

The Ethics of Student Participation in Economic Experiments: Arguments and Evidence

Robert Hoffmann^{* a} Janneke Blijlevens^a Swee-Hoon Chuah^b
Ananta Neelim^a Joanne Peryman^a Ahmed Skali^a

^a*Behavioural Business Lab, School of Economics, Finance and Marketing, RMIT University, 445 Swanston Street, Melbourne, Victoria 3000, Australia*

^b*Tasmanian School of Business and Economics, University of Tasmania, Hobart, Tasmania 7001, Australia*

8 February 2020

Abstract

Despite strong professional practice norms experimental economists are increasingly subject to sweeping ethics review processes. A central issue in these processes is the recruitment of students as “overresearched” participants. We critically discuss the potential associated ethical risks typically identified in ethics regulations. We then test the efficacy of potential design countermeasures. We find support for some (informed consent procedures, debriefings, non-differential rewards, opt-in) but not others (research outside class time, educational relevance, non-teacher researchers). The paper intends to inform economists’ (1) design choices to reduce ethical risks without sacrificing scientific integrity, and (2) justification of these choices to ethics review boards.

Key words: Human participant research, students, research ethics, informed consent
JEL-codes: A20, C90

^{*} Corresponding author. Email: robert.hoffmann@rmit.edu.au. Tel: +61 3 9925 5447. We thank Mirra Seigerman for research assistance. We are grateful to Adrian Camilleri, Catherine Eckel, Martin Kocher, Andreas Ortmann and Monica Tan for discussions.

1 Introduction

Economics has developed strong experimental design and practice norms within the profession (de Quidt et al., 2019). However the autonomy of experimental economists in designing and conducting their studies is eroding. For better or worse, experimentalists are having to pay increasing attention to the changing perceptions of other stakeholders, chiefly public policymakers, universities and experimental participants. In most places experimental research is now subject to approval from ethics review boards (ERBs) that uphold regulatory standards designed to protect research participants from ethical risks. For experimental economists this presents significant issues. ERBs represent the formal institutions introduced with the intention to supersede previous informal professional norms. Their rules are bulky and apply indiscriminately, i.e. to human participants research beyond both economics and experiments. In particular, policymakers and ERBs are frequently ignorant of experimental economics practice and the scientific reasons behind it. ERBs have also been accused of overreach and pursuit of bureaucratic interest (e.g. Haggerty, 2004; Van den Hoonaard, 2011). As result ERB processes impose significant costs on experimental economists (Page and Page, 2017).

ERBs are routinely concerned with a broad range of potential risks associated with human participant research, many of which typically do not concern experimental economists (see Greiner and Stephanides, 2019). Examples include recruitment of inherently vulnerable participants, invasive methods and deception. However one that does apply is the recruitment of students as participants (e.g. Dalziel, 1996; Clark and McCann, 2005). This is common in all social sciences including economics (de Quidt et al., 2019)¹ where students have traditionally been recruited for the convenience (i.e. relatively low opportunity cost) of both researcher and student participant (Henrich et al., 2010, p. 108) and “allowed a boom in human experimentation” (Exadaktylos et al., 2013, p. 4). This is no longer ethically uncontroversial. For example, Australia’s *National Statement on Ethical Conduct in Human Research* (2018) states that students “are vulnerable to being overresearched because of the relative ease of access to them as research populations.” In the U.S., proposed changes to the federal ‘Common Rule’ governing human participant research include a new provision to review proposed studies which may “have a detrimental effect on student achievement” (U.S. Government, 2015).

The increasing perception of the vulnerability of student participants by ERBs challenges experimental economists for whom the ability to recruit them affects research output and career progression. On the other hand, the potential ethical risks to participants should be taken seriously by all researchers, and especially when a duty of care exists beyond their research projects (i.e. as teachers). An appropriate response to the concern of using student participants in economic research must start with a systematic and critical discussion of the potential issues that concern policymakers, their validity in theory and practice, and the potential measures that can be put in place to mitigate them. On a pragmatic level, ERB processes increasingly pose challenges to project timelines and scientific integrity that

¹ These authors found that around 80% of experiments published in top economics journals between 2012 and 2017 were laboratory studies, of which “most [...] draw participants from the overall student body” (p. 390).

need to be managed. A systematic understanding of ethical issues associated with the use of student participants is therefore important in practice as well as in principle.

This paper seeks to contribute to the discussion of ethical issues in student participation in economic research. We examine the ethical risks that policy has identified in student experimentation and how can they be mitigated. We address three specific questions: (1) What are these potential ethical issues of student participation (experimentation) in theory? (2) What are the potential countermeasures that economists can (and often do) deploy in the research design and protocols to mitigate these issues? (3) Are these successful in reducing risks of harm in the eyes of potential student participants?² The overall purpose of this paper is to inform economists dealing with ERBs through a critical discussion of the ethical issues associated with using student participants along with empirical evidence regarding efficacious countermeasures.

In response to (1), we identify and examine five key types of ethical issue in student participation in economic research (section 2). Our approach is qualitative: We consulted a range of national, international as well as university ethical guidelines regarding human participant research that raise concerns in varying detail.³ We also used our collective experimental research and institutional experience to furnish additional issues that are expressed mainly between the lines of official documents and processes. In response to (2), we then identify various countermeasures economists may (and often do) take to alleviate these ethical issues. Finally, in response to (3), we present experimental evidence of the efficacy of the countermeasures in reducing risks as judged by students. Gauging the perceptions of actual and potential participants is a valid approach because of the subjective nature of ethics concerns (e.g. Milgram, 1964). The empirical study was conducted online with U.S. university students who were asked to assess a standard lab experiment presented in different versions in order to test the efficacy of our countermeasures. The design and results are presented in section 3. We discuss our findings in section 4 and offer concluding remarks in section 5.

2 Ethical Issues in Student Participation in Research

What are the potential risks of negative effects of economic research conducted with student participants? In this section we outline five potential issues identified through our qualitative analysis of ERB guidelines (see also Greiner and Stephanides, 2019). For each we state the case for an ethical concern before discussing it critically and finally outlining potential design countermeasures. Table 1 provides an overview of these two dimensions of our analysis.

² Our focus throughout is not on the important *scientific* issues surrounding student participants such as external validity that are discussed in detail elsewhere (e.g. Harrison and List, 2004; Levitt and List, 2007; Henrich et al., 2010). Scientific issues are mentioned only where they relate to ethics concerns.

³ While prevailing national guidelines govern practice in American (National Commission 1978) and Australian experimental economics centres, no central guideline seems to govern European ones. We inferred their practice from their individual institutional websites.

Countermeasures	Ethical Risks				
	2.1 Dependent Relationships	2.2 Learning Outcomes	2.3 Customer Relations	2.4 Educational Equity	2.5 Negative Effects
Debriefing		✓	✓	✓	✓
Informed Consent		✓	✓	✓	✓
Educational Relevance		✓	✓		
Opt-In	✓		✓		
Non-Teacher Researchers	✓				
Recruitment Database			✓		
Private Payment				✓	✓
Non-Class Time		✓	✓		
Non-Differential Rewards				✓	✓

Table 1

The nine countermeasures by the ethical risks they address. Matches denoted by ticks.

2.1 *Dependent Relationships*

The Issue A fundamental ethical principle of human participant research generally is consent, where participants have freely volunteered to participate (e.g. Commonwealth of Australia, 2018). When students are recruited as participants, genuine consent may be muddled by their power dependency relationships with researcher-teachers inviting them to participate. A power relationship between academic and student may create perceived pressure to participate. For instance a student who is invited by their teacher to participate in a research project may consent only to avert informal sanctions (in terms of poor grades or references) that they fear may result. According to the Australian National Statement,

pre-existing relationships between participants and researchers [...] may compromise the voluntary character of participants' decisions, as they typically involve unequal status, where one party has or has had a position of influence or authority over the other.

Similarly, according to an international list of guidelines on human participant research (U.S. Department of Health and Human Services, 2010), students are identified as a vulnerable group because they may feel coerced to participate in research by their teachers. In psychology, where participation is often expected as part of course credit, the resulting potential for dependent relationship issues has been recognised. The American Psychological Association's (APA) *Ethical Principles of Psychologists and Code of Conduct* (2010) require researchers to take steps "to protect the prospective participants from adverse consequences of declining or withdrawing from participation" including "the choice of equitable alternative activities" where the research is part of a course. Other authors have gone further and questioned the ethical merit of for-credit participation altogether (e.g. Baumrind, 1964, p. 421).

Discussion Dependent relationships may seem more harmless from the perspective of teacher-researchers who after all are the more powerful party. However, they need to be taken seriously not least due to the subjective nature of the concerns involved. When researchers also teach classes, especially large undergraduate ones, power dependency relationships with some research participants are statistically likely. Experimental economists

indirectly acknowledge that power relationships can affect a participant in their recognition of demand effects that may result when a desire to accommodate the experimenter motivates particular responses in the experiment (e.g. Zizzo, 2008; de Quidt et al., 2019). Where dependent relationships exist prior to the research, the decision to participate may be similarly influenced.

On the other hand it seems that the magnitude of the resulting ethical issue depends much on a range of specific circumstances. Chief among these is whether the research is indeed part of a particular course or instead is organised by the teacher purely for her own research independent of any course (Diamond and Reidpath, 1992). Moreover, if the class is large then students may perceive a greater degree of anonymity. The perception of detrimental non-participation may also decrease in the age and experience of the students involved. In addition to the ethical considerations, reducing the perception of dependency may be prudent scientifically to the extent that they give rise to demand effects.

