Experimental Economics (2022) 25:385–412
https://doi.org/10.1007/s10683-021-09726-7

ORIGINAL PAPER

What is considered deception in experimental economics?

Gary Charness[1] · **Anya Samek**[2] · **Jeroen van de Ven**[3]

Received: 21 January 2021 / Revised: 29 June 2021 / Accepted: 12 July 2021 /
Published online: 1 September 2021
© The Author(s) 2021

Abstract

In experimental economics there is a norm against using deception. But precisely what constitutes deception is unclear. While there is a consensus view that providing false information is not permitted, there are also "gray areas" with respect to practices that omit information or are misleading without an explicit lie being told. In this paper, we report the results of a large survey among experimental economists and students concerning various specific gray areas. We find that there is substantial heterogeneity across respondent choices. The data indicate a perception that costs and benefits matter, so that such practices might in fact be appropriate when the topic is important and there is no other way to gather data. Compared to researchers, students have different attitudes about some of the methods in the specific scenarios that we ask about. Few students express awareness of the no-deception policy at their schools. We also briefly discuss some potential alternatives to "gray-area" deception, primarily based on suggestions offered by respondents.

Keywords Experimental methodology · Deception · Attitudes

JEL Classification B49 · C80 · C91 · C92 · C93

1 Introduction

There is a strong norm against the use of deception in experimental economics (Ortmann, 2019). This is reflected in the policy of academic journals. *Experimental Economics*, for instance, does not consider studies that employ deception.[1] But precisely

[1] See https://www.springer.com/journal/10683, accessed on December 29, 2020.

✉ Anya Samek
asamek@ucsd.edu

[1] Department of Economics, University of California, Santa Barbara, Santa Barbara, CA, USA

[2] Rady School of Management, University of California, San Diego, La Jolla, CA, USA

[3] Amsterdam School of Economics, University of Amsterdam, Amsterdam, Netherlands

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

what constitutes deception is unclear. This issue is a major methodological concern for experiments both in the lab and in the field. While providing explicitly false information is a clear example of deception, there are many "gray-area" practices, such as not including potentially-relevant information or using implicitly-misleading language. Reviewers, editors, and authors may not share the same sentiments on these issues. A reviewer may assert that the experimental methodology involved deception and reject a paper; however, the authors or other reviewers may not agree that deception was used.

Several arguments can justify the norm against using deception. The first is that it can be considered unethical to deceive participants. The second is the desire to maintain experimental control. The concern is that subjects who participate in an experiment with deception may not believe the researcher in future experiments, and therefore their decisions may well be capricious and unreliable. Importantly, this can result in potential negative externalities for other researchers, including subject selection bias. Another concern related to control is control within the experiment, i.e., participants may not believe the experimenter if they believe that he or she is deceiving them.

The policy against deception is said to originate with Sidney Siegel, the noted statistician and psychologist who was active in the 1950s and 1960s.[2] In the following decades, Vernon Smith and Charles Plott were key figures who implemented the norm in experimental economics (Svorenčík, 2016). Strong views have been expressed in the literature: Wilson and Isaac (2007, p. 5) write "In economics all deception is forbidden. Reviewers are quite adamant on this point and a paper with any deception will be rejected." Gächter (2009, footnote 16) states: "Experiments … which use deception are normally not publishable in any economics journal."

Following the norm requires an agreement between authors, reviewers and editors about what constitutes deception. There is some agreement in the literature that explicit misinformation is deceptive, whereas omission of some information may not be deceptive. Hey (1998, p. 397) points out "there is a world of difference between not telling subjects things and telling them the wrong things. The latter is deception, the former is not." Hertwig and Ortmann (2008, p. 62) assert that "a consensus has emerged across disciplinary borders that intentional provision of misinformation is deception and that withholding information about research hypotheses, the range of experimental manipulations, or the like ought not to count as deception."[3]

Ortmann (2019) writes that a simple norm to not allow deceptive acts of commission but to allow only a minimal set of acts of omission should be implemented. But what are the attitudes towards practices that while not explicitly stating false information nevertheless do not provide complete information or use misleading language? Determining the information that must be provided to the subjects is also

[2] Vernon Smith stated that Siegel had two precepts: (1) Participants have to be paid, and (2) Participants have to believe what they are being told. A discussion of the history behind deception is also provided in Svorenčík (2016).

[3] In a philosophical paper, Hersch (2015) makes the argument that banning explicit, but not implicit, deception is inconsistent.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

not a trivial task, since one can hardly state everything that is known (e.g., the history of play in the game in previous studies). And do the circumstances matter? For example, is misleading language more acceptable if there is no other way to gather important data? Researchers could benefit from having a sense of what are considered acceptable methods and what methods are outside the bounds.

We conducted a study among professional experimenters and student subjects, in which we surveyed attitudes towards what might be considered deceptive practices. We did not try to define "deception" in our survey, but instead simply asked respondents to rate levels of deception in different common scenarios. We recruited professional experimenters who were listed on IDEAS/RePEc. Our global survey had a response rate of 51% (yielding 788 responses out of the 1,554 researchers that we attempted to contact). We also surveyed students (from three different universities) who had participated in experiments as undergraduates, yielding 445 responses.

As described in Sect. 2, several articles have provided discussions of what constitutes deception in experimental economics. A handful have also provided some empirical evidence through surveys of experimenters and student subjects. We believe our study adds value because it is more comprehensive than previous work, drawing from a wide pool of researchers from around the globe and offering a higher number of observations than previous studies. The higher number of observations means that we can investigate views on deception by subgroups, for example by comparing researchers in Europe to those in North America, or by comparing researchers who now serve in editor or referee roles to those who do not. One contribution of our survey is that we asked respondents to rate a number of experimental methods on several dimensions, including how deceptive they consider the methods to be and how negative their attitude is towards each method. We also asked students to indicate whether a specific method would affect their answers and/or participation in future studies.

Another contribution is that we asked respondents to suggest alternatives to the proposed methods, which could benefit researchers who consider using a particular method. We present some of these alternatives in Sect. 5. One suggestion is to gather more data where doing so is feasible and not too costly. For example, it is better to gather more data to allow for perfect stranger matching (if this is the intent) than to mislead by being vague about the matching protocol. Another suggestion is instead of surprising subjects with a re-start, tell them that more parts of the experiment will follow the initial part announced in the beginning. The strategy method is also suggested as a useful technique to obtain data from decision nodes that are not reached that often. Whenever possible and financially feasible, these recommendations help avoid even gray-area deception.

Our results show that most researchers feel it is important to avoid deception and that loss of experimental control is a key issue. However, similar to the findings of Krawczyk (2019), our responses are heterogeneous in the sense that some people are less averse to deception than others. For example, compared to North Americans, researchers from Europe find deception more unethical. We find virtually no difference in responses across experimenter reviewer roles (e.g., editors, reviewers and non-reviewers).

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

There are differences in opinion across our seven specific scenarios; in particular, the data from researchers indicate a perception that costs and benefits matter, so that such practices might in fact be appropriate when the topic is important and there is no other way to gather data.[4] A potentially-important point is that some researchers are willing to make a tradeoff between the costs and benefits of deception, so that a reasonable argument can be made for some gray-area violations being appropriate. In fact, in several of our gray-area scenarios under this condition, about half of the researchers felt that deception would be mostly appropriate if there is no other good way to answer the question.

Interestingly, we found that researcher and student respondents disagreed to some extent on which specific scenarios they found most inappropriate. Further, 75% of students stated that they would be willing to participate as often (or even more often) in lab experiments even after knowing that they were deceived in an experiment. Yet it is important to keep in mind that even though most students say that they would be likely to return to experiments when they know they have been deceived, this leaves open the possibility (mentioned in Cason and Wu, 2018) that students who return after being deceived show a selection bias and may make choices that are less consistent or even arbitrary. One must take care to avoid this potential problem, since it suggests negative spillovers to other researchers and to future research. Finally, most students (73%) were not aware of the no-deception policy at the experimental economics labs of their schools, with many students believing that the use of deception is, in fact, common practice. This suspicion towards experimenters has also been documented in other studies (for example, see Frohlich et al., 2001).

While we do not wish to try to define "deception" in experiments, we nevertheless believe that a blanket ban on policies in the gray area may be too strict. In this, we echo Cooper (2014, p. 113) who states: "only an extremist would claim that experimenters (or economists in general) should never use deception" and goes on to list four conditions that, if jointly satisfied, might serve as a guide for when deception might be allowable. We discuss this later in detail, but here content ourselves by listing one condition (Cooper, 2014 p. 113): "The value of the study is sufficiently high to merit the potential costs associated with the use of deception."

