence on Preference Stability from Wuhan." *AEA Papers and Proceedings*, 2021b, *111*, pp. 302-06.

Shalvi, Shaul; Eldar, Ori; and Bereby-Meyer, Yoella. "Honesty requires time (and lack of justifications)." *Psychological Science*, 2012, *23*(10), pp. 1264-1270.

Shavit, Tal; Sonsino, Doron and Benzion, Uri. "A Comparative Study of Lotteries-Evaluation in Class and On the Web." *Journal of Economic Psychology*, 2001, *22*(4), pp. 483-91.

Snowberg, Erik and Yariv, Leeat. "Testing the Waters: Behavior Across Participant Pools." *American Economic Review*, 2021, *111*(2), pp. 687-719.

Tinghög, Gustav; Andersson, David; Bonn, Caroline; Böttiger, Harald; Josephson, Camilla; Lundgren, Gustaf; et al. "Intuition and cooperation reconsidered." *Nature*, 2013, *498*, E1–E2.

Tognetti, Arnaud; Berticat, Claire; Raymond, Michel; and Faurie, Charlotte. "Sexual Selection of Human Cooperative Behaviour: An Experimental Study in Rural Senegal." PLoS ONE, 2012, *7*(9).

Wang, Zongyuan and Nelson, Michelle R. "Tablet as Human: How Intensity and Stability of the User-Tablet Relationship Influences Users' Impression Formation of Tablet Computers." *Computers in Human Behavior*, 2014, *37*, pp. 81-93.

Willett, Francis R.; Avansino, Donald T.; Hochberg, Leigh R.; et al. "High-performance brain-to-text communication via handwriting." *Nature*, 2021, *593*, pp. 249–254.

Appendix A: Additional Tables and Figures.

Table A1 – Design aspects held constant between online and physical laboratories.

	Subject pool[a]	*Recruitment*[b]	*Matching protocol*[c]	*Experimenter communication*[d]	*Mitigate dropouts*[e]	*Payment*[f]	*Device identified*[g]
Anderhub et al. (2001)	✕	✕	✕	✓	✓	✓	--
Shavit et al. (2001)	✕	✓	✕	--	--	✕	--
Charness et al. (2007)	✕	--	--	--	--	✓	--
Horton et al. (2011)	✕	--	--	--	--	✓	--
Amir et al. (2012)	✕	--	--	--	--	--	--
Gupta et al. (2021)	✕	--	✓	--	--	✓	--
Snowberg & Yariv (2021)	✓	--	✕	--	--	✓	--
Chesney et al. (2009)	✕	--	--	--	--	--	--
Fiedler & Haruvy (2009)	✕	--	✓	--	--	✓	--
Arechar et al. (2018)	✕	--	--	--	✓	✓	--
Hergueux & Jacquemet (2015)	✓	✓	✕	--	✓	✓	--

Notes. A "✓" ("✕") means a similar (different) design in lab and online environment. A "--" means that this aspect was not explicitly mentioned in the article.

a. The same subject pool.
b. Invitations sent in advance, no information about the task.
c. Simultaneous matching at end of session to ensure joint determination of payoffs.
d. The same communication channel with the experimenter to answer questions on instructions.
e. Option to rejoin after network disconnection.
f. Payment based on outcomes of all tasks in local currency
g. Device used to complete the online experiment identified.

Table A2 – Sample demographic statistics.

Variable	*Treatment*	Overall	Physical Computer (Public)	Physical Computer (Personal)	Physical Mobile	Online Computer	Online mobile
Gender (%)	Male	35.112	45.872	30.841	41.964	34.694	21.698
	Female	64.888	54.128	69.159	58.036	65.306	78.330
Age (mean)	Years	20.570	20.404	21.533	20.688	20.551	19.670
Monthly expenditure in RMB (%)	< 800	3.270	4.587	1.869	6.250	2.721	0.943
	800-1500	41.480	44.954	33.645	33.929	42.177	52.830
	1500-2500	47.160	45.872	50.467	51.786	47.619	39.623
	2500-4000	7.401	4.587	13.084	7.143	6.803	5.660
	> 4000	0.688	0	0.935	0.893	0.680	0.943
Operating system mobile (%)	iOS	26.606	--	--	28.571	--	24.528
	Android	73.394	--	--	71.429	--	75.472
Operating system computer (%)	Mac	5.510	0	10.280	--	6.122	
	Windows	94.490	100	89.720	--	93.878	--
Screen Height (mm)	Mean	650.76	610.385	672.299	--	688.578	--
Phone size (inch)	Mean	5.741	--	--	5.687	--	5.775

Table A3 – Decision-making using a personal versus public computer in the physical lab.