Countermeasures A number of potential protocol measures can mitigate dependent relationship effects. One is disconnecting research participation from course teaching or assessment. Further, seeking the views of student representatives on student involvement in research can inform ethical recruitment and participation of students in research (e.g. Council for International Organizations of Medical Sciences, 2002). Another is to ensure that the interaction between researcher-teachers and participants is double blind. Teacher-researchers may use research assistants to recruit for and conduct sessions to break mutual identification. Finally, to mitigate students' perception of risk from choosing not to participate in a teacher's research, participation can be based on opt-in rather than opt-out (e.g. Commonwealth of Australia, 2018), which is a requirement in some national codes. It means that a person is by default not a participant but needs to actively volunteer, for example by responding to the invitation. In our study we focus on the latter two countermeasures, i.e. the extent to which using non-teacher researchers and opt-in participation can alleviate concerns from dependent relationships.

2.2 Learning Outcomes

The Issue A second concern with student participation in research is potential harm to their learning outcomes. When research is conducted during scheduled classroom time, it "might draw sufficient time and attention away from the delivery of the regular educational curriculum [to] have a detrimental effect on student achievement" (U.S. Government, 2015, p. 27). This concern mostly relates to research with new teaching methods designed to improve pedagogy. Such research often require repeated unsuccessful trials and can negatively impact learning outcomes for the students involved.

Discussion In-class experiments are more common among psychologists including ones performed for course credit. In economics, less than 5% of experiments are conducted in class (de Quidt et al., 2019). The reasons can also be educational, when future generations of researchers are being trained. The concern over learning outcomes therefore depends much on the purpose of the experiment. In general, economic experimentation is not conducted to identify effective classroom innovations. Further, when economists do conduct

classroom (rather than research) experiments the typical objective is to improve the learning experience of current (rather than future) students (e.g. Bergstrom and Miller, 1998; Holt, 2007). There is evidence in favour of learning opportunities that these kinds of experiments afford. Course-relevant research can be used as meaningful in-class activities (Frank, 1997; Ball et al., 2006) supporting “experiential learning” (Castilla, 2014) and improving test scores (Dickie, 2006). Researchers can enhance the educational relevance of their studies as teachers through the principle of research-led teaching. Conversely they can find inspiration for their research from teaching, especially when there are student cohorts with significant work experience, such as MBAs, who highlight policy and practice issues that experiments can usefully address.

However, it is clearly neither feasible nor desirable to always and fully restrict research to areas of educational relevance to students. Research may be exploratory or ahead of received knowledge. Moreover, syllabi are increasingly standardised and changes to them are subject to demanding processes. At any rate educational relevance is hard to measure accurately especially to the extent that general knowledge or intellectual development are integral parts of a university student’s education. Pure research conducted outside the context of a particular course can also benefit student learning. Under the right circumstances participation promotes more general intellectual development by demonstrating the process of research, introducing topical research issues and generally whetting students’ appetites for the pursuit of knowledge. The large-scale U.S. survey by Russell et al. (2007) showed that undergraduates’ research opportunities promote their interest in research and can nudge them towards research degrees. Students may also learn about their own behavioural proclivities (Ifcher and Zarghamee, 2015). This is true particularly where there are post-research debriefings, as required by the American Psychological Association (2010, p. 11) ethical guidelines.

Countermeasures Instead of a single countermeasure, learning outcomes as a general objective can be addressed by several design and protocol features. Targeting students from research-relevant study programmes is one way. In our own study in this paper we consider the effects of four countermeasures when there is a risk to learning outcomes from participating in economic research. The first is to conduct research exclusively outside scheduled classes as separate activities. A second measure is to create educational relevance, i.e. a positive link between research and student learning outcomes. Economic research is almost always relevant to what *some* students are supposed to learn. A link arises if such students were targeted or self select for participation. Our third countermeasure is informed consent. Self-selection is based on informing potential participants (in general terms) about the objectives and issues of the research so they may assess a link to their learning outcomes. A fourth countermeasure is to conduct debriefings (Baumrind, 1964) that inform participants about the issues behind the research and likely findings. Where applicable, researchers might highlight links to syllabus content and learning outcomes. That said, debriefings harbour the potential for revealing research objectives to participants in future sessions in the case of experiments. In the case of experiments an alternative to verbal debriefings at the end of each session would be to deliver debriefings after the conclusion of the final session via social and electronic media.

2.3 *Customer Relations*

The Issue A third concern, especially relevant for institutions with significant or exclusively private funding, is customer relations management. Our own experience at different institutions suggests that universities are concerned with avoiding annoyance when fee-paying students are approached excessively or lose class time to research projects of little educational value to them. A related symptom of this concern is the reluctance of many institutions to grant their own researchers unfettered access to student email lists. The stated rationale is often anecdotal evidence that students complain about excessive emails from the university. This mirrors a recent trend against over-surveying in customer relations marketing (e.g. Grimes, 2012; Glazer, 2015). This kind of concern is associated with a much-debated view of college students as customers that is resisted by many academics (e.g. Eagle and Brennan, 2007).

Discussion The counterarguments to this view come from the ways in which university students differ from customers in purely commercial relationships. Students may see themselves as members of a scholarly community and have pro-social motives to contribute (Krause, 2005). Students may recognise costs and benefits to community membership, which, in cases such as research participation, have a public good character as they generate non-excludable benefits such as knowledge creation and university reputation. Participation or at least receiving solicitations to participate should therefore be an expected part of being a student.

There are also private benefits from participation. Even as customers, students stand to benefit from the additional opportunity for learning (as mentioned) but also for financial rewards to contribute to the rising costs of higher education. If researchers calibrate earnings to match students' typical opportunity cost (e.g. Friedman and Sunder, 1994), then participation can provide a relatively attractive and convenient alternative to earning income.

Countermeasures To the extent that students see themselves as consumers of education, the previous four countermeasures to learning outcome risks may reduce customer relations concerns also. In addition, in our study we examine to what extent the use of recruitment databases such as ORSEE (Greiner, 2015) can alleviate annoyance from unsolicited email invitations by sending participation invitations only to subscribed members. Furthermore, opt-in can avoid the impression that education providers are unilaterally redefining the transaction with their student customers to include research participation by default.

2.4 *Educational Equity*

The Issue The presence of incentive compatibility in economic experiments means student participants often receive differential rewards for the same commitment in terms of time and effort. ERBs are often unaccustomed to performance-related payments for student participants, to the reasons behind these and tend to query this feature. The reason is that educational equity or fairness is a fundamental principle of education (Field et al., 2007). Differential rewards, when common knowledge, may be interpreted by ERBs as jarring with student participants' or institutions' perceptions of fair treatment within

the educational environment.

Discussion ERBs are often poorly informed about what incentive compatibility means for participants in practice and why economists insist on it. Often clarification can help. For example, one may need to highlight the fact that differential rewards are neither random nor completely out of the control of participants. While rewards may be subject to chance or the decisions of other participants they reflect the decisions participants make under these conditions. At other times rewards depend on effort levels that are selected by participants with commensurate financial consequences (for an overview see Gill and Prowse, 2011). Despite continuing methodological debate, incentive compatibility remains one of the two cornerstones of experimental economics that differentiates the field from experimental psychology (e.g. Hertwig and Ortmann, 2001; Croson, 2005). The importance economists attach to it is demonstrated by the associated costs. Incentive compatibility not only raises ERB concerns but entails logistical effort of determining and delivering differential rather than flat rewards. If performance-related earnings from interactive tasks are to be delivered instantly, co-players must make decisions at the same time (Arechar et al., 2018). Using the strategy method of eliciting responses to all potential circumstances eliminates this need but can create bias (Chuah et al., 2013).

There are a number of counterarguments to the resulting concerns over fairness. The first relates to the meaning of fairness. Deutsch (1975) differentiates between two meanings, fairness as equality (same consequences irrespective of performance) and as equity (same translation of performance into consequences). While incentive compatibility may violate equality as same outcomes, it is consistent with the principle of equity. In education the idea of possibly unequal outcomes is accepted when academic performance differs ('fair isn't equal') and therefore the focus is on ensuring equal process (Welch, 2000). Indeed one of the functions of the education system is to provide labour market signals that differentiate students on the basis of performance. A related second point is that performance-related pay more generally reflects real-life interactions that university is meant to prepare students for.

Countermeasures Notwithstanding whether economists recognise the concern regarding the fairness of compatible incentives, protocol countermeasures can help alleviate it. One is informed consent, i.e. that participants have not only freely volunteered but done so based on their unimpaired understanding of all information relevant to the associated potential risks and benefits of participation (e.g. Commonwealth of Australia, 2018). Informed consent is part of the APA code and a legal requirement in many jurisdictions such as Australia. For example, an informed consent process that advertises performance-related pay in an experiment would allow participation decisions to be informed accordingly. Moreover, separating experimental participation or performance from course credit or even grades consigns any perceived unfairness to a more peripheral aspect of the educational experience. In addition, it has been shown in general educational settings that explaining the rationale behind differential rewards can address fairness concerns of students (e.g. Welch, 2000). A similar approach may work to make incentive compatibility more palatable among those student participants who harbour reservations. This can be done by experimental debriefings, for example. In addition, equality can be increased through a show-up fee that

ensures everyone leaves with something.