We do offer some recommendations. First, few students report being aware of the no-deception policy at the lab in their former schools. We believe it is a good idea to make potential subjects better aware of the policy. Second, having a more nuanced view of deception that takes into account factors such as costs and benefits seems advisable. Third, journals could be more explicit about practices that are considered to be deception. The current state of affairs exposes researchers to the idiosyncrasies of reviewers' opinions. While it seems preferable that journals harmonize their policies towards deception, we are not yet close to this point. We can only hope that the results from our survey will lend some clarity to the issue. Fourth, more research on

[4] Our reading of the researchers' comments in the free-form answer fields is consistent with this. They frequently pointed out that they needed more information to judge specific scenarios, indicating that the context matters.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

students' attitudes and behavioral reactions would be welcome, as their views appear to be divergent from those of researchers.

The remainder of this paper is organized as follows. Section 2 reviews the literature. Section 3 discusses the details of how we conducted our study. Section 4 presents the views of both researchers and students. Section 5 summarizes some alternatives to (potentially) deceptive methods. Section 6 closes with a summary and some recommendations.

2 Literature review

Textbooks and handbook chapters have provided discussion about the importance of avoiding deception in experimental economics as a way to maintain control and reduce negative spillovers. In a textbook on experimental economics, Davis and Holt (1993) began the first chapter by explaining the necessity for avoiding intentional deception, noting that (p. 23) "subjects may suspect deception if it is present […] it may jeopardize future experiments if subjects ever find out that they were deceived and tell their friends." In a handbook chapter, Ledyard (1995) noted (p. 134) "if the data are to be valid, honesty in procedures is critical." In a more recent handbook chapter, Ortmann (2019) wrote (p. 28) "until recently, editors and referees enforced the norm" but raised the concern that the norm is no longer comprehensively enforced.

While the textbooks have issued guidelines to avoid deception, a small literature has emerged arguing about the merits of such guidelines. Related work can be organized into two types. The first type provides a general discussion about deception and discusses the costs and, to a lesser degree, the benefits of its use in experiments. Such discussion has been published over the past two decades, with arguments on both sides of the debate. For example, Bonetti (1998) noted that there is little evidence that deception should be forbidden, either on the grounds of loss of control or external validity; however, McDaniel and Starmer (1998) suggested that Bonetti (1998) had under-estimated the negative externalities created by deception. Several papers have also argued that deliberate misinformation is deceptive, whereas omission of information is not necessarily so (e.g., Hey, 1998; and later, Hertwig & Ortmann, 2008; Wilson, 2016).

Ortmann and Hertwig (2002) provided a discussion of deception in *Experimental Economics*, writing about the disparity in deception practices between psychologists and economists and drawing on prior literature from the psychology field, where deception has long been common.[5] Their systematic review of psychological evidence found that having been deceived generates suspicion that may well affect the decisions made by experimental participants. They concluded (p. 111): "The prohibition of deception is a sensible convention that economists should not abandon."

[5] Policies seem to be changing in this field, however. We thank the editor for pointing out that the American Psychological Association offers guidelines (https://www.apa.org/ethics/code#807) for the use of deception that overlap to a degree with our own views.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

However, Hertwig and Ortmann (2008) defended the use of deception to some extent, studying whether deceived subjects resent having been deceived, whether any such suspicion affects decisions, and whether deception is an "indispensable tool for achieving experimental control." Here they summarized papers from the 1980s and 1990s from the psychology literature that evaluated feelings toward deception and concluded that the evidence is not clear-cut, partly because the types of deception varied across studies. They concluded that "one may decide to reserve deception for clearly specified circumstances."

More recently, Cooper (2014) wrote an op-ed piece discussing how some forms of deception might be permissible under some circumstances and formulated these four rules:

1. The deception does not harm subjects beyond what is typical for an economic experiment without deception.
2. The study would be prohibitively difficult to conduct without deception.
3. Subjects are adequately debriefed after the fact about the presence of deception.
4. The value of the study is sufficiently high to merit the potential costs associated with the use of deception.

This non-lexicographic approach implicitly considers trade-offs and gives some opening for employing deceptive tactics, but it doesn't address specifics regarding the uses of gray-area deception.

Similar discussion, with views on both sides, is happening in the area of agricultural and resource economics. Cason and Wu (2018) wrote an op-ed piece for *Environmental and Resource Economics*, addressing the use of deception in agricultural and resource economics. They recognize that "the omission of benign details of the experiment environment" has been tolerated in experimental economics, but argue that the agricultural and resource economics journals "should adhere to the wider experimental economics norms against deception." Lusk (2019) wrote about deception as it relates to food and agricultural experiments in *Food Policy*, arguing against a blanket ban and calling for a more nuanced view.

While the studies discussed above provided reasoned arguments and discussions of deception, including discussions of some of the specific scenarios that we consider (e.g. Wilson, 2016), the work closest to ours used survey evidence to understand views of deception among the broad population of researchers and student subjects. Krawczyk (2019) conducted a survey with 143 graduate students, postdoctoral researchers and professors recruited from the Economic Science Association (ESA) e-mail list and about 400 undergraduate students from the University of Warsaw. He presented several methods to respondents (several of which overlap with ours, see Sect. 3) and had respondents rate these scenarios on deceptiveness. He found that students are generally more tolerant of deception than researchers but reports a high and significant correlation between ratings of researchers and students. He concluded that there is considerable heterogeneity in responses, both among students and among researchers. Generally, making false statements is seen as worse than deception through omission, and deception is seen as worse when it

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

affects behavior or future participation. In light of these results, Krawczyk (2019) proposed a more nuanced policy towards deceptive practices as well as a typology for classifying deception by whether it is intentional or explicitly false and whether it is likely to affect subjects' behavior or willingness to participate.

Both our study and Krawczyk's (2019) study collected data on views about the deceptiveness of various commonly used methods that may be considered a deception "gray area." However, a serious difference is that we also asked researchers to rate the scenarios on important other dimensions, including how negatively the scenario is viewed generally and how negatively the scenario is viewed if there is no other way to answer the question. We further asked researchers to recommend alternatives to each practice, which we discuss in Sect. 5. In addition, we report on information from subjects about behavioral responses to each deception scenario.[6] Because the norm against deception is mostly driven by a concern about behavioral responses of subjects, we believe that these dimensions are important to study.

Krasnow et al. (2018) surveyed attitudes of psychologists and economists, measuring suspicion levels and behavior in four common economic tasks.[7] They found that (1) Psychologists are less bothered than economists by deception, and (2) Subjects are not so concerned about deception and their choices in experiments are unaffected by the possibility of deception. The results from correlating behavior in a series of experimental games and survey questions to prior participation in experiments with deception indicated that participants' present suspicion was unrelated to past experiences of deception, with suspicious participants behaving no differently than credulous participants. They concluded (p. 28): "banning all deceptive studies from economic study pools and journals cannot be justified on pragmatic grounds. It may be time to end the ban on deceptive methodology."

To shed light on the causal impact of being deceived, Jamison et al. (2008) intentionally deceived subjects in an experiment and then compared their later behavior to subjects who were not deceived. They found significant differences in selection into the later experiments as well as some differences in behavior in an ensuing game. This provides support for banning deception on the basis of reducing negative externalities. However, in a similar study conducted in a social-psychology lab, Barrera and Simpson (2012) found an effect of deception on subjects' beliefs about the use of deception but no effect on subjects' behavior in subsequent experiments.

Similarly, Krawczyk (2015) manipulated messages sent to prospective subjects aimed at reducing suspiciousness about deception, and found that they had an impact on self-reported mistrust in the experiment but not on behavior. A tentative conclusion is that deception may have a larger effect on self-reported trust or beliefs than on behavior in an experiment. However, even if effects on behavior in an experiment

[6] Krawczyk (2019) also collects behavioral responses (using different questions) but does not report on the results in his paper.

[7] About 200 economists participated in this survey. Related studies by Colson et al. (2015) and Rousu et al. (2015) surveyed undergraduate students and a small number of agricultural and applied economists. Among possible deceptive scenarios, they found that providing false or incomplete information or not making subjects aware that they were in an experiment were rated as least severe. Most respondents agreed that not making promised payments and inflicting physiological harm should be banned.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

are limited, the effects on selection into future experiments observed by Jamison et al. (2008) are important to recognize.[8]

Our paper is primarily devoted to providing attitudes towards gray-area deception, taking as a starting point that explicit deception should be avoided. We evaluate the beliefs of a large number of researchers all across the world. Because we are able to survey over 50% of all experimental economists who have published a threshold level of work on RePEc, we feel that we provide a more complete picture of the beliefs of the profession as a whole relative to e.g., the work by Krawczyk (2019); he recruits respondents from the ESA mailing list, which is a smaller and perhaps more selected sample. In addition, we survey experimental participants at three major experimental laboratories, one in Europe and two in North America. The lab in Tucson has a long and storied history and the lab at Nottingham serves the largest experimental economics group in the UK, while the UCSB lab has been quite active in the past 20 years. Related work either surveys students from one laboratory (Krawczyk, 2019) or one country (Krasnow et al., 2019).

