CIRANO
Allier savoir et décision

Persuasion Bias in Science: An Experiment on Strategic Sample Selection

ARIANNA **DEGAN**
MING **LI**
HUAN **XIE**

2019S-24
CAHIER SCIENTIFIQUE
CS

2019s-24

Persuasion Bias in Science: An Experiment on Strategic Sample Selection

Arianna Degan, Ming Li, Huan Xie

Série Scientifique
Scientific Series

Montréal
Octobre/October 2019

Centre interuniversitaire de recherche en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du gouvernement du Québec, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Quebec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the government of Quebec, and grants and research mandates obtained by its research teams.

Les partenaires du CIRANO

Partenaires corporatifs
Autorité des marchés financiers
Banque de développement du Canada
Banque du Canada
Banque Laurentienne
Banque Nationale du Canada
Bell Canada
BMO Groupe financier
Caisse de dépôt et placement du Québec
Canada Manuvie
Énergir
Hydro-Québec
Innovation, Sciences et Développement économique Canada
Intact Corporation Financière
Investissements PSP
Ministère de l'Économie, de la Science et de l'Innovation
Ministère des Finances du Québec
Mouvement Desjardins
Power Corporation du Canada
Rio Tinto
Ville de Montréal

Partenaires universitaires
École de technologie supérieure
École nationale d'administration publique
HEC Montréal
Institut national de la recherche scientifique
Polytechnique Montréal
Université Concordia
Université de Montréal
Université de Sherbrooke
Université du Québec
Université du Québec à Montréal
Université Laval
Université McGill

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

ISSN 2292-0838 (en ligne)

Persuasion Bias in Science: An Experiment on Strategic Sample Selection [*]

Arianna Degan [†], Ming Li [‡], Huan Xie [§]

Abstract/Résumé

We experimentally test a game theoretical model of researcher-evaluator interaction à la Di Tillio, Ottaviani, and Sørensen (2017a). Researcher may strategically manipulate sample selection using his private information in order to achieve favourable research outcomes and thereby obtain approval from Evaluator. Our experimental results confirm the theoretical predictions for Researcher's behaviour but find significant deviations from them about Evaluator's behaviour. However, comparative statics are mostly consistent with the theoretical predictions. In the welfare analysis, we find that Researcher always benefits from the possibility of manipulation, in contrast to the theoretical prediction that he sometimes is hurt by it. Consistent with theoretical predictions, Evaluator benefits from the possibility of Researcher's manipulation when she leans towards approval or is approximately neutral but is hurt by that possibility when she leans against approval.

Keywords/Mots-clés: Persuasion Bias, Research Conduct, Manipulation, Sample Selection, Experiment, Randomized Controlled Trials

JEL Codes/Codes JEL: C72, C92, D83

[*] We thank audiences at Concordia University, Inner Mongolia University, Shanghai University of Finance and Economics, University of Pittsburgh, University of Saskatchewan, 3rd Annual Workshop in Behavioural and Experimental Economics at Southwestern University of Finance and Economics 2017, CPEG 2019 Waterloo, Guanghua Lecture Series, Workshop on Theoretical Economics at South-western University of Finance and Economics 2017, International Conference on Economic Theory and Applications, Southwestern University of Finance and Economics 2018, North-American ESA Conference Antigua Guatemala 2018, and Workshop in Communication and Persuasion Montreal 2017. We gratefully acknowledge funding from SSHRC Insight Development Grant 430-2016-00444, "Scientific research, conflicts of interest, and disclosure policy." Noémie Cabau and Binyan Pu provided excellent research assistance. Any remaining errors are our own.
[†] Université du Québec à Montréal and CDER.
[‡] Concordia University, CIREQ, and CIRANO.
[§] Concordia University, CIREQ, and CIRANO.

1 Introduction

Recently, across the scientific community, there has been heightened concerns about ensuring the credibility and ethics of scientific research. This is in part prompted by investigative findings that spotlight problematic past and present practices pertaining to research methods and conflicts of interest.[1] In the effort to maintain and enhance credibility of scientific research, economics has the potential to make a valuable contribution, especially from the perspectives of incentive, information, and strategic interaction.

Conflicts of interest and career concerns are two important factors motivating researchers to manipulate data collection during the experiment and misrepresent the results and conclusions drawn from the experiments. For instance, in drug trials, a researcher's financial payoff is often dependent on whether or not the pharmaceutical company's drug gets approved. More generally, typical academic researchers would like to produce positively significant findings, so that their research is deemed publishable in an academic journal, a phenomenon often termed "publication bias." Even though much has been written on the nature of such deviation from the ideal ethical behaviour of a truth-finding scientist,[2] Di Tillio, Ottaviani, and Sørensen (2017a) are the first to use a game theoretic model to quantify the effects of manipulations in randomized controlled trials – how such manipulations affects the quality and usefulness of scientific research. In this paper, we report results from an experiment that closely models after their setup, and provide both optimistic and pessimistic answers to their central question, namely, how such manipulations affect the evaluator's and the researcher's expected payoffs.

Di Tillio, Ottaviani, and Sørensen (2017a) focus on randomized controlled trials

[1] In a New York Times article, O'Connor (2016) reported the finding by Kearns, Schmidt, and Glantz (2016) that the sugar industry's nutrition science funding diverted blame for heart diseases from sugar to fat. Separately, failure to replicate results in some academic studies raised serious questions about how journals and the whole of academia should safeguard against unreliable research findings (Chang and Li 2017, Open Science Collaboration, 2015). Finally, studies have found evidence of researchers manipulating data and methodology to obtain statistically significant results (Brodeur, Lé, Sangnier, and Zylberberg 2016, Head, Holman, Lanfear, Kahn, and Jennions 2015, Simonsohn, Nelson, and Simmons 2014). Such concerns, though serious, are distinct from other more egregious violations of ethics of scientific research (Jump 2011, Kolata 2018).

[2] Glaeser (2008) offers one such discussion, who takes the view that researchers' choosing variables to maximize significance should not be viewed as "a great sin" and nor will it "magically disappear." Instead, he suggests that we should embrace what he calls the researcher initiative and in the meantime deal with some of its negative effects uisng a variety of measures: just skepticism, subsidization of data creation and transmission, and development of counteracting econometric techniques, citing the work of Leamer (1983) as a promising example of the last approach.

(RCT), which is the gold standard procedure to estimate the causal effect of an intervention. They offer a concise introduction to the history of RCT, dating back to biblical times and concluding with its modern adoption and efforts to improve its validity. In particular, much effort has been made to preserve true randomization of sample selection and avoid researcher manipulation.[3]

Di Tillio, Ottaviani, and Sørensen (2017a) consider a model in which a researcher designs a randomized controlled trial to convince an evaluator to give approval of his finding. The evaluator must incur a cost to approve the finding and is only willing to do so if the finding is significant enough. This cost can also be interpreted as the level of skepticism of the evaluator or the relative desirability of a status quo option. The researcher observes private information that enables him to vary how the trial is run. Even though the evaluator does not observe the researcher's private information, she is aware of the ability of the researcher to manipulate the trial using such private information. They consider three types of possible manipulations by the researcher in conducting the RCT: selective sampling of subjects, selective assignment of subjects into control and treatment groups, and selective reporting of experiment results. They find that whether or not manipulation by a researcher would benefit the researcher and/or the evaluator depends on the evaluator's cost of acceptance. Therefore, manipulation is not necessarily detrimental to the value of scientific research.

In this paper, we experimentally test a simplified version of Di Tillio, Ottaviani, and Sørensen's (2017a) model, where the researcher, after observing private information about one of two sites, chooses one of them to run the trial.[4] That is, our experiment is exclusively focused on selective sampling. We follow Neyman (1923) and Rubin (1974), who respectively pioneered and substantively developed the potential outcomes approach adopted by Di Tillio, Ottaviani, and Sørensen (2017a). However, we deal with an experimental setup where there is only uncertainty about the treatment outcome but no uncertainty about the baseline outcome.

Selective sampling challenges the external validity of an experiment, in that the outcome of the experiment is not a reliable predictor of the average treatment effect of the population. As cited by Di Tillio, Ottaviani, and Sørensen (2017a), Allcott (2015) shows that in the Opower energy conservation program, the initial 10 sites chosen by the sponsoring company for the RCTs significantly overstate the average treatment

[3]See Imbens and Rubin (2015) and Rosenberger and Lachin (2015) for a more detailed discussion.

[4]Our experimental setup is also related to the theoretical framework of Hoffmann, Inderst, and Ottaviani (2014), applied to a different context.

effect for the next 101 sites. Thus, even with a large sample population, the outcomes of the 10 initial trial sites are a poor indicator of the average treatment effect of the policy on the general population.

We design our experiment to include environments both with and without the possibility of manipulation and under various acceptance costs of the evaluator. We find that *researchers' behaviour* is largely consistent with the equilibrium strategy as theory predicts, but there is significant difference between *evaluators' behaviour* and the theoretical predictions. Nevertheless, comparing different treatments, the comparative statics about evaluators' behaviour largely bear out the theoretical predictions. Finally, we conduct a welfare analysis and contrast our findings with the theoretical predictions of Di Tillio, Ottaviani, and Sørensen (2017a). Our analysis is based on the subjects' frequency of choices and the prior probabilities of our model setup, in order to have a fair comparison across treatments.[5] We find that the researcher always benefits from manipulation, which is in contrast to the theoretical prediction that he only benefits from manipulation from intermediate levels of acceptance cost and is hurt by manipulation for low and moderately high levels of acceptance cost.[6] The evaluator is better off under manipulation for low and intermediate levels of acceptance cost but worse off under moderately high levels of acceptance cost, which is mostly consistent with theoretical predictions.