We examine four of these countermeasures in the present study. The first two are informed consent and debriefings. We also examine non-differential or flat payment schemes that eliminate perceived unfairness from incentive compatibility. This allows us to gauge to what extent differential payments are perceived as ethically problematic by students. Fourth, we examine whether ensuring privacy of earnings and their delivery can mitigate risks by eliminating the possibility of comparison among participants.

2.5 *Negative Effects*

The Issue A final ethical concern is that the activities participants perform or are subjected to in an experiment can cause harm to themselves and/or other participants. Even in studies without physical intervention, participation may have adverse psychological effects. Controversial examples from the history of psychology are Milgram's obedience study (Baumrind, 1964; Milgram, 1964; Ifcher and Zarghamee, 2015), Wendell Johnson's stuttering experiment (Reynolds, 2003) and the Stanford prison experiment (Zimbardo, 1973).

While standard economic experiments involve no physical interventions nor deception there is still scope for negative effects. For example, a participant's own decisions in social dilemma situations or experiments with the 'dark side' of human motivations (e.g. Abbink and Herrmann, 2011) may result in an unexpected threat to self image. Some participants may also be disappointed with their earnings in absolute terms or relative to others. In hindsight participants may worry that incentive compatible risk tasks they were asked to perform contravene religious proscriptions of gambling, or take risks they otherwise would not accept (Greiner and Stephanides, 2019). Some participants may resent the decisions their co-participants make in strategic situations or fear recrimination based on their own strategic decisions.

Discussion Participants may rightfully expect that researchers are concerned with their welfare and avoid subjecting them to potential harm (Baumrind, 1964; American Psychological Association, 2010). However, there are cases where some degree of perceived harm may need to be accepted. One reason is a significant element of subjectivity in the perception of harm, as with the previous issues of dependent relationships and fairness. In practice it is impossible to conduct research to satisfy the most acute harm perceptions.

More fundamentally, there is often a trade-off between scientific and ethical objectives. The reason is a limit to how much participants can be told about an experiment without biasing responses through divulging the exact research question (List, 2008; Ifcher and Zarghamee, 2015). For example, Milgram's experiment would not have been successful had he exercised complete candour. In such cases individual adverse effects need to be weighed against the scientific benefit of the experiment (Baumrind, 1964, 1971; American Psychological Association, 2010). The research may then be justified on the grounds that the scientific value outweighs inevitable negative effects. For many, the classic studies by Milgram and Zimbardo fall into this category. On the other hand this trade-off applies to psychology more than to experimental economics, where strict non-deception of participants is practiced

(see Bonetti, 1998). Under non-deception, researchers cannot protect the scientific integrity of an experiment by keeping relevant details from participants beforehand.

Countermeasures Economists can adopt countermeasures to minimise the potential for harm from activities participants perform in experiments. As previously mentioned, appropriate informed consent procedures allow participants to avoid experiments they deem personally risky. As another measure, debriefings not only outlining the research issues but also describing the typical responses of participants can help alleviate self image threats (Milgram, 1964). Participants may find solace in learning their behaviour in an experiment corresponds to what many others do. Another measure is ensuring the anonymity and privacy of decisions and earnings. Others' mere awareness of one's choices can entail a psychic cost when disapproval is feared (Chuah et al., 2016). For example, if a participant keeps the entire stake in a dictator game, then the public payment delivery can cause discomfort from others' *gaze*, to use Sartre's (1956) term. This cost is clearly greater when a lack of anonymity allows sanctions by affected others after the experiment. Finally, to the extent that perceived unfairness of unequal rewards causes negative effects, a non-differential payment scheme can alleviate such concerns. These four previously-mentioned countermeasures (debriefing, informed consent, non-differential rewards and private payment) are tested in the present study.

3 Study

We have seen that economic experiments with students harbour certain ethical risks that could in principle be addressed with the different countermeasures we identified (table 1). In this section we report an empirical study designed to examine the effectiveness of these measures in addressing ethical risks by reducing student participants' perceptions of harm from participation in economic experiments. Our approach was to examine the extent to which different versions of a hypothetical study conform with or violate relevant ethical norms of the population affected, i.e. potential student participants. We use the incentive-compatible norm elicitation method proposed by Krupka and Weber (2013). These authors' mechanism measures 'social appropriateness' of certain actions using a pure coordination game. In particular, participants make judgements along a four-point scale of appropriateness without a neutral response and are rewarded to match the response of others. Responses elicited in this way reflect what is "collectively perceived as socially appropriate or inappropriate" (Krupka and Weber, 2013, p. 502).⁴ We use the same method to elicit participants' collective perceptions of the appropriateness (in terms of risk of harm) of a hypothetical research study based on several modular features that we vary systematically. We examine the effect of each of these features on the perception of harm from participation.

⁴ For empirical examinations of this mechanism see Erkut et al. (2015), whose study supports its validity, and Vesely (2015), who finds its incentive element does not affect responses.

3.1 Design and Implementation

The study was conducted via online questionnaire⁵ in December 2018 with 294 current university students residing in the United States, recruited and paid via Amazon Mechanical Turk (AMT). Their ages ranged between 18 and 61 (mean: 22.2) and 48% identified as female. 85% percent were U.S. citizens. Participants completed the study in approximately seven minutes. A total of 124 participants (42%) successfully matched the responses of others and were rewarded with a bonus payment. To avoid any potential university specific selection bias, AMT student participants from different universities were sampled.

The hypothetical study consisted of a fixed set of features that remained constant and was shown to every participant (see top panel, table 2). In addition there were nine individual features, each of which had a negative (greater risk of harm) and a positive (less risk of harm) version (see table 2). These features and their versions were chosen to elicit collective perceptions regarding the theoretical ethical risks discussed previously (in the guise of the negative versions) and the efficacy of the potential countermeasures (as the positive versions) we identified in response. In particular we presented participants with general instructions adapted from Krupka and Weber (2013):⁶

Your task in this survey is to assess the effects that participation in a hypothetical paid research study at your university or college would have on participants. We will provide you with different features of this paid research study and ask you to rate the potential for harm, discomfort, inconvenience or other negative consequences.

At the end of this survey, we will randomly choose one of the additional features. For this chosen feature, we will determine which rating was selected by the most respondents in this survey, i.e. the most popular rating out of the four on the scale. If you gave the same rating as the most popular rating, then you will receive a bonus of \$1. This bonus will be paid to you in addition to your participation fee.

After the presentation of the fixed feature set by itself participants were shown, on subsequent separate screens and in random order, the fixed feature set plus either a positive or a negative version of each of the nine additional features. Subsequently, they were asked about their perception of harm from the standard set plus the additional feature described. For each of these nine we asked:

Please rate a study as described above with this additional feature in terms of its potential for harm, discomfort, inconvenience or other negative consequences: (Remember: if this feature is chosen, you will receive the bonus if your rating is the same as the most popular rating).

For brevity we refer to these aspects collectively as 'harm' in the following. Participants responded along a four-point rating scale of *high potential* (coded as 1), *some potential* (2), *not much potential* (3) *no potential* (4).

The two versions of each of the features constituted our treatments. We implemented a between-participant design to test the effects of these features on participants' assessment of risk of harm. Assignment of participants to these versions was balanced in that each

⁵ For a discussion of online experiments see Arechar et al. (2018).

⁶ The full experimental materials are available from the online supplementary documents or from the corresponding author.

Fixed Features	
Imagine you have been invited to participate in a paid study to be conducted at your university/college by researchers from that institution. You are told that participation in the study is completely voluntary, i.e. it is entirely up to you whether you participate or not without any negative consequences if you do not. If you were to take part this study you would be asked to make a series of decisions in private, either by filling in a paper form or an electronic form using a computer. You may also be asked to answer standard demographic questions about yourself (such as your gender, age etc) and/or your attitudes, opinions and values. All responses you make will be completely anonymous, i.e. your data will be identified only by a random code and can never be traced back to your person.	
Individual Features	
<i>Positive Version</i>	<i>Negative Version</i>
Debriefing (DB)	
After the study is completed you will be given information about what the purpose of the study was and what the researchers found out.	After the study is completed you will not be given information about what the purpose of the study was and what the researchers found out.
Informed Consent (IC)	
Before the study you will be given a document which explains everything about the study that is relevant to your decision to participate. If you do decide to participate you will be asked to sign a document that indicates your consent to take part.	Before the study you will be told very little about what participation entails and what you will be asked to do. There will be no form for you to sign to indicate your consent to take part.
Educational Relevance (ER)	
By taking part in the study you can learn things that are relevant to what you are learning as part of your academic degree or program.	By taking part in the study you cannot learn things that are relevant to what you are learning as part of your academic degree or program.
Opt-In (OA)	
In order to participate in the study you have to actively opt in. This means that the default is that you do not take part and you have to indicate if you do want to take part.	In order not to participate in the study you have to actively opt out. This means that the default is that you take part and you have to indicate if you do not want to take part.
Non-Teacher Researcher (TR)	
The researcher is not someone who teaches you, grades your work, will write you an academic reference or someone you depend on as an academic in other ways.	The researcher is someone who teaches you, grades your work, will write you an academic reference or someone you depend on as an academic in other ways.
Recruitment Database (RD)	
The invitation to participate was emailed to you from a mailing list that you actively signed up to. This mailing list is maintained by researchers at your university/college and advertises such paid scientific studies.	The invitation to participate was an unsolicited email from the researchers that was addressed to you and other students at your university/college.
Private Payment (PP)	
The money you receive will be paid out to you in private so that other participants do not see how much you received. For example, you are given the money in another room away from other participants after the study is over.	The money you receive will be paid out to you in public so that other participants may see how much you received. For example, after the study is over, you are given the money in the same room where other participants are present.
Non-Class Time (CT)	
The study will not take place during class time, i.e. instead of or during a scheduled lecture, seminar, tutorial or other classroom activity with a teacher.	The study will take place during class time, i.e. instead of or during a scheduled lecture, seminar, tutorial or other classroom activity with a teacher.
Non-Differential Rewards (DR)	
Each participant will be paid money to participate in the study and each will receive the same amount.	Each participant will be paid money to participate in the study. The exact amount may differ from participant to participant. It depends in part on the particular decisions a participant makes and also decisions that other participants make. This is the case if participants make decisions that impact on other participants. The way decisions lead to how much money is earned will be explained in detail before decisions are made.