The size and scale of our survey makes it arguably the most representative to date. We are also the first study to examine attitudes across professional levels, finding that attitudes towards deception do not vary across professional status. In addition, we examine differences in attitudes across regions of the world. Finally, we consider the issue of costs and benefits rather than intent or operationalization as prior studies do.

3 Study design

3.1 Researcher survey

We conducted two surveys. In the first survey (conducted in the Fall of 2018), we focused on researchers in experimental economics. To create a list of potential respondents, we used IDEAS (https://ideas.repec.org/), which is the largest bibliographic database dedicated to economics that is freely available on the Internet. IDEAS is based on RePEc (http://repec.org/), which at the time of our survey in 2018 included over 2.7 million working papers or publications with 68,869 registered authors. We used the RePEc list of experimental economists, which includes 1,705 authors affiliated with 1,906 different institutions (https://ideas.repec.org/i/eexp.html). The list is compiled based on the NEP-Experimental Economics report on IDEAS, which is issued weekly and maintained by volunteers. The list includes any author who either (1) has had at least 5 papers published in the NEP-Experimental Economics report or (2) for junior authors (publishing their first paper less than 10 years ago), has at least 25% of his/her papers published in an NEP-Experimental Economics report. We chose to use IDEAS (vs, e.g., Google Scholar) because we felt it would give us the most comprehensive listing of experimental economists available.

[8] Zultan (2015) compared subjects who had and had not previously participated in psychologyexperiments and found that previous participation in a large number of psychology experiments was associatedwith a small reduction in estimated trust.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

To find author contact details, we used the author's IDEAS webpage when available. When the e-mail was not directly available, we searched for the author on Google and used the email listed on his or her professional or personal website. We were able to identify an email address for all of the 1554 researchers.

We next formulated an email invitation (available in Online Appendix 1.1) and mail-merged the invitation to send personal e-mails from Gary Charness's e-mail account to each author. Because we wished to understand whether and how beliefs are associated with impact in the profession, we also collected data on each researcher's h-index (which measures an author's impact based on citation count and productivity). We sent three different survey links—one each of the top third, middle third and bottom third of the h-index—with the same messaging.[9] This allows us to test for differences in beliefs across status in the profession.

We programmed the survey (questions presented in Online Appendix 1.2) using the Qualtrics survey software (qualtrics.com). In the first part of the survey, we asked respondents for basic background information, including the continent on which they are located, the year of their Ph.D., whether they are a graduate student, assistant, associate or full professor, and whether they have held any editorial positions. We also asked what proportion of the respondent's research is experimental, and what proportion uses laboratory versus field experiments. In the second part of the survey, we asked respondents to answer several questions about their views on deception using 7-point scales, including: the extent of respondent beliefs about whether it is unethical to deceive in experiments, whether potential loss of control due to deception is a serious problem, whether it is important to avoid deception, whether deception is useful, and how often they have observed deception in papers they have reviewed or handled as an editor.

We also presented seven scenarios describing different experimental techniques, asking respondents to rate each scenario on deceptiveness, usefulness, the degree of appropriateness if no other tools are available, and how negative their reaction would be towards the technique if they reviewed a paper that used it. Table 1 lists the seven scenarios used, with the exact text respondents saw provided in Column 2.[10]

We chose scenarios based on techniques we have seen employed in papers (lab and field) that differ in the type of deception used as well as in level of deception used (as we perceive it). For example, scenarios S1–S3—which include techniques such as surprise re-start—are fairly commonly used in economics experiments. S5—use of confederates—is fairly uncommon in economics but fairly common in psychology. S6—not informing participants that they are in an experiment and asking for unpaid effort (in the sense that they do not receive participation or other payments)—is employed in many field experiments. S4 and S7 include some degree of omission or misdirection, and are also sometimes used in lab experiments. At the end of the survey, respondents were given the opportunity to provide free-form comments. We offered $100 rewards (payable either by Amazon.com gift card, PayPal

[9] Note that anonymity was preserved, since we cannot link any responses to names (unless the respondent volunteered to give his or her name, in which case it was clear that we would be able to look up this information).

[10] Our scenarios partly overlap with those used in Krawczyk (2019). Krawczyk did not include unknown/unpaid participation and misinterpretation.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

Table 1 Scenarios

Scenario	Researcher text	Student text
S1: Subgroup re-match	In a multi-period experiment, the experimenter tells the participants that they will be randomly matched every period, but in fact the participants are only re-matched (for statistical purposes) within a subgroup of the participants	In a multi-period experiment (many periods or rounds), the experimenter tells the participants that they will be matched at random with another participant in every period. In fact, the participants are only re-matched (for statistical purposes) within a part of the group (a subgroup). This means that some participants will never be matched. Usually, participants are anonymous so they cannot tell with whom they are matched
S2: Surprise re-start	Participants in an experiment are told that there will be 10 periods in the session, but are then told that there will be another 10 periods (a surprise re-start)	Participants in an experiment are initially told that there will be 10 periods (or rounds) in the session and that the tenth period will be the last one. After 10 periods, they are then told that there will be another 10 periods. (Participants know beforehand how long the experiment will take, and the experiment does not take longer than announced)
S3: Non-representative sample	The experimenter tells the participants the average value of the choices or beliefs of "a sample of the other participants", but doesn't mention that this is not a representative sample (and states other averages to other participants)	The experimenter tells the participants the average value of the choices or beliefs of "a sample of the other participants". The experimenter doesn't mention that this is actually a specific part of the group that was selected by the experimenter (who states other averages to participants in other sessions)
S4: Unexpected data use	The experimenter uses participant responses in a way that is not revealed to the participant: for example, (1) participants are incentivized to predict behavior of other people, but are not told that these predictions will be shown to others, or (2) participant data from one part of the experiment is used to sort participants into groups in another part of the experiment	The experimenter uses participant responses in a way that is not revealed: for example, (1) you are incentivized to predict behavior of other people, but are not told that these predictions will be shown to others, or (2) participant data from one part of the experiment is used to sort participants into groups in another part of the experiment
S5: Confederates	The experimenter uses either confederates or computers that do not operate of their own volition, but instead behave as scripted by the experimenter. The experimenter does not tell subjects that confederates or computers are involved in the experiment	The experimenter uses either confederates (actors hired by the experimenter) or computers who behave as scripted or programmed by the experimenter. The experimenter does not tell participants that confederates (actors) or computers are involved in the experiment, and participants are led to believe that they are interacting with another participant like themselves

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

Table 1 (continued)

Scenario	Researcher text	Student text
S6: Unknown/unpaid participation	The experimenter conducts a field experiment that encourages people to put forth (unpaid) effort or take action, but does not inform the participants that they are in an experiment	The experimenter conducts a field experiment (an experiment outside of the laboratory) that encourages people to make (unpaid) effort or take action, but does not inform the participants that they are in an experiment
S7: Misinterpretation	The experimenter relies upon the assumption that participants will misinterpret the instructions [e.g., using the term "random" when the probabilities are actually 75% and 25% and when it is essential that they believe that this was truly random (i.e., 50%)]	The experimenter expects that people will misinterpret the instructions. For example, the experimenter uses the word "random" when the probability of an event is actually something else (75% or 25%, for example) and when it is essential that they believe that this was truly random and the probability is 50%

money transfer, or personal check) to four randomly-selected survey respondents. Respondents had to include their email address at the end of the survey to enter our lottery, but we did not use this identifying information in our analysis.

We emailed 1,554 researchers and received 63 bounce-backs. A total of 788 respondents started the survey, for a response rate of 53% (788 of 1491) excluding bounce-backs. Thirty-two respondents stopped before the second part of the survey, and we drop them from the sample. This gives us a sample of 756 respondents who are included in our analysis. Not all respondents answered all questions. In part 3 of the survey, in which we ask about attitudes towards different scenarios, we have between 669 and 684 responses for each scenario.

Table A2.1 in Online Appendix 2 provides some descriptive statistics about the researchers in our sample and explains how we categorize respondents. One third of respondents hold an (associate) editorial position at an economics journal. Most other respondents are reviewers for economics journals. The vast majority (85%) use experiments for at least 50% of their studies. Most of our respondents come from North America (28%) and Europe (61%).

3.2 Student survey

We were also interested in understanding how students who have participated in economic experiments perceive deception in economics. We surveyed subject pools from laboratories at three universities in March 2020. This survey included the University of California—Santa Barbara (Experimental and Behavioral Economics Lab), the University of Arizona (Economic Science Lab), and the University of Nottingham in the United Kingdom (CeDEx). The email script for the invitation is available in Online Appendix 1.3.[11]

The student survey followed a similar structure to the researcher survey; however, for the students we were also interested in understanding how they would react to deception as a participant. Questions are available in Online Appendix 1.4. We first asked students about their continent of origin, how many experiments they had participated in at the economics lab, and how many experiments they had participated in at other labs. Second, we asked students to rate on a 7-point scale how often they believe the lab uses deception, how often they believe other labs use deception, whether they believe it is unethical to use deception, and whether it is important to avoid deception. We also asked them if knowing that they were deceived would affect their participation (less often, no impact, more often) in the same lab or in different labs.