Other recent theoretical studies on scientific methods include those by Di Tillio, Ottaviani, and Sørensen (2017b), Min (2017), and Yoder (2016). Di Tillio, Ottaviani, and Sørensen (2017b) consider a more general framework of strategic sample selection by a researcher. To the best of our knowledge, our paper is the first experimental study on this topic.

Our paper is also related to the experimental literature on the economics of strategic information transmission, which is based on the theoretical model initiated by Crawford and Sobel (1982). Blume, Lai, and Lim (2017) provide an excellent survey of the literature. One recent paper that is somewhat related to ours is by Chung and Harbaugh (2016), who evaluate whether disclosing bias of an expert would improve or worsen the advice he provides to a decision maker. There has also been recent literature on testing the Bayesian persuasion model of Kamenica and Gentzkow (2011) in the laboratory,including papers by Au and Li (2018), Fréchette, Lizzeri, and Perego (2017)

[5]Because the realizations of random events are different in each session, the actual payoff is not a good index for welfare comparison. Please refer to Section 5 for detailed explanations.

[6]When acceptance cost is extremely high, the evaluator would never approve the researcher's finding.

and Nguyen (2017).[7] Fréchette, Lizzeri, and Perego (2017) design an experiment in which a sender commits to a disclosure strategy before learning his private information and with a certain probability gets to revise it after his private information is realized.[8] Thus, they are able to encompass cheap talk, full Bayesian persuasion, as well as partial commitment in their experiment. They show that the sender does benefit from commitment, as theory predicts. However, they find that whether information is verifiable matters for the informativeness of the sender's messages even under Bayesian persuasion, which differs from theoretical predictions. This phenomenon appears to be caused by the fact that the sender is unable to take advantage of silence or non-disclosure as an informative message when information is verifiable. In addition, informativeness decreases as commitment increases when information is verifiable and the opposite is true when information is unverifiable. This is again broadly consistent with theoretical predictions. While Fréchette, Lizzeri, and Perego (2017) instruct subjects to directly choose an explicit strategy, Au and Li (2018) and Nguyen (2017) take the traditional approach but design a game that is relatively easy for subjects to understand and play. Nguyen's (2017) design is closer to the canonical model of Bayesian persuasion with a limited space of signals. She shows that as subjects accumulate experience, their play approaches the theoretical predictions. Au and Li (2018), in contrast, allow the receiver to learn the posteriors directly, so it simplifies the receiver's inference problem. They focus on whether receiver's response demonstrates evidence of reciprocity concerns. A sender is punished for behaving in an "unfair" way, or tries to extract too much surplus.

The rest of the paper proceeds as follows. In Section 2, we describe the model and discuss the equilibrium and welfare when the sender does or does not posess private information. In Section 3, we describe the design of the experiment. In Sections 4 and 5, we present the results on the researcher's and the evaluator's behaviours, and on their welfare, respectively. In Section 6, we conclude.

[7] As Di Tillio, Ottaviani, and Sørensen (2017a) clearly articulate in their Conclusion section, their theoretical model is different from the theoretical framework of Bayesian persuasion (Kamenica and Gentzkow 2011, Rayo and Segal 2010, Kolotilin, Mylovanov, Zapechelnyuk, and Li 2017). Similarly, our experimental setup is different from experiments on Bayesian persuasion. In particular, Bayesian persuasion would allow Researcher to fully commit to a strategy, including one in which he reveals nothing to Evaluator, while our experiment does not give Researcher that option.

[8] See also Jin, Luca, and Martin (2015) for an experiment that tests disclosure of verifiable information. They show that the receiver does not always view the nondisclosure of information unfavourably towards the sender. As a result, a sender with bad information may benefit from hiding it.

2 Model

Our model is a simplified version of the selective sampling model of Di Tillio, Ottaviani, and Sørensen (2017a). There are two players: Researcher and Evaluator.[9] Researcher conducts an experiment to test the effectiveness of a drug or a policy. There are two sites with the same population size. Researcher's goal is to convince Evaluator that the average treatment effect in the whole population is sufficiently high so that Evaluator is willing to grant acceptance of the drug or the policy. Following Di Tillio, Ottaviani, and Sørensen (2017a), we simplify Researcher's experiment to one on a single subject. Researcher's choice reduces to that between two sites on which to run the experiment. Before making his choice, he has the potential to observe private information about the treatment effect of one of the sites. Researcher's use of such private information in his choice of the experimental site, or selective sampling, challenges the external validity of the experiment. In this section, we present a concise analysis of the effect of such manipulation along the lines of Di Tillio, Ottaviani, and Sørensen (2017a), assuming that Evaluator is fully rational.

Now, we introduce our formal theoretical model, closely following that of Di Tillio, Ottaviani, and Sørensen (2017a) with simplifications. Let $t \in \{L, R\}$ denote a site, with the same population size on each site. Assume that the treatment effect of each individual is homogeneous within each site, which we denote by β_t, $t \in \{L, R\}$. The average treatment effect on the whole population is therefore

$$\beta_{ATE} \quad = \quad \frac{\beta_L + \beta_R}{2}.$$

Assume that treatment effects are iid in each site and follow the Bernoulli distribution, where $\beta_t = 1$ ($t \in \{L, R\}$) with probability q and $\beta_t = 0$ with probability $1 - q$.

Researcher always prefers that his request be accepted – his payoff is 1 if Evaluator accepts his request and 0 if Evaluator rejects it. Evaluator's payoff from acceptance is equal to the difference between the average treatment effect and an acceptance cost, $\beta_{ATE} - k$, and that from rejection is normalized to 0. The cost of acceptance can be interpreted as an opportunity cost – in the case of a new drug, the effectiveness of the currently available alternative, or a real cost – in the case of a new policy project, the monetary cost of funding it.

Researcher chooses a site to conduct an experiment, the outcome of which is then

[9] To avoid confusion, we will use male pronouns for Researcher and female pronouns for Evaluator.

observed by Evaluator. We also refer to the outcome of the experiment as evidence, $v \in \{0, 1\}$. We assume that evidence is precisely equal to the treatment effect of Researcher's chosen experimental site: $v = 0$ if $\beta_t = 0$ and $v = 1$ if $\beta_t = 1$. It follows that Evaluator's best response given experimental evidence v is to accept Researcher's request if and only if $E(\beta_{ATE}|v) \geq k$, where the expectation is based on Evaluator's belief about the treatment effect on the site not chosen by Researcher after observing the experimental evidence. A conflict of interest emerges between Researcher and Evaluator because under complete information there exist instances where Evaluator finds it optimal to not accept Researcher's request.

In our theoretical analysis below, following Di Tillio, Ottaviani, and Sørensen (2017a), we also assume that Evaluator observes neither the site where the experiment is conducted nor the site about which Researcher has obtained private information. Thus, in forming an expectation about the average treatment effect, Evaluator only has the experimental evidence available.[10]

We consider two environments: No Manipulation and Manipulation. In the Manipulation environment, the timing of the game is as follows:

1. Researcher and Evaluator observe Evaluator's cost of acceptance, k ;
2. Researcher receives a private message from nature, $\beta_I \in \{\beta_L, \beta_R\}$, which reveals the *true* treatment effect on site $I \in \{L, R\}$;
3. Researcher chooses one site $t \in \{L, R\}$ to conduct the experiment;
4. Both Researcher and Evaluator observe the experimental evidence v;
5. Evaluator chooses whether to accept or reject Researcher's request.

In the No-Manipulation environment, everything is the same as in the Manipulation environment except for the absence of Step 2. So, when Researcher chooses the experimental site, he does not have any private information about the treatment effect in any site.

[10]This assumption is made to simplify the analysis. If Evaluator does not observe from which site Researcher has obtained private information, then as long as Researcher obtains private information from each site with equal probability, whether or not Evaluator observes the experimental site is inconsequential and does not affect the analysis. Even if Evaluator does observe from which site Researcher has obtained private information, as long as she does not observe the experimental site, the analysis is not affected.

In the two following subsections, we consider the characterization of the pure-strategy equilibrium in each of the two environments. Since Evaluator's optimal strategy consists of acceptance if and only if her expectation of the average treatment effect (weakly) exceeds her acceptance cost k, to characterize Evaluator's optimal strategy it is enough to determine $E(\beta_{ATE}|v)$ in the different environments, under different assumptions of her rationality.

2.1 No-Manipulation Case

First, we consider the No-Manipulation case, where Researcher has no private information. The characterization of equilibria is straightforward. Note that Evaluator does not observe the experimental site. Thus, Researcher's decision has no effect on his expected payoff regardless of Evaluator's strategy; he may randomly choose between Left and Right site to conduct the experiment.[11] Evaluator forms expectations about the average treatment effect conditional on evidence v from the experimental site. Given that Researcher does not observe any private information, Evaluator's belief about the treatment effect on the site not chosen is the same as her prior belief about it. Therefore,

$$E(\beta_{ATE}|v=0) = \frac{q}{2}$$

and

$$E(\beta_{ATE}|v=1) = \frac{1+q}{2}.$$

2.2 Manipulation Case

In the case where Researcher receives private information prior to choosing the experiment site, we focus on the equilibrium in which Researcher plays the *Intuitive Strategy* proposed by Di Tillio, Ottaviani, and Sørensen (2017a):

- If $\beta_I = 1$, then conduct the experiment on site $t = I$;
- If $\beta_I = 0$, then conduct the experiment on site $t = -I$,

where I and $-I$ are respectively the sites about which Researcher does and does not have private information. The *Intuitive Strategy* simply states that if the private information reveals a positive treatment effect of a site, then Researcher should conduct

[11] Any probability in $[0, 1]$ of choosing site L can be part of an equilibrium.

the experiment in the observed site. Otherwise, if the private information reveals zero treatment effect, then Researcher should switch to the other site to conduct the experiment.Given Researcher's strategy, Evaluator forms expectations about the average treatment effect conditional on the experimental evidence v:

$$E(\beta_{ATE}|v=0)=0$$

and

$$E(\beta_{ATE}|v=1)=\frac{1}{2-q}.$$

When $v=0$, if Evaluator is fully rational, she will be able to deduce that Researcher must have received private information that the treatment effect in the observed site is zero and consequently has chosen to conduct the experiment on the other site, $\beta_I=0$ and $t=-I$. Therefore, the average treatment effect is 0. When $v=1$, there are two possible situations: 1) Researcher has received private information $\beta_I=1$ and chosen to conduct the experiment on site $t=I$; or 2) Researcher has received private information $\beta_I=0$ and chosen to conduct the experiment on site $t=-I$. By Bayes' rule, her conditional expectation of β_{ATE} is $1/(2-q)$ (see Appendix).