Table 2

Features of the hypothetical experiment. Each participant was shown (1) the set of fixed features and (2) either the positive or the negative version for each of the individual features.

version of a feature was administered to half of the participants and that each participant received an even mix of negative and positive features on average (mean: 4.5). However we allowed the number of positive features to differ between participants (min=1, max=8) in order to assess their joint effect on overall perceptions: Participants who are shown more positive features (that were designed as countermeasures) should rate the hypothetical study overall as less harmful.

To measure this overall assessment participants were shown, after the presentation of the nine feature versions, all previously-shown features together along with the fixed feature set and asked: “Overall, how strongly do you agree or disagree that researchers should be allowed to conduct a study with all the features described above? (Remember: if this question is chosen, you will receive the bonus if your answer is the same as the most popular answer).” As in Krupka and Weber (2013), responses were recorded along a four-point rating scale of strongly disagree (coded as 1), somewhat disagree (2), somewhat agree (3) and strongly agree (4).

In this design each participant recorded ten measures of ethical perception towards the hypothetical study: one for each of the individual nine feature versions, and one overall, elicited after all features were individually presented. These constitute our dependent variables of individual feature harm assessment (denoted IA) and overall agreement (OA) respectively.

3.2 Participants’ Assessment of Individual Features (IA)

The assessment of the individual positive features (IA) averaged across all participants and all features was 3.24. For negative features, it was 2.57. This suggests that overall our countermeasures assuaged participants’ perception of harm. To test the efficacy of each of the nine individual features of the hypothetical study, we conduct univariate analyses of differences across participants who saw the positive and negative version of each of the features. The bars in the top panel of figure 1 show the IA for every feature version as the average across all participants who saw that version. T-tests show that positive versions, compared to the corresponding negatives ones, significantly reduced perception of harm for every individual feature. Cohen’s d -values show that the effect sizes are large for informed consent, debriefing and non-differential rewards and medium for the remaining six.

We then examine the same issue using regressions to check whether results from our univariate analyses are robust to individual heterogeneity in terms of age, gender and whether a participant was born in the U.S.A. or abroad. Equation 1 outlines the estimation specification for each of the regressions.

$$IA = \beta_0 + \beta_i Positive + \gamma_i \mathbf{Z}_i + \varepsilon \quad (1)$$

The dependent variable was participants’ perception of harm (IA) and the independent variables were, for each feature, whether a participant was shown the positive (coded as 1) or the negative version (coded as 0). The set of additional controls (\mathbf{Z}_i) include individual characteristics of age, gender and U.S. birth.⁷

⁷ We report Ordinary Least Squares (OLS) rather than Ordered Probit regressions here and in

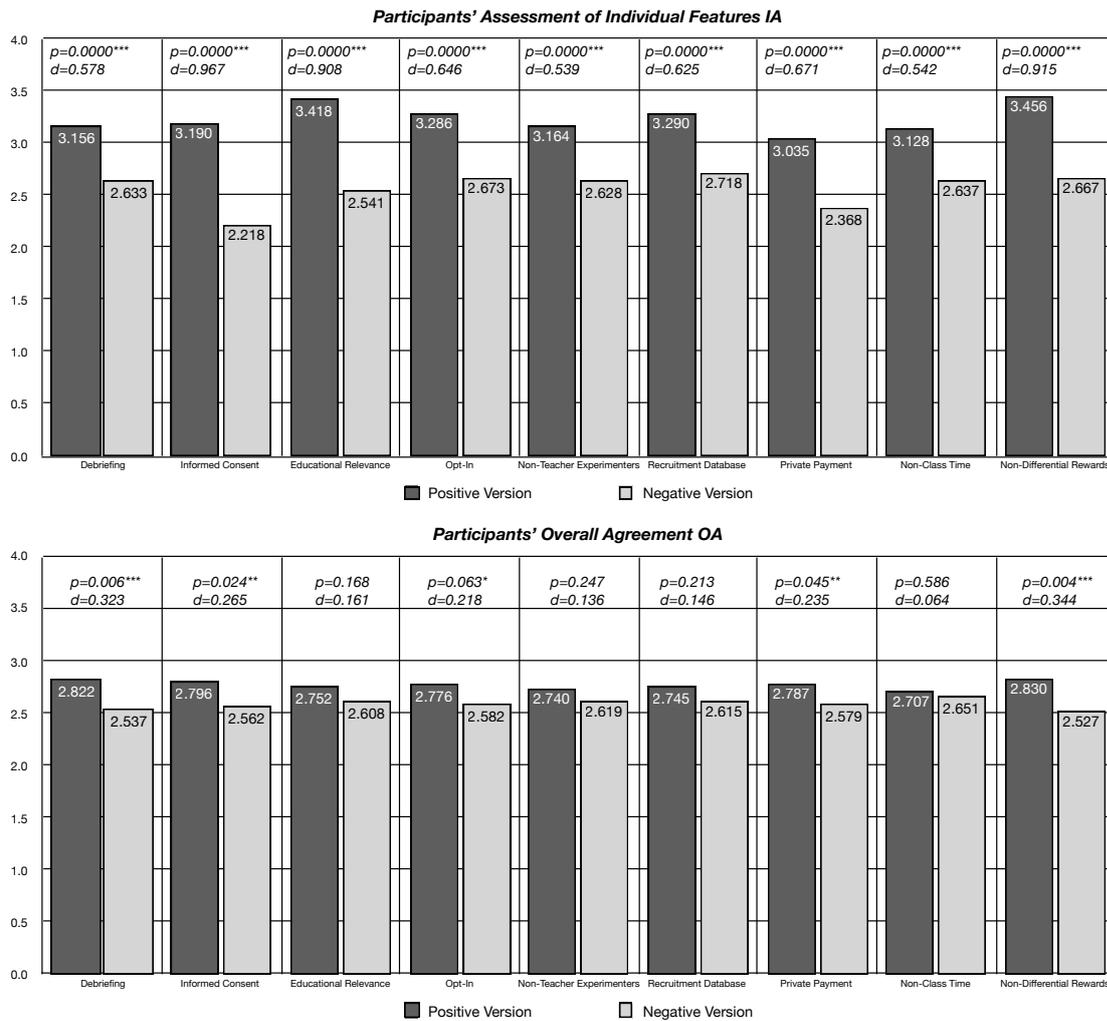


Fig. 1. Participants' assessment of hypothetical study by individual features and their versions. Significance of difference in means and effect size shown as p -values and Cohen's d . * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Results are reported in table 3. Each model (presented in separate columns) estimates participants' IA for each of the different individual features. The coefficients for the positive version dummies (β_i) are always positive and significant. Thus we conclude that all countermeasures have the desired effect in improving the participants' harm perception.

Result 1 *The individual countermeasures for every feature of the hypothetical study significantly lower participants' perception of participation risk from that feature.*

the following because some of our multivariate analyses involve combining coefficient estimates from regressions to find differences in treatment effects across gender. Combining coefficient estimates obtained from probit regressions can lead to biased estimates (Ai and Norton, 2003). For the sake of consistency, we use OLS regressions throughout. Also we are interested in relative effect sizes that are visible in standardised (beta) coefficients that Probit does not generate. Our results, in terms of the (in)significance of individual variables, are robust with respect to either estimation technique.

	DB	IC	ER	OI	TE	RD	PP	CT	DR
Age	-0.002 (0.00933)	-0.003 (0.00985)	-0.104* (0.0101)	-0.057 (0.0105)	-0.058 (0.00895)	-0.082 (0.00864)	-0.040 (0.0108)	-0.076 (0.00900)	-0.091* (0.00742)
U.S. Birth	0.024 (0.282)	0.022 (0.279)	0.016 (0.250)	0.035 (0.255)	-0.039 (0.285)	-0.005 (0.186)	0.016 (0.268)	0.015 (0.243)	0.037 (0.245)
Female	-0.005 (0.106)	0.004 (0.118)	-0.042 (0.113)	-0.113** (0.110)	-0.079 (0.117)	-0.077 (0.107)	0.048 (0.116)	-0.073 (0.106)	-0.005 (0.101)
Positive Version	0.278*** (0.106)	0.435*** (0.118)	0.415*** (0.112)	0.308*** (0.110)	0.260*** (0.116)	0.292*** (0.107)	0.317*** (0.117)	0.264*** (0.106)	0.416*** (0.101)
<i>N</i>	294	294	294	294	294	294	294	294	294
<i>R</i> ²	0.078	0.191	0.186	0.113	0.079	0.103	0.106	0.081	0.184

Table 3

Ordinary Least Squares regression results for participants' individual assessment of the respective features of the hypothetical study. Features denoted by their acronyms. Standardised beta coefficients with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.3 Participants' Overall Agreement (*OA*)

We also examined whether the countermeasures raised *OA*, the overall assessment participants made of harm, based on all the features they were presented with. *OA* averaged across all feature versions and participants was 2.68 (2.66 for males, 2.70 for females, not significantly different). If the countermeasures are successful then participants who saw more should have better perceptions of the hypothetical study. The scatterplot in figure 2 shows the relationship between participants' *OA* ratings (*Y*-axis) and the number of positive features they saw (*X*-axis). Indeed participants' *OA* is significantly and positively correlated with the number of positive features participants saw (Pearson $r = 0.302$, $p = 0.0000$).