[11] In 2018, we sent out an earlier version of the survey to subjects from the University of California-Santa Barbara, the University of Arizona, and the University of Magdeburg in Germany (MaXLab). The scenarios in that survey used slightly different wording and the sample of respondents is small (126 completed responses). For the most part, we do not report the results from this survey here to save space, but the results are quite similar to the survey we report on here. We do indicate in the main text where results differ. All data are available by request.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

In the third part of the survey, we presented students with the same seven scenarios as the researchers (randomly ordered). We kept the wording similar, but we added some brief explanations. The text of the scenarios is displayed in the third column of Table 1. Students were asked to rate the scenarios on a 7-point scale on the extent of deception used, how negative their reaction would be if this technique was used in an experiment in which they participated, how appropriate this technique would be if the issue was important and there was no other good way to answer the question, how likely they would be to participate in future experiments after they had participated in an experiment that used this technique, and how likely it would affect their answers (how much attention they would pay or the kind of answer they would give) if they knew this technique was used. At the end, students were told about the no-deception policy at their lab and asked if they had been aware of the policy.

We offered $100 rewards (payable either by Amazon.com gift card, PayPal money transfer, or personal check) to two randomly selected survey respondents from each lab. Students had to include their email address at the end of the survey to enter our lottery, but we did not use this identifying information in our analysis.

We received 445 completed responses. Table A2.2 in Online Appendix 2 provides some descriptive statistics. On average, respondents had previously participated in six experiments in the lab from which we recruited them, and in one experiment in other labs. Most respondents participated in 10 or fewer experiments in the lab from which they were recruited (87%), and in three or fewer experiments in other labs (91%).

4 Results

In this section we describe the results. For questions where we made use of a Likert scale, the answers do not always have an immediate interpretation. Given that our scale runs from "not unethical at all" to "very unethical", the midpoint (4) cannot be treated as a neutral attitude but should be interpreted as indicating that the respondent finds it at least somewhat unethical. Our main interest in those cases will be to make relative comparisons between questions or populations, implicitly assuming that different populations interpret the scale in the same way.

4.1 Researchers

Figure 1 plots the empirical CDFs of researchers' ratings for the general questions about deception. The 7-point scale runs from 1 ("not at all") to 7 ("very"). We find heterogeneity in responses, in the sense that responses are not concentrated on a single answer. About six percent of respondents view deception as not unethical at all, while 20 percent consider deception to be very unethical. The other respondents are dispersed over the remaining bins, with at least nine percent of respondents in each bin. About one third of respondents think that the loss of control is a very serious problem (a rating of 7), while 42 percent answered 5 or 6. The remaining 24 percent

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

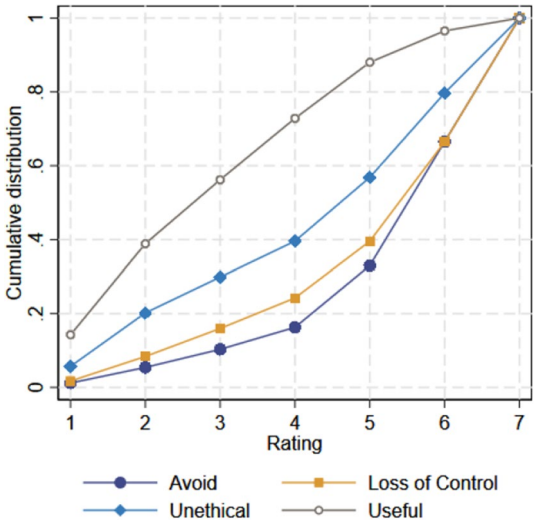


Fig. 1 Empirical CDF of researchers' ratings for questions about deception. Items are measured on a 7-point scale. Unethical: *Do you feel that it is unethical for experimenters to deceive participants in their experiments (even after debriefing)?* (1 "not unethical at all", 7 "very unethical"). Avoid: *To what extent do you feel it is important to avoid deception in experiments in practice?* (1 "not at all important to avoid", 7 "extremely important to avoid"). Loss of control: *Do you feel that the potential loss of control due to deception is a serious problem?* (1 "not a serious problem at all", 7 "very serious problem"). Useful: *How useful is deception as a tool in experimental economics?* (1 "not useful at all", 7 "extremely useful")

Table 2 Pairwise correlations of researcher views

	Unethical	Loss of control	Useful
Loss of control	0.45***	1	
Useful	−0.38***	−0.40***	1
Avoid deception	0.59***	0.70***	−0.48***

Spearman's rank correlation coefficients. ***$p<0.01$. All items are measured on a 7-point scale, from 1 "not at all" to 7 "very/extremely"

answered with at most 4. The answers to the question whether it is important to avoid deception follow roughly the same pattern. On whether deception can be useful, 12 percent of respondents answered 6 or 7 ("extremely useful"), while each of the other bins has at least 14 percent of respondents.

Table 2 shows the correlations between researchers' ratings across the four general questions. Ratings are all significantly correlated and in the expected direction. That the correlations with "useful" and other items are lower could reflect that this question is more ambiguous, and researchers may have different perceptions about what is meant by "useful."

Although we cannot show any causal links, it makes sense to view the factors "unethical," "loss of control," and "useful" as the inputs to views on whether

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

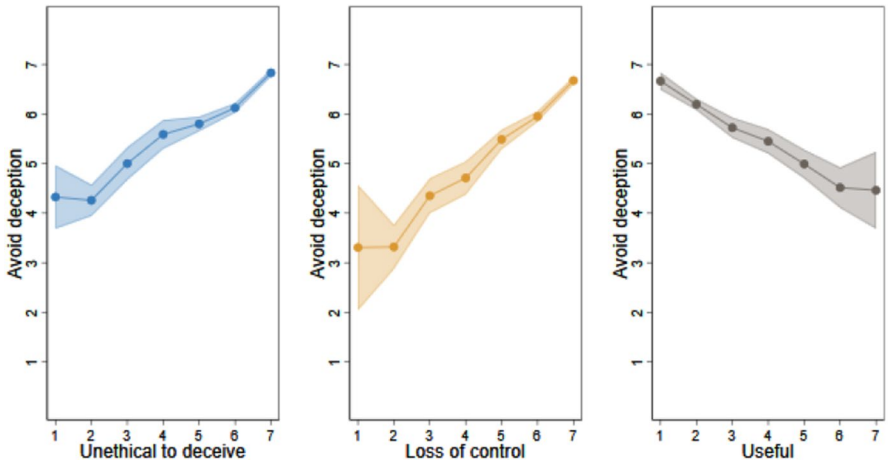


Fig. 2 Mean attitude towards the need to avoid deception by attitudes with respect to unethical to deceive, loss of control and usefulness. All attitudes are on a 7-point scale, ranging from 1 ("not at all") to 7 ("very/extremely"). The sample is the researcher respondents. Shaded areas are the 95% CIs. See the caption of Fig. 1 for the exact questions

deception should be avoided. Figure 2 plots the mean attitude towards the importance to avoid deception by each of those inputs.

Those who think it is very unethical to deceive, or who believe that loss of control is a serious problem, find it important to avoid deception. Avoiding deception is deemed less important as it becomes more useful.

We also find heterogeneity in attitudes across different groups of respondents. As mentioned before, such comparisons are only valid if those different groups interpret the answer scale in the same way. Table 3 reports mean responses by continent, where we distinguish between North America, Europe, and rest of the world.[12] Compared to North Americans, researchers from Europe find deception more unethical (mean difference of 0.56 points, $p<0.001$, two-sided Wilcoxon–Mann–Whitney test) and are more concerned about a loss of control (mean difference of 0.38 points, $p=0.055$).[13] Researchers from Europe also see more need to avoid deception, but the difference is not significant at conventional levels (mean difference of 0.32 points, $p=0.113$). In terms of perceived usefulness of deception, there is no significant difference (mean difference of 0.11 points, $p=0.675$). Figure 3 plots the mean ratings for "unethical" and "avoid." Researchers outside of Europe and North America hold views that are between those of North Americans and Europeans.

Table 3 also splits perceptions by respondent characteristics, including status in the profession (graduate student/researcher, reviewer/editor, RePEc ranking—top/

[12] North America ($N=208$) and Europe ($N=458$) make up the vast majority of respondents (89 percent). For Asia we have a further 51 respondents, for Australia/Oceania 28, for South America 9, and for Africa 2.