	Personal computer			*q*-value	*Public computer*		
	n	Mean	SD		n	Mean	SD
DG Sent [0,5]	54	1.459	1.031	1.00	56	1.661	0.944
BC Guess [0,100]	107	29.673	21.757	1.00	109	27.780	17.302
TT Match {0,1}	107	0.720	0.451	1.00	109	0.734	0.444
SH Efficient {0,1}	107	0.860	0.349	1.00	109	0.899	0.303
PD Cooperate {0,1}	107	0.355	0.481	1.00	109	0.303	0.462
RP Tolerance {1,2,…,10}	107	4.813	1.267	1.00	105	4.914	1.324
TG Sent [0,8]	26	3.038	2.418	1.00	28	3.321	2.722
TG Return [0,3*Sent]	26	2.962	5.481	1.00	28	2.893	3.705
UG Offer [0,8]	28	3.509	1.037	1.00	28	3.446	0.975
UG Accept {0,1}	28	0.893	0.315	1.00	28	0.893	0.315

Notes: Multiple hypothesis testing adjusted False Discovery Rate (FDR) *q*-values (10 comparisons) based on two-sided Wilcoxon rank-sum tests.

Table A4 – UG Offer versus DG Sent using a computer versus mobile phone device.

	UG Offer (pct. of endowment)			*p*-value	DG Sent (pct. of endowment)		
Computer							
	n	Mean	SD		n	Mean	SD
	92	0.431	0.119	< 0.001	185	0.322	0.194
Mobile phone							
	51	0.376	0.131	< 0.001	106	0.259	0.231

Notes: The *p*-values are based on one-sided Wilcoxon signed-rank tests (UG > DG).

Table A5 – UG Offer versus DG Sent in the physical lab versus online.

	UG Offer (pct. of endowment)			*p*-value	DG Sent (pct. of endowment)		
Lab							
	n	Mean	SD		n	Mean	SD
	84	0.410	0.129	< 0.001	166	0.302	0.205
Online							
	59	0.413	0.123	< 0.001	125	0.297	0.217

Notes: The *p*-values are based on one-sided Wilcoxon signed-rank tests (UG > DG).

Table A6 – Coefficient of variation analysis using a computer versus mobile phone device.

Coefficient of variation	*Computer*	*Mobile phone*	*p*-value
DG Sent [0,5]	0.630	0.762	0.103
BC Guess [0,100]	0.690	0.623	0.226
TT Match {0,1}	0.605	0.704	0.060
SH Efficient {0,1}	0.372	0.369	0.904
PD Cooperate {0,1}	1.443	1.473	0.882
RP Tolerance {1,2,…,10}	0.265	0.250	0.383
TG Sent [0,8]	0.792	0.781	0.939
TG Return [0,3*Sent]	1.565	1.204	0.357
UG Offer [0,8]	0.276	0.349	0.072
UG Accept {0,1}	0.331	0.591	<0.001

Notes: The *p*-values are based on Feltz and Miller tests for the equality of coefficients of variation between samples.

Table A7 – Coefficient of variation analysis in the physical lab versus online.

Coefficient of variation	*Lab*	*Online*	*p*-value
DG Sent [0,5]	0.655	0.706	0.516
BC Guess [0,100]	0.673	0.660	0.226
TT Match {0,1}	0.649	0.632	0.060
SH Efficient {0,1}	0.368	0.374	0.904
PD Cooperate {0,1}	1.470	1.434	0.882
RP Tolerance {1,2,…,10}	0.260	0.265	0.766
TG Sent [0,8]	0.806	0.762	0.751
TG Return [0,3*Sent]	1.480	1.348	0.725
UG Offer [0,8]	0.313	0.297	0.691
UG Accept {0,1}	0.450	0.399	0.406

Notes: The *p*-values are based on Feltz and Miller tests for the equality of coefficients of variation between samples.

Table A8 – Decision response time (seconds) using a computer versus mobile phone device.