Result 2 *Together, the countermeasures significantly lower participants' overall perception of participation risks of the hypothetical study.*

The bottom panel in figure 1 shows *OA* by feature versions, i.e. averaged for those participants who were shown a particular version. The t-tests show that participants' *OA* was significantly (at the 5% level of significance or better) raised by being shown the positive rather than the negative version of the following features: debriefing, informed consent, opt-in (marginally), private payment and non-differential rewards. Positive versions of educational relevance, non-teacher experimenters, non-class time and recruitment database did not improve participants' overall perception of harm.

However, these univariate tests do not control for the number and type of positive features each participant saw. To account for this, we analysed *OA* in a multivariate framework. All study features were simultaneously entered as independent dummy variables for being shown the positive version coded as 1. The model specifications are presented below. Equation 2 does not control for individuals characteristics whereas equation 3 does.

$$OA = \beta_0 + \beta_i \mathbf{X}_i + \varepsilon \quad (2)$$

$$OA = \beta_0 + \beta_i \mathbf{X}_i + \gamma_i \mathbf{Z}_i + \varepsilon \quad (3)$$

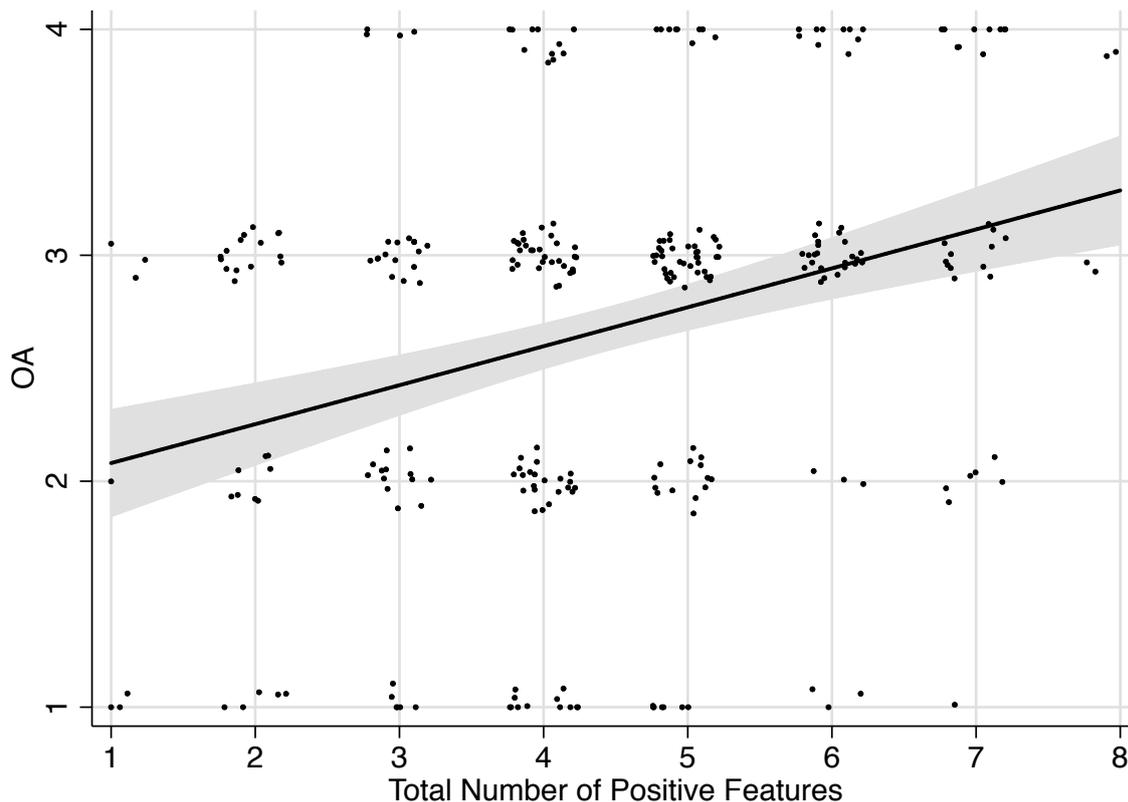


Fig. 2. Number of positive features presented to participants by their overall agreement with the study (OA). Dots around each integer coordinate represent number of observations for that coordinate (e.g. three participants with six positive features had an overall assessment of 'slightly disagree'). Fitted line shown in bold. The grey shaded area represents the 95% confidence interval.

Here β_i represents the marginal effect of each positive feature on OA controlling for the other countermeasures the participants may have been exposed to. Results are presented in Table 4. The first two columns estimate equations 2 and 3 for all participants. Unlike the case for IA , we find that only some of the countermeasures have a significant impact on overall perception of harm. Providing debriefs, an informed consent procedure and non-differential rewards significantly enhance perceptions of the hypothetical study. Opt-in marginally (i.e. at the 10% level) increases OA . The other features do not affect overall perceptions of harm. With the exception of the latter results, the significance of the independent variables is robust to the inclusion of the controls (equation 3). Additionally the dummy variable for U.S.-born participants is significant, marginally at the 10%-level.

Result 3 *Controlling for demographics, some countermeasures (debriefing, informed consent, opt-in and non-differential rewards) significantly enhance participants' overall agreement with conducting the hypothetical study.*

3.4 Gender Effects

In this section we analyse gender differences in participants' perception of ethical norms due to our treatments. In previous experimental work, female participants have been found to be more risk averse (Eckel and Grossman, 2008), less trusting (Buchan et al., 2008) and

DV: <i>OA</i>	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	Males	Males	Females	Females
DB	0.154*** (0.101)	0.152*** (0.102)	0.073 (0.146)	0.068 (0.149)	0.234*** (0.145)	0.237*** (0.143)
IC	0.153*** (0.101)	0.147** (0.103)	0.143* (0.142)	0.140* (0.143)	0.165** (0.148)	0.156* (0.149)
ER	0.081 (0.103)	0.081 (0.103)	0.065 (0.138)	0.062 (0.137)	0.089 (0.162)	0.090 (0.162)
OI	0.100* (0.101)	0.092 (0.102)	0.077 (0.143)	0.063 (0.151)	0.079 (0.148)	0.082 (0.146)
TE	0.050 (0.104)	0.043 (0.106)	0.110 (0.139)	0.108 (0.143)	0.023 (0.165)	0.009 (0.165)
RD	0.066 (0.104)	0.066 (0.105)	-0.028 (0.148)	-0.030 (0.150)	0.204** (0.152)	0.208** (0.152)
PP	0.093 (0.102)	0.088 (0.101)	-0.006 (0.141)	-0.008 (0.141)	0.210*** (0.143)	0.202*** (0.141)
CT	0.022 (0.103)	0.021 (0.103)	-0.040 (0.143)	-0.038 (0.145)	0.111 (0.149)	0.103 (0.148)
DR	0.159*** (0.102)	0.162*** (0.102)	0.239*** (0.146)	0.248*** (0.149)	0.052 (0.143)	0.041 (0.139)
Female		0.010 (0.101)				
Age		-0.017 (0.00817)		-0.019 (0.0101)		-0.025 (0.0140)
U.S. Birth		0.088* (0.194)		0.060 (0.275)		0.142* (0.278)
<i>N</i>	293.000	293.000	154.000	154.000	139.000	139.000
<i>R</i> ²	0.109	0.117	0.096	0.100	0.208	0.229

Table 4

Ordinary Least Squares regression results for participants' overall agreement with the hypothetical study. Standardised beta coefficients with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

less willing to compete (Gneezy et al., 2003) than male participants. These preferences can potentially shape gender-specific attitudes towards participating in research, particularly experiments (e.g. Cleave et al., 2013): Participation in experiments entails risks, elements of competition with co-participants and some degree of trust in the researchers. Additionally, attitudes towards individual features, like the ones relating to learning outcomes, may differ to the extent that female university students tend to be educationally more motivated (e.g. McNabb et al., 2002). These individual views may bias the ethical norm perceptions of our participants in the direction of their (gendered) individual views through social projection or false consensus effect (Wenzel, 2005). As a result the norm perceptions of males and females in the experiment may differ.