[13] After correcting for multiple hypothesis testing (such as a Bonferroni correction), the difference for "unethical" remains significant at the 1 percent level but there is no significant difference for "loss of control" between North America and Europe.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

Table 3 Attitudes by respondent characteristics

Sample	Unethical (1–7)	Control (1–7)	Useful (1–7)	Avoid (1–7)
Role(s)				
Editor (economics journal)	4.69 (0.13)	5.41 (0.12)	3.15 (0.11)	5.68 (0.10)
Reviewer (non-editor)[a]	4.67 (0.09)	5.42 (0.07)	3.37 (0.07)	5.66 (0.06)
RePEc ranking				
Bottom 500	4.53 (0.12)	5.38 (0.10)	3.66 (0.10)	5.62 (0.09)
Middle 500	4.81 (0.12)	5.50 (0.10)	3.22 (0.11)	5.73 (0.08)
Top 500	4.69 (0.12)	5.43 (0.11)	3.15 (0.11)	5.67 (0.10)
Proportion of research using experiments				
Less than half	4.49 (0.18)	4.82 (0.18)	3.84 (0.15)	5.28 (0.15)
Half or more	4.71 (0.08)	5.54 (0.06)	3.25 (0.07)	5.74 (0.06)
Type of experiments				
Mostly field[b]	4.77 (0.11)	5.38 (0.09)	3.36 (0.09)	5.66 (0.08)
Mostly lab[b]	4.62 (0.11)	5.51 (0.09)	3.27 (0.10)	5.71 (0.08)
Position				
Graduate student/Postdoctoral	4.82 (0.18)	5.48 (0.16)	3.61 (0.17)	5.75 (0.12)
Assistant/Associate Professor	4.61 (0.10)	5.43 (0.08)	3.40 (0.08)	5.62 (0.07)
Full Professor	4.71 (0.12)	5.42 (0.11)	3.12 (0.11)	5.71 (0.09)
Professional status				
Researchers	4.68 (0.07)	5.44 (0.06)	3.33 (0.06)	5.67 (0.05)
Undergraduate students	2.75 (0.08)			3.56 (0.09)
Continent (Researchers)				
North America	4.31 (0.14)	5.19 (0.13)	3.38 (0.12)	5.47 (0.12)
Europe	4.88 (0.09)	5.57 (0.07)	3.26 (0.0)	5.79 (0.06)
Rest of the world	4.54 (0.21)	5.31 (0.18)	3.60 (0.19)	5.58 (0.15)

Mean ratings. Standard errors in parentheses. The sample is researchers, except for the entry "undergraduate students." Unethical, control, avoid, useful are measured on a 7-point scale, ranging from 1 "not at all" to 7 "very/extremely". See Online Appendix 1.2 for the exact questions. [a]Reviewers for economics journals. Excludes editors. [b]Mostly field (/lab) is defined as researchers for whom more than 50 percent of experiments are in the field (/lab)

middle/bottom third) and level of engagement with experiments. Here, we classify researchers as editors if they are (associate) editor at one or more economics journals. The reviewer category includes researchers reviewing for economics journals and excludes editors. These two categories combined make up 95 percent of the sample of researchers. There are almost no systematic differences in mean attitudes between reviewer roles, the respondents' ranking in RePEc, or their proportion of research using experiments. The relation between the need to avoid deception and the other variables is also similar across these subsamples. The main difference in attitudes is between researchers and students, a topic to which we return later.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

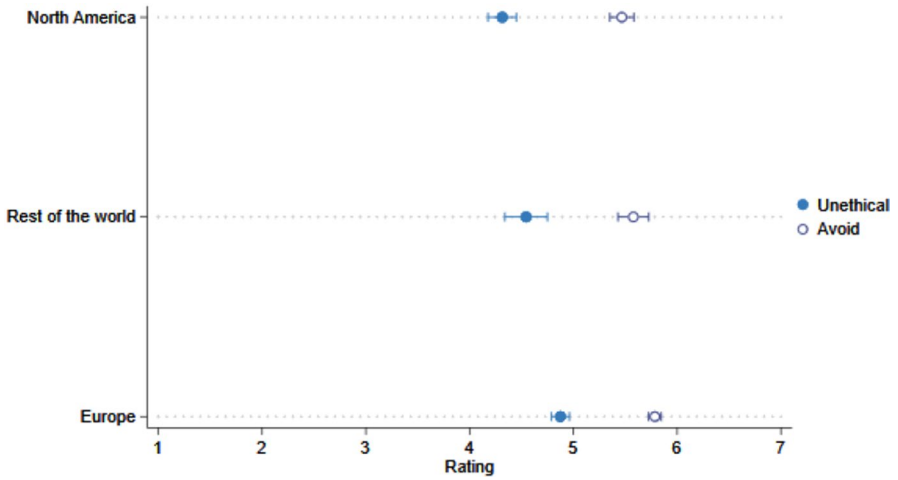


Fig. 3 Researcher attitudes about deception, by continent. This figure reports on researchers' attitudes about whether deception is unethical and whether it is important to avoid on a scale ranging from 1 ("not at all") to 7 ("very/extremely"). Error bars indicate +/− 1 s.e

Table 4 Summary statistics of valuations of scenarios by researchers

Scenario	Deceptive (1–7)	Negative (1–7)	Appropriate (1–7)	Useful (1–7)
Unexpected data use	3.18	2.94	5.19	4.96
	(0.07)	(0.07)	(0.06)	(0.06)
Subgroup re-match	3.20	3.01	5.00	4.64
	(0.07)	(0.08)	(0.07)	(0.07)
Unknown/unpaid participation	3.23	2.85	5.25	
	(0.08)	(0.07)	(0.07)	
Non-representative sample	3.76	3.42	4.76	4.40
	(0.07)	(0.07)	(0.07)	(0.07)
Surprise re-start	3.88	3.45	4.75	4.41
	(0.07)	(0.07)	(0.07)	(0.07)
Misinterpretation	4.78	4.58	3.70	
	(0.07)	(0.07)	(0.07)	
Confederates	5.33	4.79	3.88	4.07
	(0.07)	(0.07)	(0.07)	(0.07)
Total	3.91	3.58	4.65	4.50
	(0.03)	(0.03)	(0.03)	(0.03)

Mean ratings. Items are rated on a 7-point scale, ranging from 1 ("not at all") to 7 ("extremely"). Sample is researchers. See Table 1 for descriptions of the scenarios. Scenarios order by the mean rating of deceptiveness. Standard errors in parentheses. Due to a programming mistake, the survey did not display the "useful" question for two of the scenarios. Missing entries in column "useful" are due to this omission

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

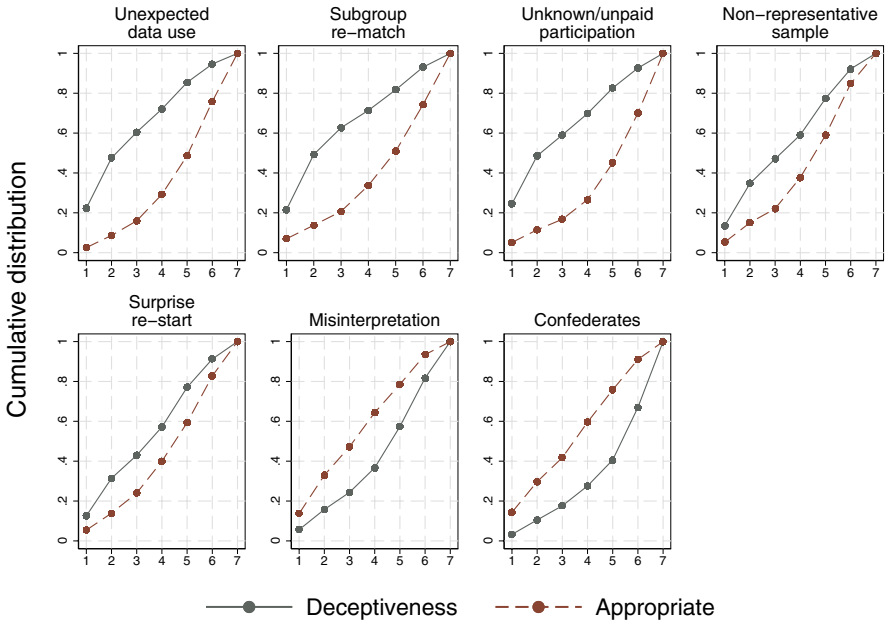


Fig. 4 Empirical CDFs of researchers' ratings of the different scenarios. Notes: Cumulative distributions of researchers' ratings of deceptiveness and appropriateness of the different scenarios

Turning to the scenario questions, we find variation in perceptions across scenarios. In what follows, we sort the scenarios by the researchers' mean rating of deceptiveness. Table 4 shows the mean ratings of the different scenarios. The use of confederates, and misinterpretation are considered to be the most deceptive techniques. Not informing subjects about their participation, unexpected data use, and matching in subgroups are considered to be the least deceptive techniques.[14] The ordering of scenarios is almost completely preserved if we look at how negatively researchers would react. The ordering is also almost completely preserved (but opposite in sign) for the other evaluations of the scenarios; how useful the form of deception is and how appropriate researchers find each scenario if the question is important and there are no alternatives at hand.