Thus, we show that Evaluator's conditional expectation of β_{ATE} is monotonic in experimental evidence v, and therefore, given any k, the probability for Evaluator to accept Researcher's request is weakly monotonic in experimental evidence v. Given this, Researcher has no incentive to deviate from the Intuitive Strategy since the distribution of experimental evidence under the Intuitive Strategy first order stochastically dominates the one under the deviation. Therefore, the Intuitive Strategy is indeed a best response.

Notice that, if we focus on *responsive equilibrium*, where Evaluator chooses different actions after observing evidence realizations $v=0$ and $v=1$, the Intuitive Strategy constitutes the unique pure-strategy equilibrium. See Appendix A for the detailed proof. Following Di Tillio, Ottaviani, and Sørensen (2017a), our paper focuses on this strategy.[12]

[12]However, there exist other equilibria in which Evaluator's action is not responsive to evidence v. In particular, for k small enough, an equilibrium exists where Researcher adopts a "Counterintuitive Strategy," whereby he switches to the other site for conducting the experiment after observing a positive treatment effect and sticks to the site after observing zero treatment effect. For very large k, where Evaluator always finds it optimal to reject, the No Manipulation strategy, where Researcher's choice is independent of his private information, constitutes an equilibrium.

Table 1: Distribution of Evidence and Conditional Expectation of Average Treatment Effect

v	No Manipulation	Manipulation (rational Evaluator)	Manipulation (naïve Evaluator)
0	$\Pr = 1-q$ (1/2) $E(\beta_{ATE}\|v) = q/2$ (1/4)	$\Pr = (1-q)^2$ (1/4) $E(\beta_{ATE}\|v) = 0$	$\Pr = (1-q)^2$ (1/4) $E(\beta_{ATE}\|v) = q/2$ (1/4)
1	$\Pr = q$ (1/2) $E(\beta_{ATE}\|v) = 1 + q/2$ (3/4)	$\Pr = 2q - q^2$ (3/4) $E(\beta_{ATE}\|v) = 1/2 - q$ (2/3)	$\Pr = 2q - q^2$ (3/4) $E(\beta_{ATE}\|v) = 1 + q/2$ (3/4)

Notes: The values in parentheses are calculated with $q = 1/2$.

2.3 Comparison of Experimental Outcomes under Manipulation and No Manipulation

In light of the equilibrium characterization above, we may summarize the probability distribution of evidence in the experiment and the expected average treatment effect associated with each evidence realization. Table 1 provides one such summary. In the table, we list the outcomes under No Manipulation and Manipulation (when Researcher employs the Intuitive Strategy) with a rational Evaluator, who makes Bayesian inferences based on Researcher's equilibrium strategy. In addition, in the last column, we present the Manipulation case with a naïve Evaluator, where Evaluator makes inferences under the (erroneous) assumption that Researcher's choice of experimental site is random. In this case, Evaluator's *perceived* expected average treatment effect conditional on evidence v is the same as in the case of No Manipulation, but the distribution of evidence is the same as in the case of Manipulation with a rational Evaluator.

From Table 1, we see that compared with No Manipulation, Manipulation (coupled with a rational Evaluator) has two effects. On the one hand, Manipulation increases the probability for Evaluator to observe positive experimental evidence. On the other hand, it decreases Evaluator's conditional expectation of the average treatment effect, regardless of the evidence realization. The second effect is due to a rational Evaluator's ability to make inferences that take into account Researcher's manipulation. In contrast, this effect is absent with a naïve Evaluator.

2.4 Welfare Analysis

Following Di Tillio, Ottaviani, and Sørensen (2017a), we provide a welfare analysis of Researcher's manipulation of the experiment in the form of strategic sample selection. In the top panel of Figure 1, we present the change in Researcher's expected payoff under Manipulation versus that under No Manipulation, where the left panel assumes rational Evaluator and the right assumes naïve Evaluator.[13] In the bottom panel, we present the same comparisons for Evaluator. In the figure and in the remaining theoretical analysis and our experiment, we will focus on the case where $q = 1/2$. In this case, it is inconsequential whether or not Evaluator observes the site on which Researcher has private information (see Footnote 10).

We first study the welfare comparison of Manipulation versus No Manipulation with a rational Evaluator. The effect of manipulation on the payoffs of Researcher and Evaluator is non monotonic in k. To see this, recall the two contrasting effects on the random distribution of evidence and the associated expected treatment effect, which we identify at the end of the previous subsection. Consider Researcher's welfare (top-left panel). For very low and intermediately high cost of acceptance, $k < 1/4$ and $k \in (2/3, 3/4)$, Researcher is worse off under Manipulation, even if Manipulation is his voluntary choice in equilibrium. When $k < 1/4$, this is because, after observing negative evidence the rational Evaluator revises her expectations downward and always rejects, while she always accepts under No Manipulation. When $k \in (2/3, 3/4)$, Researcher is hurt by Manipulation because a rational Evaluator always rejects his request, while under No Manipulation she would have accepted it after observing positive evidence.

Consider now Evaluator's welfare (bottom-left panel). Since, under Manipulation, Researcher's strategy is conditional on his private information, his choice transmits information to Evaluator. This can be strictly beneficial to Evaluator when the cost of acceptance is not too high, $k < 1/2$ and can hurt her for intermediately high acceptance cost, $k \in (1/2, 3/4)$. Under Manipulation, a rational Evaluator never incorrectly accepts when the actual average treatment effect is $\beta_{ATE} = 0$ ($\beta_L = \beta_R = 0$), as Evaluator will observe negative evidence and correctly infer the true $\beta_{ATE} = 0$. This completely accounts for the positive effect of manipulation on Evaluator for $k < 1/4$. For higher k, under both Manipulation and No Manipulation, Evaluator accepts only after observing positive evidence $v = 1$. For $k \in (1/4, 1/2)$, under perfect information, it is optimal for Evaluator to accept except when $\beta_{ATE} = 0$. Under manipulation, it is exactly what the

[13] See Appendix 6.3 for details on the calculation of welfare in the different situations.

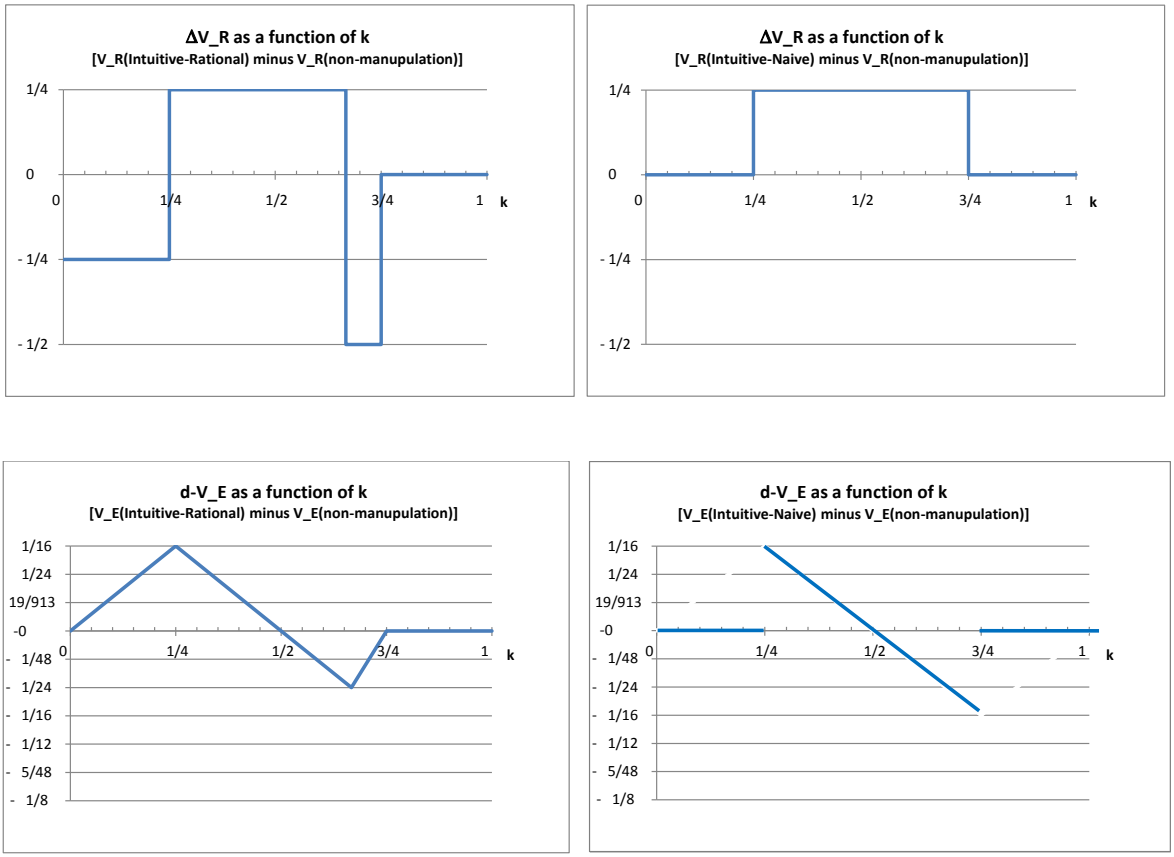

Figure 1: Welfare comparison between Manipulation and No Manipulation for the case $q = 1/2$. Note: The top row is for Researcher and the bottom one for Evaluator; the left panel is for rational Evaluator and the right one for Naïve Evaluator.