Models 3 to 6 in table 4 re-estimate equations 2 and 3 for male and female participants respectively. As previously we first estimate models for the nine positive versions before

Differences estimates for	DB	IC	ER	OI	TR	RD	PP	CT	DR
<i>IA</i>									
(Positive version - Negative version) Male	0.343***	0.498***	0.405***	0.242***	0.250***	0.347***	0.314***	0.338***	0.490***
(Positive version - Negative version) Female	0.390**	0.814***	0.901***	0.754***	0.556***	0.444***	0.669***	0.342**	0.632***
<i>OA</i>									
(Positive version - Negative version) Male	0.110	0.237*	0.103	0.092	0.184	-0.053	-0.016	-0.064	0.440***
(Positive version - Negative version) Female	0.433***	0.290*	0.166	0.149	0.021	0.380**	0.376***	0.195	0.080

Table 5

Difference estimates for effects of female gender and positive version on *IA* (top panel) and *OA* (bottom panel) for each feature denoted by acronyms. The *IA* estimates are derived from individual regressions for each characteristic outlined in equation 4. The *OA* estimates are derived from the regression outlined in equation 5. Standardised beta coefficients with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

including demographic controls. They show marked differences in the perceptions of men and women. For males, informed consent is (marginally) significant as are non-differential rewards. For female participants, debriefing and informed consent remain significant, but not opt in or non-differential rewards. Two variables that are insignificant for all participants are significant for women: recruitment database and private payment delivery. Compared to men, a greater number of positive features affect women's perceptions of the ethical appropriateness of the hypothetical study. This is further evidenced by the much greater R^2 value for women that attest to the greater explanatory power of the countermeasures on women's perceptions.

To provide more evidence we ran two further regressions for *IA* and *OA* respectively (equations 4 and 5).

$$IA = \beta_0 + \beta_i \text{Positive} + \delta_1 \text{Female} + \theta_i \text{Female} \times \text{Positive} + \gamma_i \mathbf{Z}_i + \varepsilon \quad (4)$$

$$OA = \beta_0 + \beta_i \mathbf{X}_i + \delta_1 \text{Female} + \theta_i \mathbf{X}_i \times \text{Female} + \gamma_i \mathbf{Z}_i + \varepsilon \quad (5)$$

For the *IA* regression, we are interested in the difference estimates across gender due to our treatment. β_i captures the effect of countermeasure i for male participants. Thus $\beta_i > (<) 0$ implies that countermeasure i has a positive (negative) effect on male participants' perceptions. The effect of a countermeasure i on *IA* for females is given by $\beta_i + \theta_i$. As before, the set of additional controls (\mathbf{Z}_i) include individual characteristics of age and U.S. birth. For the *OA* regression, β_i and $\beta_i + \theta_i$ capture the effect of each countermeasure i for male and female participants respectively. The difference estimates from the *IA* and *OA* regressions are presented in Table 5. Tables 7 and 8 in the appendix show the full estimations.

The results for the *IA* regressions show that all nine countermeasures have a positive impact on perceptions for both male and female participants. Additionally, the magnitude of the positive impact is greater for female participants throughout. Turning to the *OA* estimates, we find only informed consent had a positive impact on the perceptions of both male and female participants. Our countermeasures of educational relevance, opt-in, non-teacher researcher and non-class time had no effect on perceptions of either gender. Finally, non-differential rewards were important only for male participants; debriefing,

recruitment database and private payment delivery were important for female participants.

Result 4 *Individual countermeasures have a greater positive impact on females' overall assessment of the appropriateness of the hypothetical study. Females' overall perception is influenced by different features than males. Experimental countermeasures are therefore gender-specific to an extent.*

4 Discussion

What are the implications of our findings for designing economic experiments? As we have argued, tradeoffs exist between ethical risks and the scientific integrity of research. This means that the design of ethics-conscious research features needs to be informed by the opportunity costs in terms of both economic and scientific costs. Table 6 contains an overview of costs and benefits of the different features. In the table we summarise the results for each of the countermeasures (ordered by importance) in terms of univariate effect size and marginal effects from regressions for *IA* and *OA*. The table also summarises economic and scientific downsides to each countermeasure. We now discuss each of these countermeasures in this light.

Informed Consent The most important feature we identified is informed consent, i.e. that participants decide to volunteer based on relevant information given to them beforehand. In our regression it is significant overall, for females and marginally for males, and is associated with the second-highest and highest effect size in the univariate analysis of *OA* and *IA* respectively. For experimental economists, the downside of this feature is that the information provided could cause demand effects or allow participants to gain task-related experience that makes them unrepresentative. This is an issue especially where participation is motivated by financial gain providing an incentive for such preparation. Knowledge of the exact tasks may allow certain kinds of participants to self select into the study, causing bias. In practice it is possible to devise general informed consent documents that give participants enough information without biasing them.

Debriefing There are good reasons for experimental economists to conduct debriefings, overall the second most effective feature of our study. Debriefing is a highly significant countermeasure in all our analyses and has the largest effect size for overall agreement. What are the downsides? Writing and disseminating suitable materials can be done at modest cost and inconvenience to researchers, especially if electronic media are used. More significant is the danger that debriefs are leaked to future participants and can bias them in two ways (Friedman and Sunder, 1994, p. 13, 30): One is what might be called encumbered participants, i.e. that they gain experience of the tasks that makes them unrepresentative and/or generates demand effects (de Quidt et al., 2019, p. 389). The other is that this knowledge causes demand effects, i.e. responses in line with what participants think experimenters want. Experimental economists would need to be mindful of releasing debriefs only after data collection has been completed.

Non-Differential Rewards We found evidence that incentive compatibility or performance-related rewards contribute to a lower overall perceived appropriateness of research. We saw this result was driven by the responses of male participants. Our data do not reveal

Counter-measures	Effect				Downsides	
	<i>IA</i>	<i>OA</i>			Economic Costs	Scientific Issues
	<i>d</i>	<i>d</i>	Significance	β		
Informed Consent	1.384	0.497	Overall, males (marginally), females	0.153	Creation of consent materials; reading takes participants time	Demand effects, encumbered participants, selection bias
Debriefing	0.894	0.533	Overall, females	0.154	Creation and dissemination of materials	Demand effects, encumbered participants
Non-Differential Rewards	0.858	0.137	Overall, males	0.159	Determining and delivering performance-related pay is logistically more challenging	Salience of rewards
Opt-In	0.996	0.338	Overall (marginally)	0.100	Increased participant recruitment effort	Selection bias
Private Payment	1.246	0.223	Females	0.093	Greater logistical demands (e.g. assistants and infrastructure)	Response bias from desirability
Recruitment Database	0.480	0.196	Females	0.066	Database recruitment is administratively more costly in the short run due to scale economies	Demand effects, encumbered participants
Educational Relevance	1.172	0.447	Not significant	0.081	Targeted recruitment increases effort and reduces potential participant pool	Demand effects, encumbered participants
Non-Teacher Researcher	0.870	0.283	Not significant	0.050	Recruitment and training of assistants	Data quality
Non-Class Time	0.870	0.131	Not significant	0.022	Increased participant recruitment effort	Selection bias, demand effects, encumbered participants

Table 6

Costs and benefits of countermeasures in terms of effects on participants ethics concerns, economic costs and implications for scientific integrity. Effect sizes given as Cohen's *d* and Ordinary Least Squares beta coefficients (β).

the underlying reasons, which might include greater self-image or equality concerns in males. However, incentive compatibility is an essential feature of experimental economics. An awareness that differential rewards can be an issue may inform design choices experimentalists make at low cost. Again, related countermeasures such as the private delivery of payments, informed consent and debriefing documentation that explains incentive compatibility can help.

Opt-In Opt-in is a significant influence on *OA* in our univariate analyses. In the regressions opt-in was marginally significant overall. Implementing opt-in has two downsides, both of which seem modest when put into perspective: Participants who actively seek participation may be unrepresentative, but the available evidence does not support this (e.g. Cleave et al., 2013). Note that conversely, under opt-out, unwilling participants who

did not take the exit option may bias responses in a different way. The other is recruitment cost: *Ceteris paribus*, opt-out should generate more participants.

Private Payment Our private payment feature related to the possibility of participants seeing each other's rewards (rather than experimenter blindness to earnings). As all other features, the effect on private payment of the individual feature assessment *IA* was significant. However, private earnings delivery was significant in our analyses of overall agreement only for females. Experimental economists often choose to conduct payment privately for this and other reasons. It is not costly (unless a separate room is used) and has clear scientific benefits, such as avoiding response biases when participants fear their actions are revealed or envy caused from the money they receive.

Recruitment Database A case for using recruitment databases for ethical reasons again comes only from the perceptions of our female participants. In addition, all participants preferred the opportunity of signing up to a recruitment database to receiving unsolicited e-mail invitations in the significant effects on *IA*. Again there are also good scientific reasons why researchers may want to establish a recruitment database irrespective of ethics considerations (see Greiner, 2015, for more detail).

Educational Relevance While educational relevance was insignificant overall, it had relatively high univariate effect sizes on *IA* and *OA*. There are several steps economists can take to enhance educational relevance at relative ease, such as designing recruitment, informed consent and debrief materials with educational relevance in mind. In all these cases the timing of research (especially experiment) matters in order to avoid participants learning enough about the research and purposes so as to be unrepresentative or create demand effects.

Non-Teacher Researcher We found no evidence that dependent relationships between teacher-researcher and participant significantly influences overall perceptions of study appropriateness. However in practice ERBs may require researchers to address them. In theory this problem is simple to solve by recruiting research assistants to liaise with students. Further, advertising the names of researchers would allow students to avoid research conducted by their own teachers. Moreover, recruitment efforts could take place outside the context of a class or course and studies take place outside class time (see the relevant paragraph below). The problem for economists is mainly cost such as the expense of identifying and training suitable assistants. Some teacher-researchers may worry about the quality of data collected and the lack of quality control if they were banned from labs.