	Computer		q-value	*Mobile phone*	
	Mean	SD		Mean	SD
DG Sent [0,5]	23.535	11.780	0.436	25.792	14.310
BC Guess [0,100]	34.317	29.249	0.019*	29.431	28.578
TT Match {0,1}	27.441	21.080	0.062	23.119	13.932
SH Efficient {0,1}	13.650	13.736	0.808	13.596	14.708
PD Cooperate {0,1}	12.309	14.109	<0.001*	9.532	13.583
RP Tolerance {1, 2,…,10}	49.066	30.966	0.019*	44.463	27.164
TG Sent [0,8]	34.409	37.808	0.251	26.545	24.560
TG Return [0,3*Sent]	36.833	28.997	0.573	31.929	18.222
UG Offer [0,8]	16.913	11.599	0.403	21.569	18.315
UG Accept {0,1}	16.273	14.505	0.469	15.696	7.596

Notes: Multiple hypothesis testing adjusted False Discovery Rate (FDR) q-values (10 comparisons) based on two-sided Wilcoxon rank-sum tests, *$q < 0.05$.

Table A9 – Decision response time (seconds) in the physical lab versus online.

	Lab		q-value	*Online*	
	Mean	SD		Mean	SD
DG Sent [0,5]	25.265	14.013	0.654	23.152	10.873
BC Guess [0,100]	35.866	32.345	0.013*	28.099	23.524
TT Match {0,1}	26.826	17.174	0.015*	24.514	20.731
SH Efficient {0,1}	13.793	14.793	0.723	13.419	13.163
PD Cooperate {0,1}	11.262	13.584	0.873	11.273	14.476
RP Tolerance {1, 2,…,10}	49.311	31.548	0.290	44.783	26.856
TG Sent [0,8]	32.049	30.053	0.723	30.788	37.825
TG Return [0,3*Sent]	30.915	18.286	0.290	40.125	31.821
UG Offer [0,8]	18.750	14.210	0.793	18.322	14.949
UG Accept {0,1}	16.113	14.693	0.654	15.969	8.391

Notes: Multiple hypothesis testing adjusted False Discovery Rate (FDR) q-values (10 comparisons) based on two-sided Wilcoxon rank-sum tests, *$q < 0.05$.

Table A10 – Instructions reading time (seconds) using a computer versus mobile phone device.

	Computer		q-value	*Mobile phone*	
	Mean	SD		Mean	SD
DG Sent [0,5]	28.157	42.745	0.874	25.292	14.153
BC Guess [0,100]	60.628	28.644	0.874	60.248	28.642
TT Match {0,1}	49.532	25.044	0.874	47.138	22.912
SH Efficient {0,1}	59.074	32.236	0.874	61.569	34.312
PD Cooperate {0,1}	43.419	27.348	0.703	45.142	24.420
RP Tolerance {1, 2,…,10}	32.512	20.335	0.189	34.817	18.302
TG Sent [0,8]	34.957	18.072	0.874	34.727	23.846
TG Return [0,3*Sent]	32.978	19.591	0.874	32.607	16.512
UG Offer [0,8]	25.467	12.955	0.874	24.098	9.888
UG Accept {0,1}	24.773	11.813	0.874	28.375	15.848

Notes: Multiple hypothesis testing adjusted False Discovery Rate (FDR) q-values (10 comparisons) based on two-sided Wilcoxon rank-sum tests.

Table A11 – Instructions reading time (seconds) in the physical lab versus online.

	Lab		q-value	*Online*	
	Mean	SD		Mean	SD
DG Sent [0,5]	28.416	44.688	0.884	25.384	14.923
BC Guess [0,100]	62.686	29.005	0.098	57.632	27.910
TT Match {0,1}	48.655	20.294	0.479	48.605	28.666
SH Efficient {0,1}	54.043	21.922	0.012*	67.747	42.188
PD Cooperate {0,1}	44.579	26.053	0.777	43.399	26.608
RP Tolerance {1, 2,…,10}	32.512	16.797	0.884	34.498	22.736
TG Sent [0,8]	35.427	19.658	0.871	34.182	21.265
TG Return [0,3*Sent]	33.683	19.274	0.861	31.750	17.335
UG Offer [0,8]	26.107	12.884	0.479	23.373	10.327
UG Accept {0,1}	26.963	13.374	0.479	25.188	13.891

Notes: Multiple hypothesis testing adjusted False Discovery Rate (FDR) q-values (10 comparisons) based on two-sided Wilcoxon rank-sum tests, $*q < 0.05$.