rational Evaluator does in equilibrium. However, under No Manipulation, Evaluator rejects whenever evidence is negative $v = 0$, but when the other site's treatment effect is 1 and therefore $\beta_{ATE} = 1/2$, this results in incorrect rejection. In contrast, for $k \in (1/2, 2/3)$, when $\beta_{ATE} = 1/2$, it is optimal for Evaluator to reject, which Evaluator does when she observes negative evidence $v = 0$ under No Manipulation. However, under Manipulation, for $k \in (1/2, 2/3)$ Evaluator always accepts after observing positive evidence $v = 1$, which results in incorrect acceptance. The negative payoff difference for $k \in (2/3, 3/4)$ comes from the fact that a rational Evaluator always rejects under Manipulation, which is correct when $\beta_{ATE} = 1/2$ but incorrect when $\beta_{ATE} = 1$ ($\beta_L = \beta_R = 1$).

We now turn to the welfare comparison of Manipulation versus No Manipulation,

with a naïve Evaluator. Evaluator uses the same acceptance rule conditional on evidence in the two situations. Thus, Researcher is always weakly better off under Manipulation. For low and high level of k, Evaluator's acceptance rule is independent of evidence, prescribing to always accept for $k \leq 1/4$ and to always reject for $k > 3/4$. There is no change in payoffs for either Researcher or Evaluator. Researcher is strictly better off for intermediate levels of k, $k \in (1/4, 3/4)$, where Evaluator accepts only conditional on positive evidence, because under Manipulation the probability of observing positive outcomes is higher. Perhaps surprisingly, a naïve Evaluator under manipulation can do better than under No Manipulation for $k < 1/2$. This is because in states where the real average treatment effect is $1/2$, under No Manipulation there is a 0.5 probability (when she observes negative evidence) that Evaluator incorrectly rejects, while under Manipulation she always accepts, as in the corresponding states she always observes positive evidence. The opposite holds for higher k, where under Manipulation, it is the naïve Evaluator who incorrectly accepts in these states.

Note that Evaluator potentially suffers a loss from being naïve when Researcher engages in manipulation. Consider the bottom right and left panels. For low levels of acceptance cost, $k < 1/4$, a rational Evaluator takes into account the manipulation incentives of Researcher, who, despite the conflicts of interest, transmits useful information to Evaluator. A rational Evaluator correctly deduces that negative evidence implies that both sites have treatment effects of zero and revises her expectation accordingly. In contrast, a naïve Evaluator fails to take manipulation into account and always accepts regardless of evidence. For acceptance cost $k \in [1/4, 2/3)$, a rational Evaluator and a naïve Evaluator employ the same acceptance rule conditional on evidence, and therefore have the same payoff. The same holds for high acceptance cost $k > 3/4$, where both reject regardless of evidence. However, for $k \in [2/3, 3/4)$ after positive evidence a rational Evaluator rejects while a naïve Evaluator accepts. A naïve Evaluator fails to infer that the expected treatment effect on the site not chosen for the experiment should be revised lower than the ex ante expectation, because Researcher may have observed negative information about the other site, and therefore incorrectly accepts.

To summarize our welfare analysis, note that (1) Researcher may benefit or get hurt from manipulation through strategic sample selection when he faces a rational Evaluator, while he always weakly benefits from such manipulation when he faces a naïve Evaluator; (2) a rational Evaluator benefits from manipulation when the acceptance cost is relatively low but is hurt by it when the acceptance cost is intermediately high

(up to the level of expected value of the average treatment effect after observing positive evidence under No Manipulation); however, a naïve Evaluator gives up the gain when acceptance cost is low and suffers worse damages for the upper range of intermediately high cost.

3 Experimental Design

In our experiment, we fix $q = 1/2$ and choose the other parameters as follows. We first calculate the cutoffs of the value of k for Evaluator's equilibrium strategy by assuming that players are risk neutral and Researcher follows the Intuitive Strategy as in the model.

$$\text{rational Evaluator's equilibrium strategy} = \begin{cases} \text{accept} & \text{if } k \leq 0.67 \,\&\, v = 1; \\ \text{reject} & \text{if } k \leq 0.67 \,\&\, v = 0; \\ \text{always reject} & \text{if } k > 0.67. \end{cases}$$

$$\text{naïve Evaluator's equilibrium strategy} = \begin{cases} \text{always accept} & \text{if } k \leq 0.25; \\ \text{accept} & \text{if } 0.25 < k \leq 0.75 \,\&\, v = 1; \\ \text{reject} & \text{if } 0.25 < k \leq 0.75 \,\&\, v = 0; \\ \text{always reject} & \text{if } k > 0.75; \end{cases}$$

Recall that Evaluator's equilibrium strategy under No Manipulation is the same as under Manipulation with a naïve Evaluator. In addition, given that previous experimental studies show that most people are risk averse, we also calculate the cutoffs for k by assuming that both players have a constant relative risk aversion (CRRA) utility function $u(m) = m^r$ with $r = 0.5$, where m is the monetary payment received by an agent.[14]

[14] As summarized by Holt and Laury (2002), estimates for relative risk aversion in the literature vary between 0.3-0.7 in different decision tasks.

Table 2: Predictions

	$k_1 = 10$	$k_2 = 40$	$k_3 = 70$
$v = 1$	rational E accept	rational E accept	**rational E reject**
	naïve E accept	naïve E accept	**naïve E accept**
$v = 0$	**rational E reject**	rational E reject	rational E reject
	naïve E accept	naïve E reject	naïve E reject

$$\text{rational Evaluator's equilibrium strategy} = \begin{cases} \text{accept} & \text{if } k \leq 0.65 \,\&\, v = 1; \\ \text{reject} & \text{if } k \leq 0.65 \,\&\, v = 0; \\ \text{always reject} & \text{if } k > 0.65. \end{cases}$$

$$\text{naïve Evaluator's equilibrium strategy} = \begin{cases} \text{always accept} & \text{if } k \leq 0.125; \\ \text{accept} & \text{if } 0.125 < k \leq 0.73 \,\&\, v = 1; \\ \text{reject} & \text{if } 0.125 < k \leq 0.73 \,\&\, v = 0; \\ \text{always reject} & \text{if } k > 0.73; \end{cases}$$

Based on these calculations, we choose values of k, as in Table 2, leading to different Evaluator's best responses and welfare outcomes.

In our theoretical model, Evaluator observes neither the experiment site nor the site on which Researcher has received private information. Therefore, Evaluator's strategy only depends on the experimental evidence v. In our experiment, the probability for Researcher to observe private information from Left site is set to $m = 1/2$, but it is not explained in the instructions, as from a theoretical point of view the value of such probability is irrelevant. The theoretical analysis would not change if Evaluator observed the site where the experiment takes place when $m = 1/2$. However, this may no longer be true if m takes other values, where, among other things, a pure-strategy equilibrium may fail to exist.

Our experiment uses both between-subject and within-subject design. Within each session, we varied the value of acceptance cost k and whether Researcher receives private information that enables him to manipulate sample selection. Each session started with a No Manipulation treatment, followed by a Manipulation treatment.

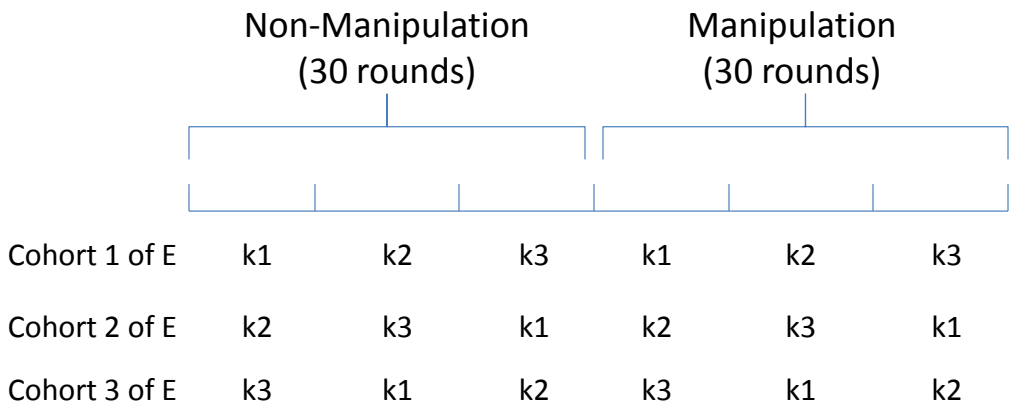

Figure 2: Structure of Sessions

We chose this order on purpose to allow subjects to first learn to play the game in the easier environment (No Manipulation) and then switch to the more complicated one (Manipulation). Each treatment consisted of 30 rounds. There were 3 practice rounds before each treatment started. Subjects only received the instructions for the Manipulation treatment right before it started, but they were informed at the beginning of the session that there would be two parts of the experiment. For each session, Evaluators made decisions under all the three levels of acceptance cost, k, in both No Manipulation and Manipulation treatments. In order to facilitate Evaluator's learning process, each Evaluator experienced the same value of k for a duration of 10 consecutive rounds before switching to a new value of k. At the same time, we kept the distribution of Evaluators possessing different values of k constant in every round, as shown in Figure 2.