Figure 4 shows the empirical CDFs of how deceptive the different scenarios are considered to be and how appropriate each method is. Taken over all scenarios, the mean rating for appropriateness of deception when no alternatives are available is 4.65 (on a 7-point scale), again with heterogeneity. For instance, when asked about the use of a surprise-restart, 40 percent rate this as 4 or lower and 60 percent rate this as 5 or higher. The scenarios that are rated as most deceptive are considered to

[14] Mean ratings for unexpected data use, subgroup re-match, and unknown/unpaid participation are statistically indistinguishable from each other, as are the ratings for non-representative sample and surprise re-start (Wilcoxon signed rank test, $p > 0.260$ in all those cases). Ratings differ significantly between non-representative sample and unknown/unpaid participation ($p < 0.001$), misinterpretation and surprise re-start ($p < 0.001$), and confederates and misinterpretation ($p < 0.001$).

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

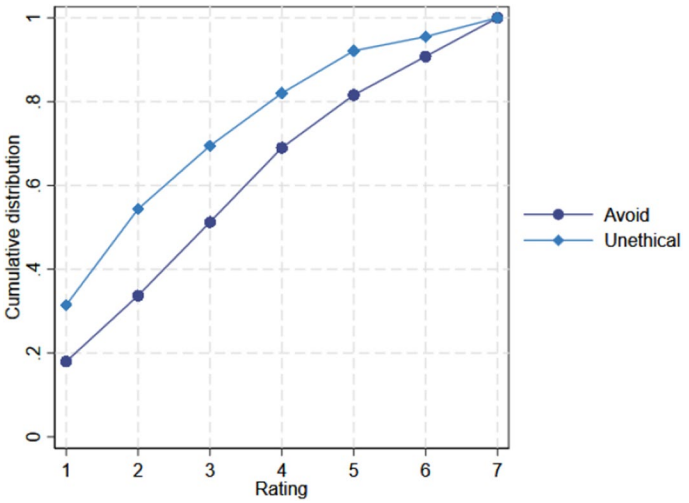


Fig. 5 Empirical CDF of students' ratings on "importance to avoid" and "how unethical." (7-point scale, ranging from "not at all" to "very/extremely")

be the least appropriate to use. For most scenarios, over 40 percent of respondents rate the appropriateness as 6 or 7. The exceptions are a misinterpretation and the use of confederates or bots, where about 20 percent of respondents rate the appropriateness as 6 or 7.

Result 1 Researchers are heterogeneous in how they view "gray-area" deception, with differences across regions and scenarios. The use of confederates (or bots) and misrepresentation of information are rated as the most deceptive of the scenarios.

4.2 Students

Figure 5 plots the empirical CDFs of students' attitudes. Students, like researchers in our sample, show heterogeneity in their attitudes towards deception. Compared to researchers, they perceive deception as being less unethical (mean rating of 2.75 vs 4.68, $p < 0.001$, two-sided Wilcoxon–Mann–Whitney test) and think it is less important to avoid (mean rating of 3.56 vs 5.67, $p < 0.001$, two-sided Wilcoxon–Mann–Whitney).[15] As with our researcher results, students from the US labs find deception less unethical than students from the UK lab (mean rating of 2.6 vs 3.4, $p < 0.001$, two-sided Wilcoxon–Mann–Whitney test).

Interestingly, students appear to be largely unaware of the existence of no-deception policies in economics labs. Only 27 percent of respondents indicate that they know about the no-deception policy in the lab from which they were recruited. Even

[15] In our first wave of the student survey (see footnote 9), the mean ratings for unethical and avoid are 2.88 and 3.52 respectively.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

Table 5 Summary statistics for valuations of scenarios by undergraduate students

Scenario	Deceptive (1–7)	Negative (1–7)	Appropriate (1–7)	Continue participation (1–7)	Affects answers (1–7)
Unexpected data use	3.72	3.04	4.99	4.81	4.45
	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Subgroup re-match	2.88	2.23	5.57	5.38	3.30
	(0.08)	(0.07)	(0.07)	(0.07)	(0.09)
Unknown/unpaid participation	4.53	3.83	4.67	4.17	4.43
	(0.09)	(0.09)	(0.08)	(0.09)	(0.09)
Non-representative sample	3.86	2.81	5.08	5.05	3.92
	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)
Surprise re-start	4.18	3.93	4.82	4.40	4.05
	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)
Misinterpretation	4.30	3.23	4.88	4.82	4.43
	(0.08)	(0.09)	(0.08)	(0.08)	(0.09)
Confederates	4.20	2.94	5.29	4.99	4.33
	(0.09)	(0.08)	(0.07)	(0.08)	(0.09)
Total	3.95	3.14	5.04	4.80	4.13
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)

Mean ratings. Items are rated on a 7-point scale, ranging from 1 ("not at all") to 7 ("extremely"). Sample is undergraduates. See Table 1 for descriptions of the scenarios. Standard errors in parentheses

excluding those (79) who had never participated in an experiment, this percentage is 31 percent. When asked how often they think their lab uses deception, 21 percent answered "never" (1) and 23 percent answered 5 or higher (on a 7-point scale, from "never" to "very often"). Experience does not seem to change this; there is only a small negative correlation between the number of times respondents participated in an experiment and how often they think deception is used (Spearman rank correlation $\rho = -0.090$, $p = 0.063$). Among those who participated in at least five experiments, 37 percent are aware of the no-deception policy.

We also detect a positive correlation between the number of times respondents participated in an experiment in other labs and how often they think deception is used in the surveyed lab (Spearman rank correlation $\rho = 0.11$, $p = 0.027$). While this cannot be given a causal interpretation, this may point to a negative externality from other labs. This result is intuitive if we assume that the other labs in which students participate sometimes use deception.

Is deception a problem? 25 percent of respondents reported that they would participate less often in that lab if they knew they were being deceived. The students' mean attitude towards the different scenarios is a very strong predictor of their self-reported likelihood of participating in the future. Using a scenario as the

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

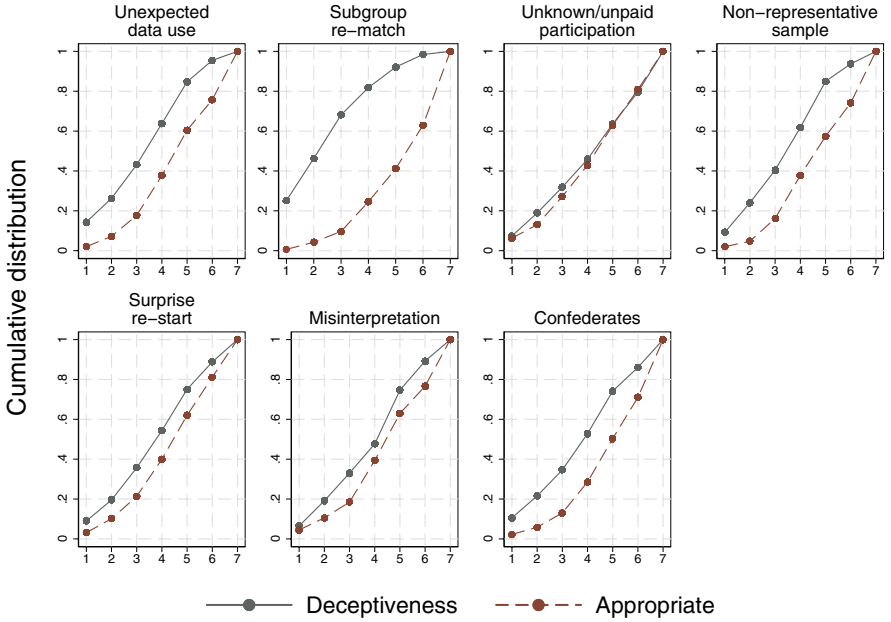


Fig. 6 Empirical CDFs of students' ratings of the different scenarios. *Notes*: Cumulative distributions of students' ratings of deceptiveness and appropriateness of the different scenarios

independent unit of observation, the correlation between a student's negative reaction towards a scenario and the likelihood of participating again is −0.93 (Spearman's rank correlation coefficient, $p = 0.003$). Of course, in addition to having only a modest number of observations, a caveat here is that there might be a selection problem; those who participated in our survey may have been those who are less disturbed by the thought that they may have been misled and this is a concern regarding data interpretation. The mean evaluations are reported in Table 5.