Table A12 – Linear regression analysis of decision-making in the experiment controlling for decision response and instruction reading time.

	Dependent Variable					
	DG Sent	BC Guess	RP Tolerance	UG Offer	TG Sent	TG Return
	(1)	(2)	(3)	(4)	(5)	(6)
Online	-0.005	-1.056	-0.228*	-0.102	-0.191	-0.034
	(0.148)	(2.035)	(0.132)	(0.215)	(0.577)	(0.532)
	[0.974]	[0.974]	[0.860]	[0.974]	[0.974]	[0.974]
Mobile	-0.092	-0.972	-0.463***	-0.531**	0.304	0.35
	(0.171)	(2.002)	(0.136)	(0.223)	(0.729)	(0.565)
	[0.847]	[0.847]	[0.007]	[0.094]	[0.847]	[0.847]
Online * Mobile	-0.147	-2.738	0.238	0.306	-0.17	0.042
	(0.261)	(3.125)	(0.208)	(0.336)	(0.965)	(0.745)
TG Sent						1.463***
						-0.105
Ln(decision time)	-0.255*	-3.929***	0.191	-0.211**	-0.135	0.131
	(0.151)	(0.998)	(0.149)	(0.101)	(0.283)	(0.295)
Ln(instruction time)	-0.039	-3.678**	-0.047	-0.127	0.616*	-0.406
	(0.133)	(1.565)	(0.124)	(0.126)	(0.358)	(0.413)
Constant	4.467***	63.977***	5.132***	4.829***	3.993	-1.822
	(0.803)	(12.345)	(0.931)	(0.981)	(3.134)	(3.326)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	291	581	567	143	148	146

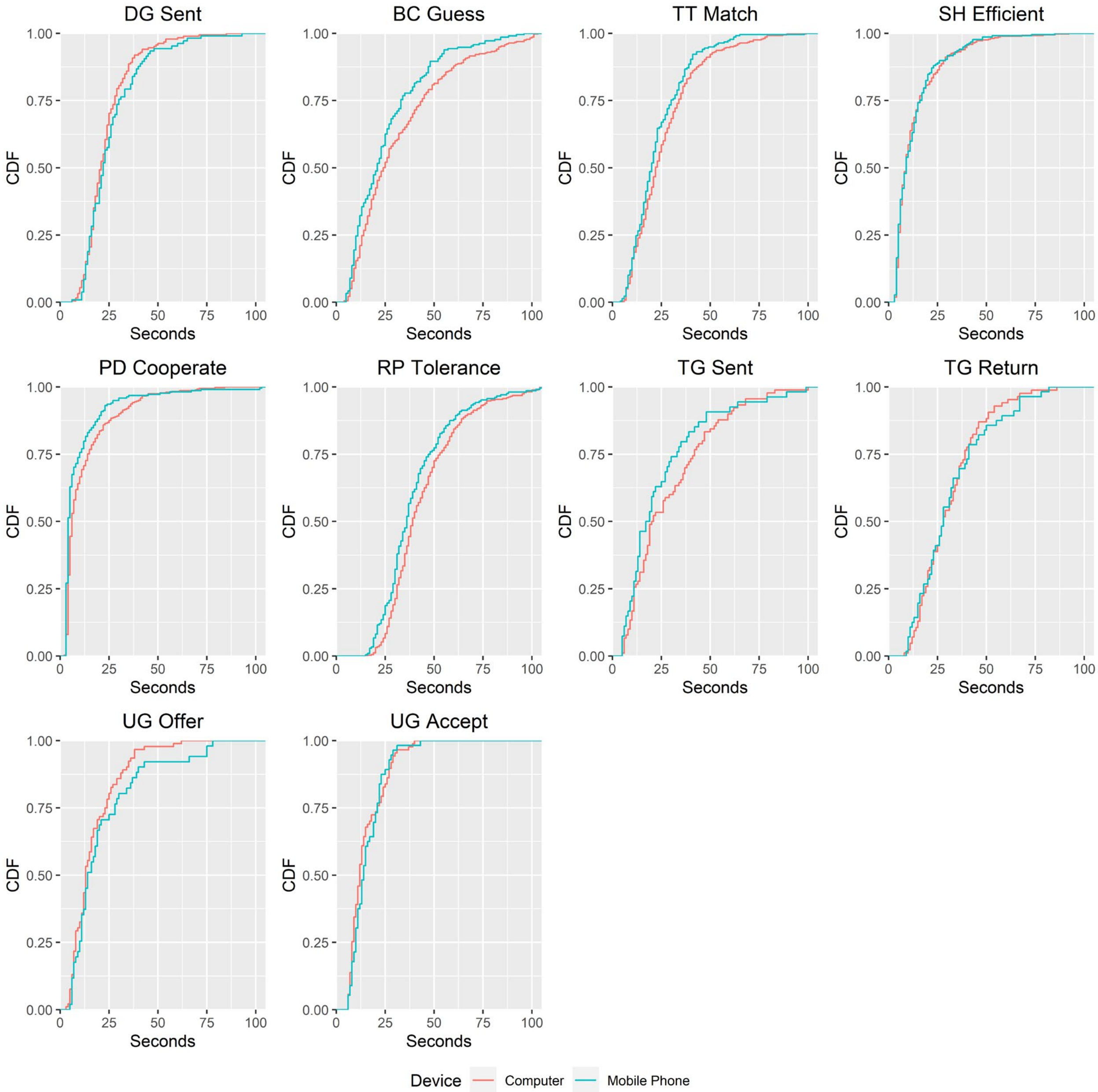
Notes: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors are shown in the parentheses, calculated using the Huber/White sandwich estimator of variance. Multiple hypothesis testing adjusted False Discovery Rate (FDR) q-values in square brackets (10 comparisons) for online and mobile treatment dummies. Control variables include subject age, gender, monthly expenditure, and academic major.