In order to identify the possible effect of other-regarding preferences and/or strategic uncertainty on Evaluators' behaviour, we introduce in our experimental design a third treatment variable, namely, whether the role of Researcher is played by human subjects who make conscious decisions or robots who always follow the intuitive strategy. We refer to these two treatments as Human Researcher treatment and Robot Researcher treatment, respectively. In the Human Researcher treatment, there were 12 subjects in each session, who were randomly assigned as Researcher (Player A) and Evaluator (Player B) at the beginning of the session and remained in the same role until the end of the experiment.

During a session, in each round Evaluators and Researchers were randomly and

anonymously paired with each other. Then, each pair of matched Evaluator and Researcher were both told the value of k of the Evaluator in their match. In order to remove framing effects, we described the game in each round using a neutral context: "There are two bins, LEFT BIN and RIGHT BIN. In each bin there are 50 balls of a same colour, RED or BLUE. The computer will randomly draw the colour of the balls in each bin at the beginning of each round, with 50% chance being RED and 50% chance being BLUE.....The two bins together represent a project. Each RED ball has a value of 1 point and each BLUE ball has no value. Therefore, the value of the project is equal to the total number of RED balls in the two bins.....Player A will have to choose one bin, LEFT BIN or RIGHT BIN. Then the computer will reveal the colour of the balls in the chosen bin to the matched Player A and Player B. After that, Player B will have to make a decision on whether to IMPLEMENT or NOT IMPLEMENT the project." In addition, we described the cost of acceptance for Evaluator, k, as Evaluator's endowed income, which she will have to forgo if choosing IMPLEMENT and will keep if choosing NOT IMPLEMENT.

At the end of each session, two rounds from each of the No Manipulation and Manipulation treatments were randomly chosen for the actual payment. Although in each round subjects were shown the history of play and payoffs from each previous round in that treatment, they were only informed of which rounds were chosen after both treatments were finished. We chose this design to control for income effects. We conducted the experiment at the Bell experimental economics lab at the research centre CIRANO (Montreal, Canada). At the end of the session, subjects were paid privately, in cash, their earnings from the four chosen rounds as well as a show-up fee of $10, using an exchange rate of 10 points=$1CAD. In the rare case where a subject's total earnings, including the show-up fee, is less than $15, the subject receives $15 per CIRANO lab regulations.

4 Results

We conducted 3 sessions under the Human Researcher treatment, with 18 pairs of Researchers and Evaluators, and 1 session under the Robot Researcher treatment, with 18 Evaluators. Rigorously speaking, with random matching between Researchers and Evaluators in each round, data from each session is considered to be an independent observation. In what follows, we present statistical tests using individual-level data

Table 3: Subjects' Payment

	Average Earnings	Min	Max	No. of Obs.	Session
Human Researcher	\$25	\$0	\$40	18	1-3
Evaluator (Human)	\$25.39	\$14	\$34	18	1-3
Evaluator (Robot)	\$23.72	\$10	\$35	18	4

as independent observations. Although this choice of approach relaxes the criterion on independent observation to some extent, we consider it to be acceptable since in our experiment subjects were anonymous throughout the session, randomly matched to each other in each period, and interact with each other in a finite horizon.[15]

Table 3 summarizes the payment for Researchers and Evaluators, excluding the show-up fee.[16] The average earnings between Researchers and Evaluators are not significantly different, by two-tailed Wilcoxon Mann-Whitney test ($p = 0.51$, 18 vs. 36 obs.). Neither is there a significant difference in Evaluators' earnings between Human Researcher treatment and Robot Researcher treatment, by two-tailed Wilcoxon Mann-Whitney test ($p = 0.48$, 18 vs. 18 obs.).

Figure 3 presents the distribution of earnings for the subjects. The distribution of Evaluators' earnings is more skewed compared to the distribution of Researchers' earnings, which is probably due to the fact that Evaluators have an endowment of income but Researchers do not, so Researchers' payoff is more dependent on the realization of random events.

4.1 Human Researchers' behaviour

First, we calculate Researchers' frequency of choosing Left Bin and Right Bin and find that there is no systematic preference for one of the two bins due to the different positions of the left and right button on the computer screen.[17]

[15]Kandori (1992) shows that cooperation is possible to achieve in equilibrium when a finite population is anonymously and randomly paired to each other to play a Prisoner's Dilemma game in an infinite horizon. Our game differs from the Prisoner's Dilemma game and, in addition, finite interactions should prevent the folk-theorem type of results like those shown by Kandori (1992).

[16]Notice that by the experimental design, Researcher's earnings can only take the value of \$0, \$10, \$20, \$30, or \$40.

[17]In the No Manipulation treatment, the average frequency of choosing the Left Bin is 47.6%, and in the Manipulation treatment it is 52.6%. We further calculate these two frequencies for each Researcher and find no significant difference between the Manipulation and No Manipulation treatments (two tailed matched-pair signed-rank test, $p = 0.5$, 18 obs.).

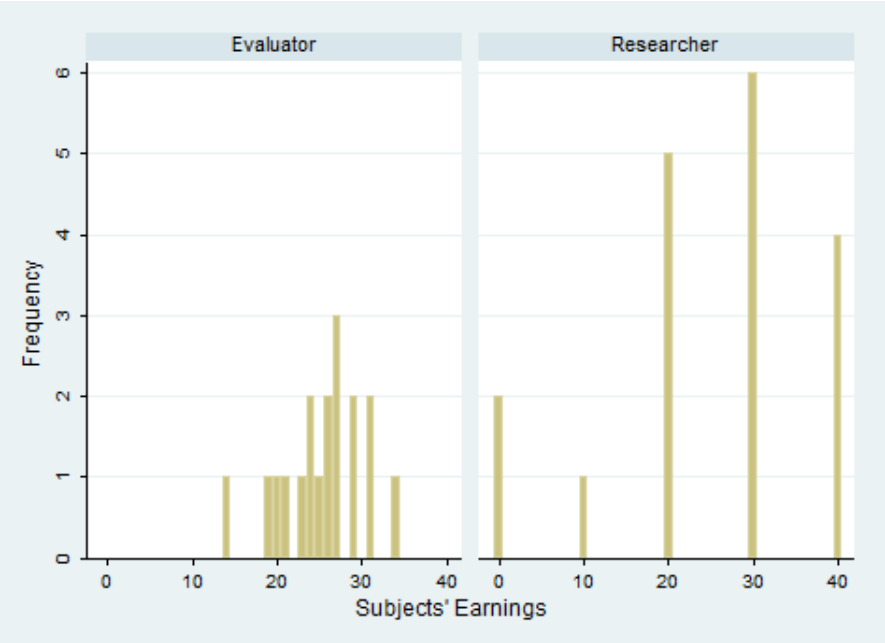

Figure 3: Distribution of Payment (Evaluators vs. Researchers)

We then calculate Researchers' frequency of following the *Intuitive Strategy* in the Manipulation treatment as defined before. First, the average frequency is as high as 83.9%. Second, we calculate the individual frequency for each Researcher and find that the majority of Researchers behave closely to the equilibrium predictions. Figure 4 shows the distribution of Researchers' individual frequency of following the equilibrium strategy. We see that one-third of Researchers follow the Intuitive Strategy in all the 30 rounds, and two-thirds of them follow the Intuitive Strategy at least 90% of time. Third, we further calculate Researchers' average frequency of following the Intuitive Strategy given different contents shown in the message, as presented in Table 4. We see that the conditional frequency is between 80% and 86%, indicating that the message content is not an important factor in influencing whether Researchers follow the equilibrium strategy or not.[18] We also examine Researchers' behaviour across rounds, which is presented in Figure 8 in Appendix B. From the graph we do not observe obvious learning effects over repetitions of the game.

Finally, we conduct a random-effects Probit regression on whether or not Researcher's choice is consistent with the intuitive strategy by using Researchers' choice data in the Manipulation treatment of the Human Researcher sessions (i.e., sessions 1-3). The re-

[18]The frequency of each different message is between 23.9% and 26.5%, not far from the mean of 25%.

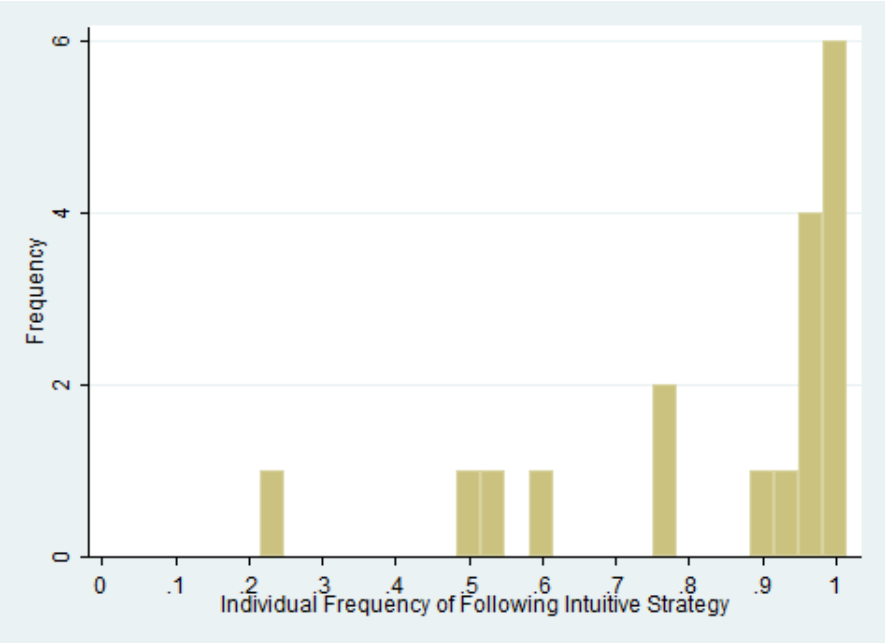

Figure 4: Distribution of Researchers' Individual Frequency of Following Intuitive Strategy in Manipulation Treatment