Figure 6 shows the empirical CDFs of ratings for the different scenarios. The mean rating for deceptiveness over all scenarios is 3.95. There is again heterogeneity in how deceptive scenarios are considered to be. Not knowing that they are part of an experiment and misinterpretation are considered the most deceptive techniques. Matching in subgroups is seen as the least deceptive technique.[16] Roughly the opposite ordering is observed if we look at the likelihood that a student would participate again.

[16] There are significant differences in mean ratings between subgroup re-match and unexpected data use (Wilcoxon signed rank test, $p < 0.001$), between non-representative sample and surprise re-start ($p = 0.001$), and between unknown/unpaid participation and misrepresentation ($p = 0.047$). Differences between unexpected data use and non-representative sample, and between surprise re-start, confederates, and misinterpretation are not statistically different ($p > 0.224$ in all those cases).

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

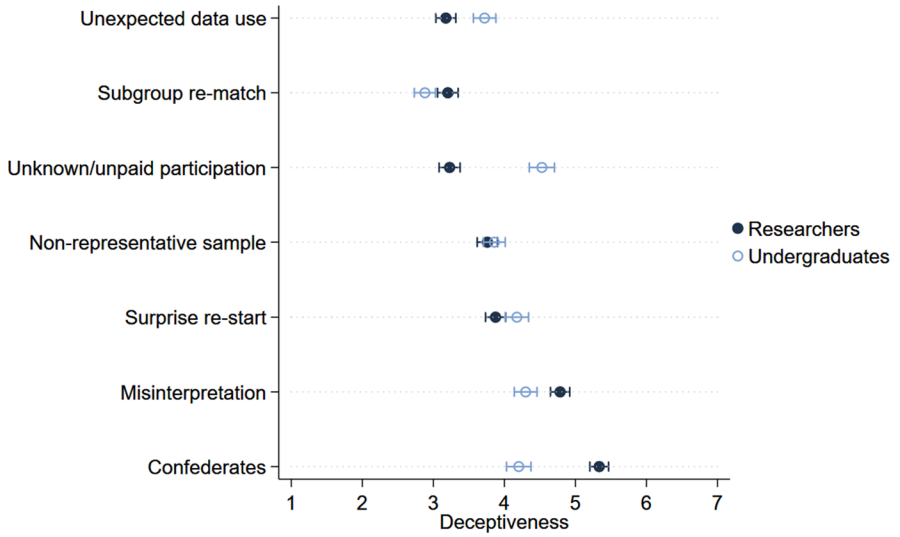


Fig. 7 Student attitudes about deception, by scenario. See Table 1 for descriptions of scenarios. Deceptiveness is rated on a 7-point scale where 1 is "not deceptive at all" and 7 is "extremely deceptive." Error bars indicate the 95% CI

Result 2 Students are heterogeneous in their attitudes towards deception. Some, but not all, indicate that they are less likely to continue participating after being deceived. Many students believe deception takes place in economics labs and the majority is unaware of the no-deception policies that are in place.

4.3 Differences between researchers and students

We next focus on some differences between researchers and students. We have already noted that students are on average more lenient towards deception. Yet interestingly, taken over all scenarios, the mean rating for deceptiveness among students (3.95) is very comparable to that of researchers (3.91). Nevertheless, the mean rating for the appropriateness of using deception is 5.04 (on a 7-point scale), somewhat higher than the rating by researchers (4.65). This difference is statistically significant ($p < 0.001$, two-sided Wilcoxon–Mann–Whitney test).

Figure 7 shows some other interesting patterns appear when we compare attitudes across the different scenarios. Students have very different views than researchers on some techniques. Two techniques stand out in particular in this respect. Compared to researchers, students see unknown/unpaid participation as more deceptive (mean rating of 3.2 vs 4.5), and the difference is significant ($p < 0.001$, two-sided Wilcoxon–Mann–Whitney test). The opposite is true for the use of confederates or bots; generally considered as more deceptive by researchers than by students (mean rating of 5.3 vs 4.2, $p < 0.001$). Students also have a more negative attitude toward

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

unknown/unpaid participation (3.8 vs 2.9, $p<0.001$) and are less negative towards the use of confederates (2.9 vs 4.8, $p<0.001$). At least visually, the consistency in attitudes across researchers and students appears weak. Statistically we do not find a significant relationship with regards to how unethical a method is (Spearman's rank correlation 0.57, $p=0.180$) nor how negatively a method is perceived to be ($\rho=-0.04$, $p=0.939$), but this is based on a very small number of independent observations (7).[17]

Krawczyk (2019) also found that students are more forgiving than researchers towards the use of confederates, but did not include a scenario about unknown/unpaid participation. Furthermore, he also found that both students and researchers are relatively lenient towards the use of subgroup re-matching. Unlike us, respondents in his survey rate a non-representative sample as relatively deceptive (rated around 4.8 after rescaling to a 7-point scale). He concluded that there is a high consistency of deceptiveness ratings between students and researchers.[18]

What can account for the differences in reported attitudes between researchers and students? Researchers and students may have genuinely different attitudes. Some students may be less concerned about some techniques because they care mostly about being paid and any negative externalities are not an issue to them. It is also possible that these two groups interpret the scale or questions differently. Researchers may be better aware of the implementation and use of several techniques, something that we cannot exclude.[19] Indeed, the question "Do you feel that it is unethical for experimenters to deceive participants in their experiments (even after debriefing)?" may have been interpreted by researchers as meaning that the deception would continue to occur after debriefing subjects. The wording of the question for students was less ambiguous. This could help to explain the difference between researchers and students.[20]

Result 3 Compared to researchers, students appear to be more lenient towards deception in general and find it less important to avoid it. There are, however, some important differences in their attitudes towards some standard methods; students appear more concerned about not knowing that they are taking part in an experiment, and less concerned about the use of confederates or bots.

[17] Figure A2.1 in Online Appendix 2 replicates Fig. 7 using the sample of students that saw the first version of our survey (see also footnote 9). The main difference is that students in that sample rate the re-start scenario as more deceptive. The first version of the survey may have prompted respondents to believe that they had to stay longer in the lab than anticipated. We believe that this could explain the students' attitude toward this scenario, and therefore made clear in the second version of the survey that the experiment would not take longer than anticipated.

[18] Krawczyk (2019) reports a correlation coefficient of 0.86. The Spearman's rank correlation coefficient of deceptiveness ratings between scenarios is 0.627, with a p-value of 0.039 (own calculations based on his Table 1).

[19] One option would be to provide respondents with some specific examples of each technique. We did not opt for this because we did not want to prime respondents on specific examples.

[20] We thank a referee for pointing this out.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

Table 6 Alternatives to various techniques

Scenario	Possible alternatives (frequency mentioned)
Subgroup re-matching	Collect larger sample (29) Use vague wording/omission (27)
Surprise re-start	Multipart design with new instructions at the start of a part (78) offer exit option with payment before the re-start (15) probabilistic re-start (22) Leave number of rounds unspecified (13)
Non-representative sample	Subsamples (33) Alternative wording (10) Strategy method (7)
Unexpected data use	Use vague wording/omission (11) Probabilistic use (6) Unspecified use (4) Offer exit option (3)
Confederates	Conditional Information Lottery method (28) Collect larger sample (23) Use real script of previous subject (7) Strategy method (5)
Unknown/unpaid participation	Keep cost to participant low (4) Compensate participant (4) Donate wage-equivalent to charity (1)
Misrepresentation	Compound lottery (2) Strategy method (1)

See Table 1 for a description of the scenarios. Based on the free-form comments provided by respondents (researchers). The number in parentheses is the number of researchers mentioning the alternative

5 Alternatives

We asked researchers to give alternatives for each of the techniques described in the respective scenarios. Table 6 lists the most frequently mentioned alternatives and conditions under which the technique can be used. This is based on our own reading of the comments, and not meant as a systematic analysis. Nor should it be taken as an indication that we endorse the proposed alternatives. The suitability of any alternative depends on the research question.

5.1 Subgroup re-matching

A number of respondents propose being vague about the way that subjects are re-matched ("at the start of each round, you will be re-matched"). They believe this technique is fine as long as the researcher does not outright lie about it, on the grounds that omitting information is not very deceptive. The most often-mentioned alternative is to collect a larger sample. This is best, of course, but may not be feasible or affordable. In general, our data indicate that subgroup re-matching does not seem to be a big problem for subjects.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

5.2 Surprise re-start

The most often proposed alternative is to have a design with several parts, with new instructions at the start of a part. The instructions for the second part can simply be "you will play another 10 rounds like this." Some others propose to give subjects the option to leave after the first part, without losing any payments. Other alternatives mentioned are to tell subjects that there is a probability that the experiment will continue after the first part, or to leave the number of rounds unspecified. The point is to avoid being misleading or vague in the instructions and experimental protocol unless it serves a clear purpose.