Table A13 – Logistic regression analysis of decision-making in the experiment controlling for decision response and instruction reading time.

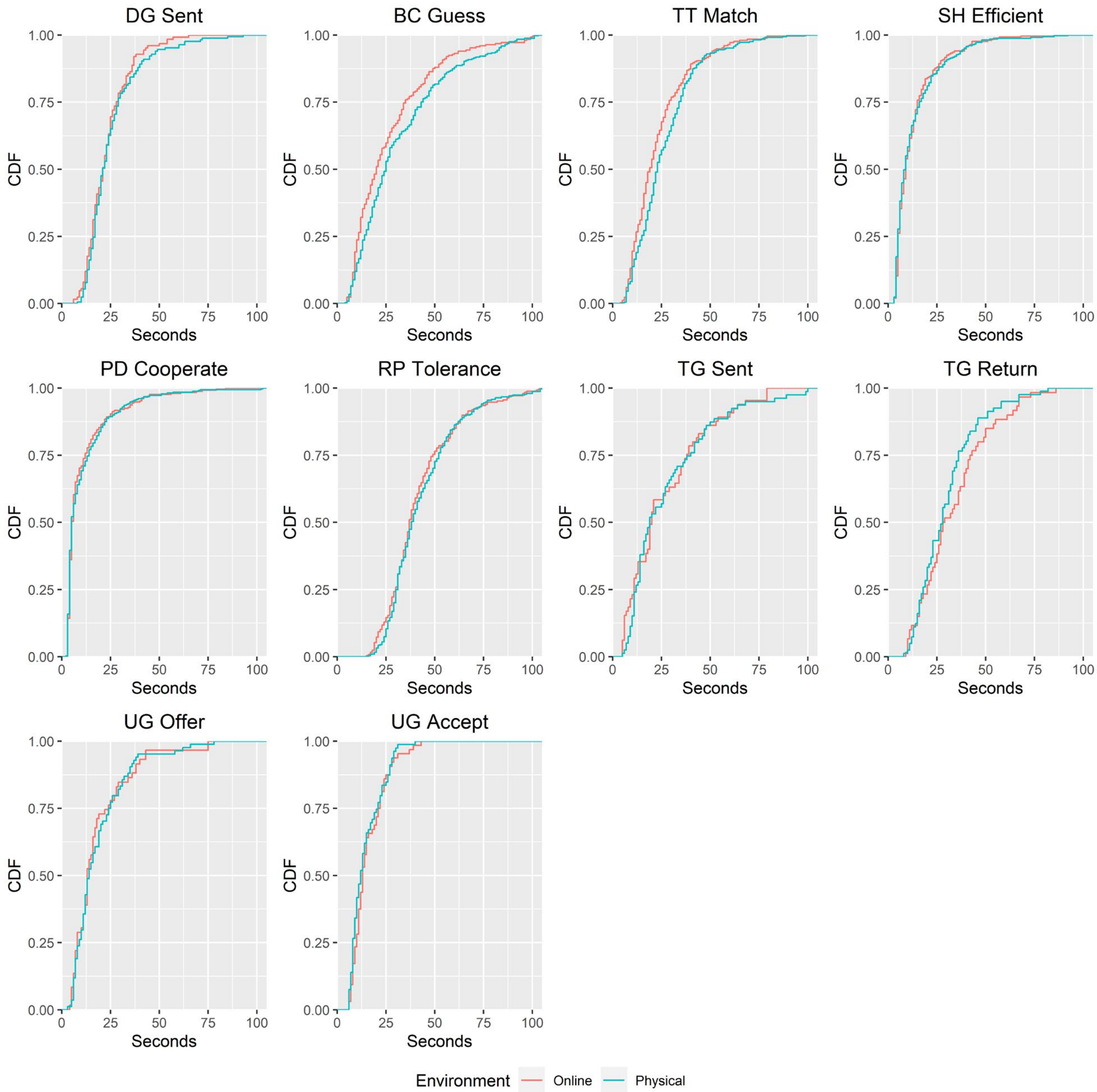
	Dependent Variable			
	TT Match (7)	PD Cooperate (8)	SH Efficient (9)	UG Accept (10)
Online	0.156	0.049	-0.014	0.656
	(0.254)	(0.245)	(0.331)	(1.405)
	[0.974]	[0.974]	[0.974]	[0.974]
Mobile	-0.317	-0.063	0.071	-0.621
	(0.264)	(0.271)	(0.381)	(1.039)
	[0.764]	[0.852]	[0.852]	[0.847]
Online * Mobile	-0.017	0.071	-0.177	-0.047
	(0.392)	(0.391)	(0.553)	(1.660)
UG Offer				2.205***
				(0.378)
Ln(decision time)	0.002	-0.481***	-0.336*	0.916
	(0.168)	(0.132)	(0.179)	(0.658)
Ln(instruction time)	-0.02	-0.576***	-0.553*	-0.12
	(0.194)	(0.154)	(0.300)	(0.820)
Constant	-0.26	-0.443	5.116**	-8.874
	(1.639)	(1.752)	(2.326)	(6.520)
Control variables	Yes	Yes	Yes	Yes
Observations	581	581	581	144