Table 4: Researchers' Frequency of Following the Intuitive Strategy Conditional on Message Content in the Manipulation Treatment

	All Periods	Message= Left Red	Message= Left Blue	Message= Right Red	Message= Right Blue
Freq. of Message	N/A	25.2%	24.4%	23.9%	26.5%
Conditional Freq. of Intuitive Strategy	83.9%	84.6%	84.1%	80.6%	86.0%

sult is consistent with Table 4 and Figure 8 in Appendix B. Since we adopted random matching between Researchers and Evaluators in each session, the standard errors have been adjusted to allow for clustering of observations by session, and are calculated using the *gllamm* package in Stata. Our focus is on examining the effect of the acceptance cost k, the content of the private message shown to the Researchers (whether the balls in Left/Right Bin are Red/Blue), and the learning effect (period number) on the probability that Researchers follow the Intuitive Strategy. The dependent variable is a dummy variable which is equal to 1 if Researcher's choice of the bin in a period is consistent with the prediction by the Intuitive Strategy given the content of the private message he receives. In the specification, we include the variables Period, k, a dummy

Table 5: Random-Effect Probit Regression on Individual Researcher's Choice of Using Intutive Strategy

Dependent variable	Coefficient
Period	0.002 (0.019)
k	0.002 (0.002)
LeftBin	-0.501 (0.598)
RevealRed	-0.354 (0.813)
LeftRed	0.469 (1.131)
Constant	1.743 (0.518)
No. of Obs.	540

Notes: Robust standard errors clustered at session level are provided in parentheses.

variable LeftBin, which equals 1 if the private message shows the colour of the LEFT bin and 0 if the private message shows the colour of the RIGHT bin, a dummy variable RevealRed, which equals 1 if the private message shows the colour is Red and 0 if the private message shows the colour is Blue, and an interactione dummy variable LeftRed, which equals 1 if the private message shows the colour in the LEFT bin is Red, i.e., if both LeftBin=1 and RevealRed=1. Consistent with other analysis, the marginal effects of the Probit regression in Table 5 shows that none of the independent variables in the regression has a significant effect at the 10% significance level.

Finding 1 *Researchers follow the Intuitive Strategy in the Manipulation treatment to a large extent. Furthermore, the likelihood that Researchers follow the Intuitive Strategy in the Manipulation treatment is not significantly affected by the message content, the value of k, or experience.*

4.2 Evaluators' behaviour

We now turn to the analysis of Evaluators' behaviour. Table 6 shows the average frequencies with which an Evaluator chooses Implement. The top part of the table corresponds to the Human Researcher treatment and the bottom part the Robot Researcher treatment. Within each part, the average frequencies are reported for each realized experimental evidence v (Red or Blue), each value of k (10, 40, or 70) , and each environment (No Manipulation or Manipulation). To facilitate comparisons, we list the theoretically predicted frequency ("Model") alongside the one calculated from the experiment ("Data") and provide the p-value of the two-tailed matched-pair signed rank test between these two frequencies in each (k, v) cell. We see that, in general,

Table 6: Average Frequency of Implementation

Human Researcher Treatment									
No Manipulation (Part One)									
	$k = 10$			$k = 40$			$k = 70$		
	Data	Model	p	Data	Model	p	Data	Model	p
$v = 1$ (Red)	0.917	1	0.046	0.906	1	0.046	0.510	1	*0.001*
$v = 0$ (Blue)	0.600	1	*0.001*	0.273	0	*0.003*	0.075	0	0.026
Manipulation (Part Two)									
	$k = 10$			$k = 40$			$k = 70$		
	Data	Model	p	Data	Model	p	Data	Model	p
$v = 1$ (Red)	0.915	1	0.084	0.903	1	0.084	0.461	0	*0.000*
$v = 0$ (Blue)	0.431	0	*0.002*	0.125	0	0.084	0.085	0	0.084
Robot Researcher Treatment									
No Manipulation (Part One)									
	$k = 10$			$k = 40$			$k = 70$		
	Data	Model	p	Data	Model	p	Data	Model	p
$v = 1$ (Red)	0.981	1	0.317	0.942	1	0.084	0.654	1	*0.002*
$v = 0$ (Blue)	0.875	1	0.026	0.209	0	*0.005*	0.102	0	0.084
Manipulation (Part Two)									
	$k = 10$			$k = 40$			$k = 70$		
	Data	Model	p	Data	Model	p	Data	Model	p
$v = 1$ (Red)	0.974	1	0.084	0.994	1	0.317	0.463	0	*0.002*
$v = 0$ (Blue)	0.373	0	*0.005*	0.162	0	0.026	0.020	0	0.317

Note: $p-$values less than 0.01 are in italics.

Evaluators' behaviour is significantly different from the theoretical prediction ($p < 10\%$ for all tests), and in some cases, very significantly so ($p < 1\%$, in italics). In the Robot Researcher treatment, with the absence of strategic uncertainty on Researchers' behaviour and Evaluators' other-regarding preferences about Researchers' payoffs, we find very similar patterns in Evaluators' behaviour to that under the Human Researcher treatment.

Analysing in Table 6 Evaluator's behaviour in each cell and across cells, we identify the following four types of deviations by Evaluators from the theoretical predictions:

- Case 1: Given Blue evidence, Evaluators exhibit over-implementation when the model predicts no implementation, in all treatments (Manipulation or No Manipulation, Robot or Human Researcher). The extent of over-implementation decreases with the value of k, in both frequency of implementation and p-value.

- Case 2: Given Blue evidence, in the only cell where the model predicts implementation ($k = 10$ and No Manipulation) Evaluators exhibit under-implementation.

- Case 3: Given Red evidence, Evaluators exhibit significant under-implementation only under No Manipulation and $k = 70$.

- Case 4: For the same level of acceptance cost k and evidence v, Evaluators can exhibit over-implementation and under-implementation in the two different manipulation treatments, both in the Human and Robot Research treatments. In particular, when $k = 70$, there is under-implementation after Red in the No Manipulation treatment but over-implementation after Red in the Manipulation treatment. Similarly, when $k = 10$ there is under-implementation after Blue evidence in the No Manipulation but over-implementation after Blue in the Manipulation treatment.

When we compare the frequencies of implementation in Table 6 within treatments (Red vs. Blue) and across treatments (No Manipulation vs. Manipulation), we find that the comparative statics go in the same direction as the theoretical predictions. To address this point formally, using Evaluators' individual frequency of choosing Implement given the value of k, we conduct two-tailed matched-pair signed rank tests for within-treatment comparison and across-treatment comparison, as presented in Table 7 (18 observations for each test). The p-values are put in italics when the comparative statics is inconsistent with the theory. We find that the comparative statics results of Evaluators' behaviour are mostly consistent with the theoretical predictions and that in the Robot Researcher treatment they deviate much less from the theoretical predictions than in the Human Researcher treatment. We summarize our findings on Evaluators' behaviour as follows.

Finding 2 *Compared with the theoretical predictions, Evaluators exhibit both over-implementation and under-implementation (Table 6). However, the overall comparative statics are consistent with the model predictions, especially in the Robot treatment (Table 7).*

To summarize our analysis of Evaluators' behaviour in the experiment, we find significant deviations from the theoretical predictions in terms of the *levels* of the frequency Evaluators choose Implement, as shown in Table 6. We should note, however, that when the theory predicts a *certain* acceptance or rejection decision, it is inevitable

Table 7: p-value of Matched-pair Signed Rank Tests on Evaluators' Frequency of Implementation

Human Researcher treatment			
	$k = 10$	$k = 40$	$k = 70$
Red vs. Blue (No Manipulation)	*0.003*	0.000	0.002
Red vs. Blue (Manipulation)	0.002	0.000	*0.002*
No Manipulation vs. Manipulation (Red)	0.979	0.968	*0.184*
No Manipulation vs. Manipulation (Blue)	*0.274*	*0.036*	0.547
Robot Researcher treatment			
	$k = 10$	$k = 40$	$k = 70$
Red vs. Blue (No Manipulation)	0.105	0.000	0.002
Red vs. Blue (Manipulation)	0.001	0.000	*0.003*
No Manipulation vs. Manipulation (Red)	0.564	0.084	0.037
No Manipulation vs. Manipulation (Blue)	0.004	0.407	0.564

Note: entries in italics indicate inconsistency from theory predictions.

that the experimental outcomes will differ from the theoretical predictions, due to a variety of factors, including initial learning through experimentation and possible cognitive errors. In contrast, the *changes* in the levels of the frequency of implementation are in the most part consistent with theory, as demonstrated by Table 7.

To shed further light on these differences, recall our theoretical discussion on rational and naïve Evaluators in the Manipulation treatment. Note that, as shown in Table 2, these two types of Evaluators behave differently only in two configurations: ($k = 10$, $v = Blue$) and ($k = 70$, $v = Red$).[19] A rational Evaluator should behave as theory predicts, but a naïve Evaluator cannot anticipate that Researcher will use the Intuitive Strategy under the Manipulation treatment. Thus, her behaviour will be similar to that under the No Manipulation treatment, in which she believes that Researcher chooses the site randomly. The comparative statics for the Human Researcher treatment in Table 7 is consistent with some subjects in the Evaluator role behaving in a naïve way. Specifically, the tests that compare the No Manipulation and Manipulation treatments for the configurations ($k = 10$, $v = Blue$) and ($k = 70$, $v = Red$) show that their differences are not significant at the 10% level, which is inconsistent with a rational Evaluator but consistent with a naïve Evaluator. Interestingly, the p-value of these two tests in the Robot Researcher treatment do become significant, $p = 0.004$ and $p = 0.037$

[19]Recall that both types of Evaluators are predicted to have the same behaviour under No Manipulation anyway.