5.3 Non-representative sample

Drawing random subsamples is a commonly-proposed alternative. Some propose to use a different wording (telling subjects that "a group of other subjects choose X" instead of "on average subjects choose X") or to use the strategy method. Wording choices are a bit delicate and not everyone will agree. The strategy method is in fact not without controversy, but any treatment effect found to date with the strategy method has also been present using the direct-response method (see Brandts & Charness, 2011).

5.4 Unexpected data use

Some respondents indicate that they do not perceive this as deceptive as long as the researcher does not explicitly lie about this. In their view, omitting information about the use of data in a later part of the experiment is not very deceptive. Some propose telling subjects that there is a probability that their choices can be used in later parts. Another method is to tell subjects that their choices will be used later, but not specify in what way (so that subjects cannot "game" the system by distorting their decisions). Some respondents suggest offering subjects the opportunity to leave the experiment without any loss of payment.

5.5 Confederates or computer players

The "Conditional Information Lottery" method proposed by Bardsley (2000) is often mentioned. Bardsley's proposal is to camouflage a true task amongst some other fictional tasks. Subjects are informed that some of their opponents or the information they receive is fictitious, but are not told in which rounds this is the case. Only the real task is payoff relevant. Some respondents propose collecting a larger sample. Other alternatives are to use the strategy method, or to have a confederate or computer player use the script/decisions of a real subject from a previous session.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

5.6 Unknown/unpaid participation

For this technique, not many alternatives are mentioned. Some respondents suggest keeping the cost low, i.e., not consuming too much of a participant's time. Some others suggest compensating participants afterwards. One respondent suggests donating the equivalent of the participant's wage to a charity.

5.7 Misrepresentation

Again, not that many alternatives are mentioned for this technique. One respondent suggests using the strategy method. A few respondents suggest using a compound lottery. For instance, in the case of two events, one can implement the two events with probabilities 0.25 and 0.75 for some subjects, and with probabilities 0.75 and 0.25 for other subjects, so that on average the probability of each event is 0.5.

6 Discussion

Our findings inform researchers in experimental economics about attitudes towards "gray-area" deception, such as the omission of relevant information or misleading language. The attitudes towards such practices and views about whether such gray-area deception is unethical are heterogeneous. Perhaps even more interesting is that few student respondents were aware of the no-deception policy in their laboratory. Students appear less concerned about deception, except where it directly affects payment and the time spent in the experiment. However, the selection issue raised by Cason and Wu (2018) is well taken, since we may be observing a biased sample, and it is also possible that students interpret the questions and/or scale differently.

Our findings lead us to several prescriptions. First, if we value and wish to have experimental control, it would be useful to ensure that subjects are aware of the no-deception policy and what it means. Second, costs and benefits do matter. For example, there is support in the experimental community for gray-area deception when the data are important, particularly if this will not greatly impact any subject pool in the future. This calls for a careful lab policy, since it should be clear to participants what they can expect in terms of the use of deception. A lab policy that hinges on cost/benefits would be confusing for students to interpret. Hence it may be more appropriate to maintain a clear no-deception policy in labs where studies are run frequently. Studies using gray-area deception when it cannot be avoided could be run outside of those labs, minimizing the risk of contaminating the subject pool for future experiments. Third, we suggest that reviewers recognize that their views are not shared universally. To avoid researchers getting exposed to idiosyncrasies of reviewers' opinions, journals could be more explicit about practices that are considered deceptive. Fourth, more research on students' attitudes and behavioral reactions would be welcome, as their views are not entirely aligned with those of researchers.

We hope that our results lead to a great deal of discussion and perhaps some convergence on the issue of gray-area deception. While we are certainly not arguing for

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

telling outright lies to the subjects (although Krasnow et al., 2018, suggest that even this is not harmful), we would certainly welcome a more nuanced view. Our study by no means covers all experimental techniques. To give just one example: we sent out our survey to three different groups (based on their *h*-index) which we then used in our analysis (without telling our respondents). This is akin to, e.g., Niederle and Vesterlund (2007) who form groups based on gender but do not let the participants know that this is one of the variables that is being studied. We believe this practice can be justified on the grounds that anonymity is preserved, but perhaps other researchers feel differently about this. More studies and discussion on such points are most welcome.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10683-021-09726-7.

Acknowledgements We acknowledge valuable comments from Tim Cason, David Cooper, Erik Eyster, Guillaume Fréchette, Alexandros Karakostas, Andreas Ortmann, and Bertil Tungodden. We also acknowledge Charles Noussair and Karim Sadrieh for allowing us to contact former students in their subject pools. We owe considerable thanks to Tarush Gupta, whose technical abilities made this survey possible, and the students at the Behavioral and Experimental Economics research group who collected contact data for the researchers we surveyed. Last but not least, we thank everyone who took the time to complete our survey.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Bardsley, N. (2000). Control without deception: Individual behaviour in free-riding experiments revisited. *Experimental Economics, 3*(3), 215–240.

Barrera, D., & Simpson, B. (2012). Much ado about deception: Consequences of deceiving research participants in the social sciences. *Sociological Methods & Research, 41*(3), 383–413.

Bonetti, S. (1998). Experimental economics and deception. *Journal of Economic Psychology, 19*(3), 377–395.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: A first survey of experimental comparisons. *Experimental Economics, 14*, 375–398.

Cason, T. N., & Wu, S. Y. (2018). Subject pools and deception in agricultural and resource economics experiments. *Environmental and Resource Economics, 73*, 1–16.

Colson, G., Corrigan, J. R., Grebitus, C., Loureiro, M. L., & Rousu, M. C. (2015). Which deceptive practices, if any, should be allowed in experimental economics research? Results from surveys of applied experimental economists and students. *American Journal of Agricultural Economics, 98*(2), 610–621.

Cooper, D. J. (2014). A note on deception in economic experiments. *Journal of Wine Economics, 9*(2), 111–114.

Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton University Press.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press

Frohlich, N., Oppenheimer, J., & Moore, B. (2001). Some doubts about measuring self-interest using dictator experiments: The costs of anonymity. *Journal of Economic Behavior and Organization, 9*(2), 271–290.
Gächter, S. (2009). Improvements and future challenges for the research infrastructure in the field 'experimental economics'. RatSWD working paper No. 56.
Hersch, G. (2015). Experimental economics' inconsistent ban on deception. *Studies in History and Philosophy of Science Part A, 52*, 13–19.
Hertwig, R., & Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior, 18*(1), 59–92.
Hey, J. D. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology, 19*(3), 397–401.
Jamison, J., Karlan, D., & Schechter, L. (2008). To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments. *Journal of Economic Behavior & Organization, 68*(3–4), 477–488.
Krasnow, M. M., Howard, R. M., & Eisenbruch, A. B. (2019). The importance of being honest? Evidence that deception may not pollute social science subject pools after all. *Behavior Research Methods, 52*, 1–14.
Krawczyk, M. (2015). Trust me, I am an economist. A note on suspiciousness in laboratory experiments. *Journal of Behavioral and Experimental Economics, 55*, 103–107.
Krawczyk, M. (2019). What should be regarded as deception in experimental economics? Evidence from a survey of researchers and subjects. *Journal of Behavioral and Experimental Economics, 79*, 110–118.
Ledyard, J. O. (1995). Public goods: A survey of experimental research. In A. Roth & J. Kagel (Eds.), *Handbook of experimental economics.* Princeton University Press.
Lusk, J. L. (2019). The costs and benefits of deception in economic experiments. *Food Policy, 83*, 2–4.
McDaniel, T., & Starmer, C. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology, 19*(3), 403–409.
Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *The Quarterly Journal of Economics, 122*(3), 1067–1101.
Ortmann, A. (2019). Deception. In *Handbook of research methods and applications in experimental economics*. Edward Elgar Publishing.
Ortmann, A., & Hertwig, R. (2002). The costs of deception: Evidence from psychology. *Experimental Economics, 5*(2), 111–131.
Rousu, M. C., Colson, G., Corrigan, J. R., Grebitus, C., & Loureiro, M. L. (2015). Deception in experiments: Towards guidelines on use in applied economics research. *Applied Economic Perspectives and Policy, 37*(3), 524–536.
Svorenčík, A. (2016). The Sidney Siegel tradition: The divergence of behavioral and experimental economics at the end of the 1980s. *History of Political Economy, 48*(1), 270–294.
Wilson, B. J. (2016). The meaning of deceive in experimental economic science. In *The Oxford handbook of professional economic ethics.* Oxford University Press.
Wilson, R. K., & Isaac, R. M. (2007). Political economy and experiments. In *The political economist: Newsletter of the section on political economy*.
Zultan, R. (2015). Do participants believe the experimenter? Unpublished manuscript.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

https://doi.org/10.1007/s10683-021-09726-7 Published online by Cambridge University Press