Notes: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors are shown in the parentheses, calculated using the Huber/White sandwich estimator of variance. Multiple hypothesis testing adjusted False Discovery Rate (FDR) q-values in square brackets (10 comparisons) for online and mobile treatment dummies. Control variables include subject age, gender, monthly expenditure, and academic major.

Figure A1 – Distributions of decision response times by device and laboratory environment.



Panel (a): Device.



Panel (b): Laboratory environment.

Notes: Decision response times of more than 105 seconds are excluded for purposes of exposition.

Appendix B: Instructions (translated from original Chinese)

Thank you for participating in this experiment! This experiment includes 8 tasks. Each task is different and you will do each task once. In each task, we will pay you the corresponding payment according to your decision. Your final earnings include two parts: participation fee and payment according to your performance in each task. You can see the results and payment of each task after finishing all tasks.

Please do not communicate with others during the experiment, and do not interrupt the experiment. Randomly interrupting the experiment will result in the invalidation of the experimental data, causing serious losses to the laboratory, and also affect your final earnings. In the process of experiment, once you make the decision, you cannot go back to revise it. Please make the decision carefully.

If the experiment cannot be completed successfully due to unavoidable circumstances (such as network interruption, other participants interrupting the experiment, etc.), we will pay you 10 RMB as your participation fee.

We will transfer you the payment after the end of the experiment through official account "Ancademy". You can withdraw it (enter "Ancademy"-Assistant-Account-YANZHI-Withdrawal) to your WeChat account. If you haven't followed official account "Ancademy", please follow it as soon as possible. In this way, we can transfer you the payment!

Please fill in your mobile phone number in the box below, so that we can contact you if there is any problem during the experiment.

Task 1

The introduction of task 1

In this task, all participants will be randomly divided into groups of two. Two participants in each group will be randomly assigned to roles. One is participant P1, and the other one is participant P2.