Table 8: p-value from Two-tailed Mann-Whitney Tests on Evaluators' Frequency of Implement (Human vs. Robot Researcher Treatments)

	No Monipulation (Part One)		
	$k = 10$	$k = 40$	$k = 70$
$v = 1$ (Red)	0.171	0.598	0.325
$v = 0$ (Blue)	0.008	0.572	0.528
	Manipulation (Part Two)		
	$k = 10$	$k = 40$	$k = 70$
$v = 1$ (Red)	0.865	0.258	0.732
$v = 0$ (Blue)	0.631	0.432	0.324

respectively, more consistent with the model prediction for rational Evaluators. The different effects of manipulation between the Human Researcher and Robot Researcher treatments in the ($k = 10$, $v = Blue$) and ($k = 70$, $v = Red$) cells suggest that uncertainty about Researcher's strategy may also play a role, in addition to any played by naïveté of Evaluator. Strategic uncertainty may also explain the other differences in the comparative statics tests in Table 7 between the Human and Robot Researcher Treatment.

Finally, we conduct a two-tailed Mann-Whitney test on the individual Evaluator's frequency of implementation between Human Researcher and Robot Researcher treatments. Overall, on the basis of this test we find no significant differences between Evaluators's frequency of implementation in the two treatments. Combined with the analysis above, we summarize the following finding.

Finding 3 *Evaluators' frequency of implementation is not significantly different between Human Researcher and Robot Researcher treatments (Table 8). However, the difference in Evaluator's behaviour between Manipulation and No Manipulation is more consistent with a rational player in Robot Researcher treatment than in Human Researcher treatment.*

5 Welfare Analysis

An important objective of our research is to experimentally test the effect of Researcher's strategic sample selection on Researcher's and Evaluator's payoffs. In this section, we conduct an analysis of subjects' welfare based on their behaviour in the

experiment, in the hope of answering experimentally whether the possibility of manipulations by Researcher is welfare-enhancing/hurting to Researcher and Evaluator. We then compare our findings with the theoretical predictions we drew in Subsection 2.4, which follows similar analysis by Di Tillio, Ottaviani, and Sørensen (2017a). Difficulties arise in our welfare comparisons as players' actual actions and payoffs depend on random realizations, which differ across treatments and sessions. In order to conduct a fair comparison, we first propose a procedure for calculating a welfare index which relies on individuals' actual choice frequencies but the ex ante probability of random realizations instead of the actual realizations. Then, we compare this welfare index between the No Manipulation and Manipulation treatments for both Researchers and Evaluators.

5.1 Construction of Welfare Measures

To construct welfare measures for Evaluator and Researcher, we use a two-step approach. First, based on the experimental data, we calculate Researcher's individual frequency of following the Intuitive Strategy and Evaluator's individual frequency of implementation given k and v, as well as the session averages of these individual frequencies. Second, based on these frequencies, we compute the expected payoff of Researcher and Evaluator, using the ex ante distribution of Red ($v = 1$) and Blue ($v = 0$). So our constructed index provides a welfare measure that is not influenced by the random realizations in a specific session, but is determined by the behaviour of other subjects in his/her session, which is not always consistent with theoretical predictions.

5.1.1 Measuring Researcher's Welfare

We start by constructing a measure of Researcher's welfare. In each session, given the data for each possible realized pair of acceptance cost and evidence (k, v), we can calculate the session-level individual Evaluators' average frequency of implementation. Denote them as $q(k\,,\, Red,\, No)$, $q(\,k,\, Blue,\, No)$, $q(\,k,\, Red,\, Man)$, $q(\,k,\, Blue,\, Man)$, where No and Man are shorthands for No Manipulation and Manipulation treatments, respectively.

Under No Manipulation, the ex ante probability that a colour is drawn from a Bin is

$$p(Red \mid No) = p(Blue \mid No) = 0.5.$$

Therefore, under No Manipulation the expected frequency of implementation for each Researcher i given Evaluator's acceptance cost k is

$$\begin{aligned} U_i^R(\ k,\ No) &= p(Red \mid No) * q(\ k,\ Red,\ No) + p(Blue \mid No,) * q(\ k,\ Blue,\ No) \\ &= 0.5q(\ k,\ Red,\ No) + 0.5q(\ k,\ Blue,\ No). \end{aligned}$$

Since in the experiment Researcher's payoff is simply 100*(frequency of implementation), we can use this ex ante frequency of implementation as an index of welfare that removes the effect of random realizations from Researcher's actual payoff in the experiment. Note that $U_i^R(\ k,\ No)$ is the same for every Researcher in one session since it is based on the ex ante realization and the session level frequency of implementation.

Under Manipulation, a Researcher's expected payoff also depends on the probability that he adopts the Intuitive Strategy, which is denoted by γ_i.[20] Then, using the ex ante probability of the realization of each colour, we can calculate for each Researcher i the probability of $v = Red$ and $v = Blue$, given γ_i:

$$\begin{aligned} p_i(Red \mid Man) &= p(Red)\gamma_i + p(Red)(1-\gamma_i)p(Red) + p(Blue)\gamma_i p(Red) \\ &= 0.5\gamma_i + 0.25, \\ p_i(Blue \mid Man) &= p(Blue)(1-\gamma_i) + p(Red)(1-\gamma_i)p(Blue) + p(Blue)\gamma_i p(Blue) \\ &= 0.75 - 0.5\gamma_i. \end{aligned}$$

Therefore, under Manipulation the expected frequency of implementation for each Researcher i, given k is

$$\begin{aligned} U_i^R(\ k, Man) &= p_i(Red \mid Man) * q(\ k, Red,\ Man) + p_i(Blue \mid Man) * q(\ k, Blue,\ Man) \\ &= (0.5\gamma_i + 0.25) * q(\ k, Red,\ Man) + (0.75 - 0.5\gamma_i) * q(\ k, Blue,\ Man). \end{aligned}$$

To summarize, $U_i^R(\ k,\ No)$ and $U_i^R(\ k, Man)$ are our constructed welfare indices for Researcher under No Manipulation and Manipulation, respectively.

[20] This probability can also be calculated conditional on k and/or the message content. However, since our regression confirms that the likelihood for Researcher to adopt the Intuitive Strategy does not significantly depend on these variables, we calculate only one probability for each individual Researcher for simplicity.

5.1.2 Measuring Evaluator's Welfare

Now, we turn to Evaluator's welfare. To the extent that in some economic environments, like the drug approval process, Evaluator is acting on behalf of a constituency, it is important to compare Evaluator's welfare under No Manipulation and Manipulation. Under No Manipulation, Researcher's Strategy does not affect Evaluator's expected payoff. Under Manipulation, we take the session-level average of individual Researchers' frequency of using the Intuitive Strategy, which we denote by γ. Then, to measure each Evaluator's expected payoff, for each possible realized pair of acceptance cost and evidence $(k,\, v)$, we take the Evaluator's average frequency of implementation based on the data in each session. In particular, for each Evaluator j, we denote her individual frequency of implementation given $(k,\, v)$ for the No Manipulation and the Manipulation treatment by $q_j(\, k,\, v,\, No)$ and $q_j(\, k,\, v,\, Man)$, respectively.

Based on the above construction, we calculate the expected payoffs of each Evaluator j. Under No Manipulation, conditional on the evidence observed by Evaluator, they can be written as

$$\begin{aligned} U_j^E(\, k,\, Red,\, No) &= 75 * q_j(\, k,\, Red,\, No) + k * (1 - q_j(\, k,\, Red,\, No)); \\ U_j^E(\, k,\, Blue,\, No) &= 25 * q_j(\, k,\, Blue,\, No) + k * (1 - q_j(\, k,\, Blue,\, No)). \end{aligned}$$

Therefore, Evaluator j's ex ante expected payoff under No Manipulation is

$$U_j^E(\, k,\, No) = 0.5U_j^E(\, k,\, Red,\, No) + 0.5U_j^E(\, k,\, Blue,\, No),$$

where we used the prior probability of evidence Blue and Red.

Under Manipulation, Evaluator j's expected payoff conditional on evidence $v = Red$ and $v = Blue$ are respectively:

$$\begin{aligned} U_j^E(\, k,\, Red,\, Man) &= \beta(Red,\, \gamma) * q_j(\, k,\, Red,\, Man) + k * (1 - q_j(\, k,\, Red,\, Man)); \\ U_j^E(\, k,\, Blue,\, Man) &= \beta(Blue,\, \gamma) * q_j(\, k,\, Blue,\, Man) + k * (1 - q_j(\, k,\, Blue,\, Man)), \end{aligned}$$

where $\beta(Red,\, \gamma)$ and $\beta(Blue,\, \gamma)$ are the expected numbers of red balls given that Researcher's estimated frequency of using the Intuitive Strategy is γ and evidence is

$v = Red$ or $v = Blue$, respectively. They can be calculated as follows:

$$\begin{aligned}
\beta(Red,\, \gamma) &= \frac{100p(Red)p(Red) + 50p(Red)\gamma p(Blue) + 50p(Blue)\gamma p(Red)}{0.5\gamma + 0.25}, \\
&= \frac{100 + 100\gamma}{2\gamma + 1}; \\
\beta(Blue,\, \gamma) &= \frac{50p(Red)(1-\gamma)p(Blue) + 50p(Blue)(1-\gamma)p(Red) + 0p(Blue)p(Blue)}{0.75 - 0.5\gamma}, \\
&= \frac{100(1-\gamma)}{3 - 2\gamma}.
\end{aligned}$$

Therefore, Evaluator j's payoff under Manipulation is

$$\begin{aligned}
U_j^E(\, k, Man) &= p(Red\,|Man)U_j^E(Red,\, Man,\, k) + p(Blue\,|Man)U_j^E(Blue,\, Man,\, k) \\
&= (0.5\gamma + 0.25)U_j^E(Red,\, Man,\, k) + (0.75 - 0.5\gamma)U_j^E(Blue,\, Man,\, k),
\end{aligned}$$

where $p(v\,|Man)$ is the ex ante probability that evidence v is observed given that Researcher follows the Intuitive strategy with average probability γ.