Non-Class Time Conducting research during class time can lower the cost of recruitment (Friedman and Sunder, 1994, p. 43) but involves a risk of responses biased by participants' knowledge or motivation to please the experimenter. We did not find this feature to be a significant effect on *OA* over all participants either in univariate or multivariate analysis. When economics experiments are conducted during class time, again the countermeasures for other features can alleviate some of the associated issues such as opt-in, educational relevance and dependent relationships. Student participants may find class time experiments more acceptable to the extent that participation is not assumed (and alternative activities

are offered, as discussed previously) and there is a clear relevance to the syllabus.

Gender Effects While there is no gender difference in overall appropriateness perceptions of research, we found that male and female participants respond to different countermeasures. Females' perceptions are influenced by a greater number of features, suggesting they are more sensitive to different aspects of study participation. These differentials matter in terms of sample selection, especially when research are conducted that explore gender differences or target one specific gender.

5 Concluding Remarks

We presented a first dedicated examination, both theoretical and empirical, of ethical issues involved in student participation in empirical research studies. We assessed five areas of potential ethical risk identified in policy guidelines. The present study shows these can be effectively addressed by some of the design countermeasures that were identified.⁸

Two points are worth re-emphasising before we discuss implications. The first is that our paper is not an assessment of the ethics of actual experimental economics practice. Instead we set out to examine the *potential* ethical issues that could arise when experiments are conducted with students. As a result, the ethical issues we discuss do not come from our own assessment of experimental practice but rather what is contained in typical ERB policy. Whether these issues are recognised by participants is an empirical question that we examine in our study.⁹ The results provide evidence-based responses to ethical queries from ERBs that economists increasingly face. We believe this kind of systematic and objective approach is an effective and credible way to engage these issues. We see providing economists with information about ERB concerns and counterarguments as a key motivation of this paper, similar to Greiner and Stephanides (2019). Having said that it is also important to acknowledge that experimental economists do already have a number of norms, some of which relate ethical issues.¹⁰ However the manner has been largely informal and decentralised. As a result, these ethical concerns and the efficacy of some of the best practices adopted have not been well articulated in the behavioural and experimental economics literature.

We believe our work is important and timely because the ethics of student experiments is likely to grow rather than diminish in importance in the future. Recruiting students will continue to be an important component of economic method despite the increasing recruitment of non-students (Henrich et al., 2010). Moreover, the increasing interest in field experiments, randomised controlled trials, neuroeconomics, emotions, and the dark side of human motivation will raise the potential for ethical issues in future economic experiments

⁸ It should be noted that our participants are opt-in student volunteers in an online labour market where paid study participation is common. As a result they may be biased towards positive views of experiments.

⁹ Greiner and Stephanides (2019, p. 317) discuss some results of such an exercise conducted via the mailing list of the Economic Science Association.

¹⁰ de Quidt et al. (2019) survey some of these norms (regarding experimental design, setting and reporting standards). While many of these have merely scientific motivations, others (like non-deception) also have clear ethical relevance.

more generally (Barrett and Carter, 2010; Ifcher and Zarghamee, 2015). Institutional scrutiny of human (and especially student) participants in research is likely to rise in step.

What is the way forward? We believe more debate is needed. In our experience many experimental economists have not engaged with the ethics debate and maintain ambivalent attitudes towards ERB processes. Many believe that, compared to psychology, our own topics and methods harbour fewer ethical issues in general and from student participation in particular (e.g. Blomfield, 2012; Ifcher and Zarghamee, 2015). Compared to psychology, there has been little debate of the ethical issues of experimental economics within the discipline (Barrett and Carter, 2010; Blomfield, 2012; Ifcher and Zarghamee, 2015). The standard “how-to” references date from the 1990s (Friedman and Sunder, 1994; Davis and Holt, 1993) and contain virtually no references to ethical issues.¹¹ This is also true for the key methodological contributions (Hertwig and Ortmann, 2001; Croson, 2005; Guala, 2005; Ariely and Norton, 2007; Bardsley et al., 2009; Croson and Gächter, 2010; Henrich et al., 2010). Potential issues such as deception are debated mainly in terms of its scientific rather than ethical implications (e.g. Ortmann and Hertwig, 2002). One important actual controversy, the Brisbane bus driver experiment (see Ortmann, 2015), was discussed in the media but not in professional outlets of experimental economics.

We believe that more debate is in the self interest of experimental economics. Ethics review, when carried out well, can be valuable to research. One reason is behavioural: Researchers can be too close to their own research or lack information and experience to assess ethical risks appropriately. In addition, it is not clear to what extent the researcher conducting an experiment is an appropriate judge of whether its scientific importance can offset associated ethical risks to the extent that she may be subject to selection and self serving biases. While researchers should be routinely concerned about the welfare of participants this is especially true when a professional care relationship exists, as is the case with students.

A debate over ethics should result in the creation of a professional code of conduct for experimental economics as exists for psychology. While the lack of a code is understandable due to the later development of experimental economics, its recent high profile and success demand more attention be paid by its practitioners in this area. Greater outside scrutiny will no doubt be the result. Such a code would protect not only participants but also researchers should adverse effects arise. Such a common standard would also help the profession better inform ERBs and facilitate gaining ethics clearance for economic experiments in timely fashion.

Conversely our work demonstrates that ERBs should design their own processes based on better information and evidence. The dangers of overzealous ethics review processes are clear to the extent that we found little support for some of their concerns. ERBs are often unaware of the difference in practice between disciplines (such as economics and psychology) and need to be suitably appraised. This echoes views of ERB overreach or misplaced concerns. For example, commentators like Haggerty (2004, p. 392) and Van den

¹¹ Since the first submitted version of this paper such discussion has appeared in print (Greiner and Stephanides, 2019).

Hoonard (2011), speak of an institutional 'ethics creep' to encompass practices and enforce conventions beyond the original aim of regulators that "now poses dangers to the ability to conduct university based research." Other authors like Baron (2015) as well as Page and Page (2017) suggest systematically biased and erroneous decision making by ethics review boards (ERBs) that unnecessarily hinders scientific progress and impose unjustifiable costs. To safeguard scientific progress, ethics policies should be based on better evidence. It is also worth noting that addressing ethics concerns impact on researchers unequally in that the associated financial burdens are harder for less well resourced projects and researchers to bear.

References

- Abbink, K., Herrmann, B., 2011. The moral costs of nastiness. *Economic Inquiry* 49 (2), 631–633.
- Ai, C., Norton, E. C., 2003. Interaction terms in logit and probit models. *Economics Letters* 80 (1), 123–129.
- American Psychological Association, 2010. *Ethical Principles of Psychologists and Code of Conduct*.
- Arechar, A. A., Gächter, S., Molleman, L., 2018. Conducting interactive experiments online. *Experimental economics* 21 (1), 99–131.
- Ariely, D., Norton, M. I., 2007. Psychology and experimental economics: A gap in abstraction. *Current Directions in Psychological Science* 16 (6), 336–339.
- Ball, S. B., Eckel, C. C., Rojas, C., 2006. Technology improves learning in large principles of economics classes: Using our WITS. *American Economic Review (Papers and Proceedings)* 96 (2), 442–446.
- Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., Sugden, R., 2009. *Experimental Economics: Rethinking the Rules*. Princeton University Press, Princeton.
- Baron, J., 2015. Some fallacies of human-subjects protection and some solutions. *Cortex* 65, 246–254.
- Barrett, C. B., Carter, M. R., 2010. The power and pitfalls of experiments in development economics: some non-random reflections. *Applied economic perspectives and policy* 32 (4), 515–548.
- Baumrind, D., 1964. Some thoughts on ethics of research: After reading Milgram's "behavioral study of obedience." *American Psychologist* 19 (6), 421–423.
- Baumrind, D., 1971. Principles of ethical conduct in the treatment of subjects: Reaction to the draft report of the committee on ethical standards in psychological research. *American Psychologist* 26 (10), 887.
- Bergstrom, T., Miller, J. H., 1998. *Experiments with Economics Principles*. McGraw-Hill.
- Blomfield, M., 2012. Ethics in economics: lessons from human subjects research. *Erasmus Journal for Philosophy and Economics* 5 (1), 24–44.
- Bonetti, S., 1998. Experimental economics and deception. *Journal of Economic Psychology* 19 (3), 377–395.
- Buchan, N. R., Croson, R. T., Solnick, S., 2008. Trust and gender: An examination of behavior and beliefs in the investment game. *Journal of Economic Behavior & Organization* 68 (3-4), 466–476.
- Castilla, C., Jul. 2014. Field Experiments in a Course on Behavioral Economics: Nudging Students Around Campus. *The Journal of Economic Education* 45 (3), 211–224.
- Chuah, S. H., Hoffmann, R., Lerner, J., 2013. Elicitation effects in a multi-stage bargaining experiment. *Experimental Economics* 17, 335–345.
- Chuah, S. H., Hoffmann, R., Lerner, J., 2016. Perceived intentionality in 2×2 experimental games. *Bulletin of Economic Research* 68 (S1), 78–84, (B).
- Clark, E., McCann, T. V., 2005. Researching students: an ethical dilemma. *Nurse Researcher* 12 (3), 42–51.
- Cleave, B. L., Nikiforakis, N., Slonim, R., 2013. Is there selection bias in laboratory experiments? The case of social and risk preferences. *Experimental Economics* 16 (3),

372–382.