At the beginning of the task, participant P1 has 5 RMB. Participant P1 decides how much money (X) to pass on to participant P2. The remaining money is owned by participant P1.

Payment calculation formula: Participant P1: 5-X; Participant P2: X

Your choice

You are the participant P1, please decide how much money to pass on to participant P2.

I will pass to participant P2: [¥]

Your choice

You are the participant P2. You don’t need to make any decision in this task.

Task 2

The introduction of task 2

In this task, all participants will be randomly divided into groups of four. You and all other team members are required to choose an integer between 0 and 100 (including 0 and 100). Half of the average value of all selected numbers is the target value, and the participant whose integer is closest to the target value wins. If more than one participant selects the target value, they are both winners.

For example, four participants in one group respectively select A, B, C, D. Half of the average of the four number is (A+B+C+D) *1/4*1/2, so participant whose number closest to this number wins.

If one participant wins, he will gain 8 RMB in this experiment and other participants gain 0 RMB; If more than one participant wins, the winners will divide 8 RMB equally.

Your Choice

Please enter an integer between 0 and 100 (including 0 and 100):

Task 3

The introduction of task 3

In this task, you need to randomly choose one integer between 0-9 first, and add the integer to the last number of your student number. Please keep the ones digit of the sum in mind. For example,

If you choose 6, and the last number of your student number is 1, you will get number7.

If you choose 4, and the last number of your student number is 9, then the sum of these two is 13. Keep the ones digit of 13, you will get 3.

After the above process is completed, the system will randomly generate a number between 0 and 9 and display it on the screen. You need to tell us whether the number generated by the system is the same as the number you got in the previous process.

If they are the same, you will get the reward of 5 RMB,

If they are different, you will get nothing.

Your Choice

The random number generated by the system is: 1

Is the random number generated by the system the same as the number you got in advance?

- ○ Yes
- ○ No

Task 4

The introduction of task 4

In this task, all participants are randomly divided into two-person groups, and you need to make a decision that has two options: options A and B, where your decision and that of the other participant jointly determine your payment in this task.

The payoff matrix corresponding to your decision and that of the other participant is as follows. In each cell, the first number (in bold) is your payoff, and the second number is the payoff of the other participant. That is,

If you choose A, and the other participant choose A, you will gain 3 RMB, the other participant will gain 3 RMB;

If you choose A, and the other participant choose B, you will gain 3 RMB, the other participant will gain 0 RMB;

If you choose B, and the other participant choose A, you will gain 0 RMB, the other participant will gain 3 RMB;

If you choose B, and the other participant choose B, you will gain 8 RMB, the other participant will gain 8 RMB;

		The other participant	
		A	B
You	A	**￥3.00** ￥3.00	**￥3.00** ￥0.00
	B	**￥0.00** ￥3.00	**￥8.00** ￥8.00

Your Choice

According to the payoff matrix, your choice is:

○ A

○ B

Task 5

The introduction of task 5

In this task, all participants are randomly divided into two-person groups, and you need to make a decision that has two options: options C and D, where your decision and that of the other participant jointly determine your payment in this task.

The payoff matrix corresponding to your decision and that of the other participant is as follows. In each cell, the first number (in bold) is your payoff, and the second number is the payoff of the other participant. That is,

If you choose C, and the other participant choose C, you will 6 RMB, the other participant will gain 6 RMB;

If you choose C, and the other participant choose D, you will gain 0 RMB, the other participant will gain 9 RMB;

If you choose D, and the other participant choose C, you will gain 9 RMB, the other participant will gain 0 RMB;

If you choose D, and the other participant choose D, you will gain 3 RMB, the other participant will gain 3 RMB;

		The other participant	
		C	D
You	C	**￥6.00** ￥6.00	**￥0.00** ￥9.00
	D	**￥9.00** ￥0.00	**￥3.00** ￥3.00

Your Choice

According to the payoff matrix, your choice is:

○ C

○ D

Task 6

The introduction of task 6

Hereinafter, you are presented with nine pairs of options listed on the screen, each of which is a lottery, and you have to choose an option between "option A" and "option B".

"option A" has a 50-50 chance of getting 9 RMB and a 50-50 chance of getting 3 RMB.

“option B” has a certain amount of money.

After you have made all your choices, the system will randomly select one of the nine pairs of options, and depending on which option you choose, A or B, the system will randomly determine your reward in this task in a specified probability.

For example, the system randomly selects the ith pair of options,

If you choose option A in the ith pair of options, you will have a 50-50 chance of getting 9 RMB and a 50-50 chance of getting 3 RMB.

If you choose option B in the ith pair of options, you will a certain amount of money that is determined by option B.

Your choice

Option A	Option B
○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥3.00 for sure(Fixed income)
○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥3.75 for sure(Fixed income)
○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥4.50 for sure(Fixed income)
○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥5.25 for sure(Fixed income)

○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥6.00 for sure(Fixed income)
○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥6.75 for sure(Fixed income)
○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥7.5 for sure(Fixed income)
○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥8.25 for sure(Fixed income)
○a 50-50 chance of getting 9 RMB; a 50-50 chance of getting 3 RMB	○get ￥9.00 for sure(Fixed income)
Submit	

Task 7 (Trust)

The introduction of task 7

In this task, all participants will be randomly divided into groups of two people. One is participant P1, and the other one is participant P2.

At the beginning of the task, participant P1 has an endowment of 8 RMB. Participant P1 decides how much money (X) to pass to participant P2. The amount of money passed on triples before it is handed over to participant P2. After participant P2 receives three times as much money, he decides how much money (Y) to pass on to participant P1.

Payment calculation formula: Participant P1: 8-X+Y; Participant P2: 3X-Y

Your choice

Your role for this task is participant P1. Now you have an endowment of 8 RMB, please decide how much money you are willing to pass to participant P2.

Please enter a number between 0 and 8: [￥]

Your choice

Your role for this task is participant P2. Participant P1 passed on ￥4.00 to you, so you actually receive ￥12.00. Therefore, now you have ￥12.00, how much money are you willing to pass to participant P1?

Please enter a number between 0 and ￥12.00: [￥]

Task 7 (Ultimatum)

The introduction of task 7

In this task, all participants will be randomly divided into groups of two people. One is participant P1, and the other one is participant P2.

At the beginning of the task, participant P1 has an endowment of 8 RMB. Participant P1 decides how much money (X) to pass to participant P2. Participant P2 can accept or reject the proposal.

If participant P2 choose to accept, the two participants in this group will receive the corresponding amount of money according to the delivery scheme of participant P1.

If participant P2 choose to reject, the two participants in this group both receive 0 RMB.

Payment calculation formula:

When participant P2 accept, participant P1: 8-X; participant P2: X;

When participant P2 reject, participant P1: 0; participant P2: 0;

Your choice

Your are participant P1. how much money are you willing to pass to participant P2? [￥]

Your choice

Your are participant P2. Participant P1 decide to pass on ￥3.00 to you.

Please choose to accept or reject the delivery scheme proposed by participant P1?

○ Accept

○ Reject

Task 8: Questionnaire

Please fill in the following questions truthfully.

How old are you, please?

What is your gender, please?

- O Male
- O Female
- O other

What is your monthly allowances, please?

- O Less than 800 RMB
- O 800-1500 RMB
- O 1500-2500 RMB
- O 2500-4000 RMB
- O More than 4000 RMB

What is the annual income of your family, please?

- O Less than 30000 RMB
- O 30000-100000 RMB
- O 100000-200000 RMB
- O 200000-400000 RMB
- O More than 400000 RMB

Which category of the following includes your major, please?

- Philosophy
- Economics
- Law
- Pedagogy
- Literature
- History
- Natural Science
- Engineering
- Agronomy
- Medicine
- Management
- Art
- Others

What is your mobile phone brand?

Where are you?

- Home
- Shopping mall
- Classroom
- Library
- Dormitory
- Others

Which equipment do you use to participate in the experiment?

- ○ Desktop
- ○ Laptop
- ○ PAD
- ○ Smartphone
- ○ Others

The final earnings of the experiment

The participation fee for this experiment is 10 RMB.

The payment you get from the decision throughout the experiment is:

￥ today.

Therefore, your total payment for the entire experiment, including the participation fee, is:

￥ today.

Thank you for your participation. We will transfer you the payment after the end of the experiment through official account “ancademy”. You can withdraw it (enter "ancademy"-Assistant-Account-YANZHI-Withdrawal) to your WeChat account. If you haven’t followed official account “ancademy”, please follow it as soon as possible. In this way, you can receive the payment and withdraw it in time!