- Commonwealth of Australia, 2018. National Statement on Ethical Conduct in Human Research 2007 (Updated 2018).
- Council for International Organizations of Medical Sciences, 2002. International Ethical Guidelines for Biomedical Research Involving Human Subjects.
- Croson, R., 2005. The method of experimental economics. *International Negotiation* 10, 131–148.
- Croson, R., Gächter, S., 2010. The science of experimental economics. *Journal of Economic Behavior & Organization* 73, 122–131.
- Dalziel, J. R., 1996. Students as research subjects: Ethical and educational issues. *Australian Psychologist* 31 (2), 119–123.
- Davis, D. D., Holt, C. A., 1993. *Experimental Economics*. Princeton University Press.
- de Quidt, J., Vesterlund, L., Wilson, A. J., 2019. Experimenter demand effects. In: Schram, A., Ule, A. (Eds.), *Handbook of Research Methods and Applications in Experimental Economics*. Ch. 16, pp. 384–400.
- Deutsch, M., 1975. Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues* 31 (3), 137–149.
- Diamond, M. R., Reidpath, D. D., 1992. Psychology ethics down under: A survey of student subject pools in Australia. *Ethics & Behavior* 2 (2), 101–108.
- Dickie, M., 2006. Do classroom experiments increase learning in introductory microeconomics? *The Journal of Economic Education* 37 (3), 267–288.
- Eagle, L., Brennan, R., 2007. Are students customers? TQM and marketing perspectives. *Quality Assurance in Education* 14 (1), 44–60.
- Eckel, C. C., Grossman, P. J., 2008. Men, women and risk aversion: Experimental evidence. *Handbook of Experimental Economics Results* 1, 1061–1073.
- Erkut, H., Nosenzo, D., Sefton, M., 2015. Identifying social norms using coordination games: Spectators vs. stakeholders. *Economics Letters* 130, 28–31.
- Exadaktylos, F., Espín, A. M., Brañas Garza, P., 2013. Experimental subjects are not different. *Scientific Reports* 3 (1213), 1–6.
- Field, S., Kuczera, M., Pont, B., 2007. *Ten Steps to Equity in Education*. OECD Publishing.
- Frank, B., 1997. The impact of classroom experiments on the learning of economics: an empirical investigation. *Economic Inquiry* 35 (4), 763.
- Friedman, D., Sunder, S., 1994. *Experimental Methods: A Primer for Economists*. Cambridge University Press, Cambridge, UK.
- Gill, D., Prowse, V. L., 2011. A novel computerized real effort task based on sliders. Available at SSRN 1732324.
- Glazer, R., June 15 2015. Feedback fatigue: Stop over-surveying your customers. *Forbes*.
- Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics* 118 (3), 1049–1074.
- Greiner, B., 2015. Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association* 1, 114–125.
- Greiner, B., Stephanides, M., 2019. Subject pools and recruitment. In: Schram, A., Ule, A. (Eds.), *Handbook of Research Methods and Applications in Experimental Economics*. Ch. 16, pp. 313–334.
- Grimes, W., March 16 2012. When businesses can't stop asking, 'How Am I Doing?'. *The*

- New York Times.
- Guala, F., 2005. *The Methodology of Experimental Economics*. Cambridge University Press, Cambridge, UK.
- Haggerty, K. D., 2004. Ethics creep: Governing social science research in the name of ethics. *Qualitative Sociology* 27 (4), 391–414.
- Harrison, G. W., List, J. A., 2004. Field experiments. *Journal of Economic Literature* 42 (4), 1009–1055.
- Henrich, J., Heine, S. J., Norenzayan, A., 2010. The weirdest people in the world? *Behavioral and Brain Sciences* 33 (2-3), 61–135.
- Hertwig, R., Ortmann, A., 2001. Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences* 24 (03), 383–403.
- Holt, C. A., 2007. *Markets, Games, & Strategic Behavior*. Pearson.
- Ifcher, J., Zarghamee, H., 2015. Experimental economics' ethics. In: E., S., Searing, D. (Eds.), *Practicing Professional Ethics in Economics and Public Policy*. Springer.
- Krause, K., 2005. *Understanding and promoting student engagement in university learning communities*, University of Melbourne.
- Krupka, E. L., Weber, R. A., 2013. Identifying social norms using coordination games: why does dictator game sharing vary? *Journal of the European Economic Association* 11 (3), 495–524.
- Levitt, S. D., List, J., 2007. What do laboratory experiments tell us about the real world? *Journal of Economic Perspectives* 21 (2), 153–174.
- List, J. A., 2008. Informed consent in social science. *Science* 322, 672.
- McNabb, R., Pal, S., Sloane, P., 2002. Gender differences in educational attainment: The case of university students in England and Wales. *Economica* 69 (275), 481–503.
- Milgram, S., 1964. Issues in the study of obedience: A reply to Baumrind. *American Psychologist* 19 (11), 848–852.
- Ortmann, A., 2015. Bus driver research by UQ's Frijters and Mujcic deserves publication. *The Australian* March 4.
- Ortmann, A., Hertwig, R., 2002. The Costs of Deception: Evidence from Psychology. *Experimental Economics* 5, 111–131.
- Page, L., Page, K., 2017. Human studies: Reforms overdue for ethical reviewing. *Nature* 544 (7649), 161–162.
- Reynolds, G., 2003. The stuttering doctor's "Monster Study". *NY Times Magazine*, 36–9.
- Russell, S. H., Hancock, M. P., McCullough, J., 2007. Benefits of undergraduate research experiences. *Science* 316 (5824), 548–549.
- Sartre, J.-P., 1956. *Being and Nothingness*. Philosophical Library.
- The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. US Government Printing Office.
- U.S. Department of Health and Human Services, 2010. *International compilation of human research protections*.
- U.S. Government, September 8 2015. Federal policy for the protection of human subjects: Proposed rules. *Federal Register* 80 (173).
- Van den Hoonaard, W. C., 2011. *The seduction of ethics: Transforming the social sciences*. University of Toronto Press.

- Vesely, S., 2015. Elicitation of normative and fairness judgments: Do incentives matter? *Judgment and Decision Making* 10 (2), 191.
- Welch, A. B., 2000. Responding to student concerns about fairness. *Teaching Exceptional Children* 33 (2), 36–40.
- Wenzel, M., 2005. Motivation or rationalisation? Causal relations between ethics, norms and tax compliance. *Journal of Economic Psychology* 26 (4), 491–508.
- Zimbardo, P. G., 1973. On the ethics of intervention in human psychological research: With special reference to the Stanford prison experiment. *Cognition* 2 (2), 234–256.
- Zizzo, D. J., 2008. Experimenter demand effects in economic experiments. *Experimental Economics* 13 (1), 75–98.

Appendix

DV: <i>IA</i> for	DB	IC	ER	OI	TE	RD	PP	CT	DR
Age	-0.004 (0.010)	-0.010 (0.010)	-0.105* (0.010)	-0.057 (0.011)	-0.057 (0.009)	-0.082 (0.009)	-0.041 (0.011)	-0.074 (0.009)	-0.095** (0.007)
U.S. Birth	0.020 (0.283)	0.021 (0.280)	0.016 (0.249)	0.042 (0.252)	-0.039 (0.286)	-0.007 (0.185)	0.016 (0.268)	0.019 (0.238)	0.046 (0.243)
Female	0.063 (0.154)	0.071 (0.184)	-0.052 (0.178)	-0.182** (0.166)	-0.090 (0.176)	-0.021 (0.165)	0.045 (0.163)	0.005 (0.158)	0.074 (0.153)
Positive Version	0.343*** (0.136)	0.498*** (0.163)	0.405*** (0.149)	0.242*** (0.146)	0.249*** (0.155)	0.347*** (0.135)	0.314*** (0.164)	0.338*** (0.136)	0.490*** (0.132)
Positive Version × Female	-0.116 (0.212)	-0.116 (0.240)	0.017 (0.226)	0.119 (0.219)	0.018 (0.235)	-0.096 (0.214)	0.005 (0.232)	-0.131 (0.212)	-0.134 (0.204)
Constant	2.519*** (0.371)	2.099*** (0.350)	2.934*** (0.362)	2.891*** (0.371)	3.088*** (0.347)	3.056*** (0.292)	2.418*** (0.402)	2.813*** (0.304)	2.764*** (0.304)
<i>N</i>	294	294	294	294	294	294	294	294	294
<i>R</i> ²	0.083	0.195	0.186	0.118	0.080	0.106	0.106	0.087	0.190

Table 7

Ordinary Least Squares regression results for participants' individual assessment of the respective features of the hypothetical study. Features denoted by their acronyms. Standardised beta coefficients with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

DV: <i>OA</i>		
Age	0.027	(0.009)
U.S. Birth	0.124 *	(0.255)
Female	-0.091	(0.324)
Age	-0.026	(0.009)
U.S. Birth	0.102*	(0.193)
Female	-0.278	(0.314)
Positive Version for:		
DB	0.062	(0.149)
IC	0.133*	(0.143)
ER	0.058	(0.138)
OI	0.051	(0.147)
TE	0.103	(0.142)
RD	-0.030	(0.150)
PP	-0.009	(0.142)
CT	-0.036	(0.145)
DR	0.247***	(0.148)
Positive Version × Female for:		
DB	0.155	(0.207)
IC	0.026	(0.206)
ER	0.030	(0.212)
OI	0.028	(0.207)
TE	-0.076	(0.218)
RD	0.206**	(0.212)
PP	0.185*	(0.200)
CT	0.122	(0.207)
DR	-0.173	(0.203)
Constant	1.883***	(0.329)
<i>N</i>	293	
<i>R</i> ²	0.164	

Table 8

Ordinary Least Squares regression results for participants' overall agreement with the hypothetical study. Standardised beta coefficients with standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.