To summarize, $U_j^E(\, k,\, No)$ and $U_j^E(\, k, Man)$ are our constructed welfare indices for Evaluator under No Manipulation and Manipulation, respectively.

5.2 Results of Welfare Comparison

Now, we conduct our welfare analysis using the welfare measures we constructed above. Table 9 reports the average of Researcher's welfare and the p-value of two-tailed matched-pair signed rank tests which compare Researcher's welfare index between No Manipulation and Manipulation treatments. We find that Researcher is better off by manipulation when $k = 40$ and $k = 70$ and is not worse off when $k = 10$. This is inconsistent with the theoretical model with a rational Evaluator, which predicts that Researcher becomes worse off by manipulation when $k = 10$ and $k = 70$. It is however consistent with the prediction of the theory with a naïve Evaluator. As we have shown in Tables 6 and 7, Evaluators do not sufficiently discount positive evidence in the Manipulation treatment when k is large. Neither do they sufficiently take into account the implication of negative evidence in the Manipulation treatment; that is, they frequently fail to observe either of these two facts: (1) Researcher engages in manipulation; (2) manipulation by Researcher in the form of the Intuitive Strategy means that the observation of negative evidence on the experimental site implies that the treatment effect on the other

Table 9: Researcher's Welfare Comparison

	$k = 10$		$k = 40$		$k = 70$
No Manipulation	75.87		58.95		29.23
Manipulation	75.48		64.51		33.65
p-value	0.53		0.03		0.05
Number of Obs.	18		18		18

Table 10: Evaluator's Welfare Comparison

Human and Robot Researcher Treatment					
	$k = 10$		$k = 40$		$k = 70$
No Manipulation	46.15		54.22		69.93
Manipulation	48.22		57.32		68.36
p-value	0.005		0.001		0.004
Number of Obs.	35		35		35
Human Researcher Treatment					
	$k = 10$		$k = 40$		$k = 70$
No Manipulation	44.30		53.79		69.60
Manipulation	46.25		56.17		67.97
p-value	0.221		0.009		0.098
Number of Obs.	18		18		18
Robot Researcher Treatment					
	$k = 10$		$k = 40$		$k = 70$
No Manipulation	48.38		54.74		70.32
Manipulation	50.59		58.69		68.83
p-value	0.002		0.028		0.021
Number of Obs.	17		17		17

site is also zero. This offsets any theoretically predicted negative effect on Researcher.

Similarly, Table 10 reports the Evaluator's average welfare and the p-value of two-tailed matched-pair signed rank tests for Evaluator's welfare between No Manipulation and Manipulation treatments. We find that Evaluator becomes better off under the Manipulation treatment than under the No Manipulation treatment when $k = 10$ and $k = 40$, but becomes worse off when $k = 70$. This finding is consistent with the theoretical prediction. We summarize in the following finding the results of our welfare analysis.

Finding 4 *Researcher's welfare significantly improves under the Manipulation treat-*

ment compared with the No Manipulation treatment. Evaluator's welfare significantly improves for acceptance cost levels $k = 10$ and $k = 40$ but decreases for acceptance cost $k = 70$ under the Manipulation treatment compared with the No Manipulation treatment.

6 Conclusion

Scientists' manipulation during the procedure of data collection and analysis in experiments is generally viewed unfavourably – it challenges the validity of their experimental findings. The history of RCTs is a history of the struggle to keep them truly randomized.[21]

In their simple yet insightful theoretical analysis, Di Tillio, Ottaviani, and Sørensen (2017a) offer a cautionary tale on such received wisdom. In their model, Researcher tries to persuade Evaluator to accept the finding of a scientific study by convincing her that the treatment effect is large enough, where Researcher and Evaluator are both fully rational. They show that Researcher is hurt by the possibility of manipulation through strategic sample selection, when Evaluator's acceptance cost is very low or very high, or in other words, Evaluator is ex ante strongly for or against acceptance. In these instances, one might expect to see conscious efforts by Researcher to refrain from engaging in it, if he could. Evaluator, on the other hand, may benefit from Researcher's manipulation, when she has low or medium cost of acceptance. In contrast, both Evaluator and Researcher may be hurt when Evaluator has a high cost of acceptance, as manipulation would make Evaluator discount positive findings so much that it eliminates the possibility of convincing Evaluator.

In this paper, we report results of an experiment directly based on Di Tillio, Ottaviani, and Sørensen's (2017a) theoretical model. Our experimental design tests the theoretical predictions when such manipulation is feasible and not feasible to the Researcher, and when Researcher is played by a human subject or a robot. Our results largely confirm the theoretical predictions of Researcher's behaviour: that they engage in manipulation in the form of an "Intuitive Strategy." However, Evaluator's behaviour demonstrates significant deviations from the theoretical predictions, even though the comparative statics is consistent with the theoretical predictions. Our welfare analy-

[21] As mentioned in the Introduction, please refer to Di Tillio, Ottaviani, and Sørensen's (2017a) historical account, and the references they cite, which include Chalmers (1999); Fisher (1925, 1926, 1935); Hart (1999); Neyman (1923), among others.

sis offers a mixed message on the consequences of manipulation. First, we find that Researcher always benefits from the possibility of manipulation, in contrast to the theoretical prediction that he is hurt by it with a low or high acceptance cost of Evaluator. Second, consistent with theoretical predictions, Evaluator is hurt by the possibility of Researcher's manipulation when her acceptance cost is high but benefits from that possibility with a low or medium acceptance cost.

Our study is a first step in experimentally testing the effect of manipulation on the research evaluation process. The natural next step is to test other forms of manipulation and provide further guidance on related public policy.

7 Appendix A: Proof for Equilibrium

In this Appendix, we prove that the Intuitive Strategy by Researcher is an equilibrium strategy.

7.1 Intuitive Strategy

We first discuss how the evaluator calculates $E(\beta_{ATE}|v)$ if Researcher follows the Intuitive Strategy:

1. If $v = 0$, then

$$\beta_I = 0 \quad \text{and } \beta_t = \beta_{-I} = 0 \quad \text{with probability } (1-q)^2.$$

Therefore,

$$\begin{aligned} \Pr(v=0) &= (1-q)^2, \\ E(\beta_{ATE}|v=0) &= 0. \end{aligned}$$

2. If $v = 1$, then there are two possible cases:

$$\begin{array}{llll} \text{case 1} & \beta_I = \beta_t = 1 & \beta_{-I} \in \{0,1\} & \text{with probability } q \\ \text{case 2} & \beta_I = 0 & \beta_t = \beta_{-I} = 1 & \text{with probability } q(1-q), \end{array}$$

Therefore,

$$\begin{aligned} \Pr(v=1) &= q(2-q), \\ E(\beta_{ATE}|v=1) &= \frac{q(1+q)/2+q(1-q)/2}{q(2-q)} = \frac{1}{2-q}. \end{aligned}$$

It is straightforward to verify that $E(\beta_{ATE}|v=1) > E(\beta_{ATE}|v=0) = 0$. This means that, for any given k, the evaluator's strategy is weakly monotone in experimental evidence v, and strictly so if $k < 1/(2-q)$.

We now show that the proposed Intuitive Strategy for Researcher is indeed an equilibrium strategy, given the strategy of Evaluator. In order to show this, we check that Researcher has no incentives to deviate from the Intuitive Strategy given any possible experiment evidence.

- when $\beta_A = 1$ the probability distribution of the possible outcomes is:

	Intuitive Strategy: $t = I$ w.p.	Deviation: $t = -I$ w.p.
$v = 0$	0	$1-q$
$v = 1$	1	q

Therefore, the distribution of outcomes under the Intuitive Strategy first-order stochastically dominates the one under the deviation. So Researcher does not want to deviate.

- when $\beta_I = 0$ the probability distribution of the possible outcomes is:

	Intuitive Strategy: $t = -I$ w.p.	Deviation: $t = I$ w.p.
$v = 0$	$1-q$	1
$v = 1$	q	0

Therefore, the distribution of outcomes under the Intuitive Strategy first-order stochastically dominates the one under the deviation. So again Researcher does not want to deviate.

7.2 Uniqueness of Intuitive Strategy (in Responsive Equilibrium)

Definition 5 *A Strategy is Responsive when Evaluator's action differs conditional on different evidence v. A Responsive equilibrium is one where Evaluator plays a Responsive Strategy.*

We now show that in a Responsive equilibrium Researcher cannot adopt an alternative strategy to the Intuitive Strategy. Namely, he cannot adopt a Counter-intuitive Strategy or Non-manipulative Strategy. Thus, the Intuitive Strategy is the unique Responsive equilibrium strategy.

Definition 6 *The Counter-intuitive Strategy is as follows:*

- *If $\beta_I = 1$, then conduct the experiment in $-I$, i.e., $t = -I$.*
- *If $\beta_I = 0$, then conduct the experiment in I, i.e., $t = I$.*

We now discuss how Evaluator calculates $E(\beta_{ATE}|v)$ if Researcher follows this Counter-intuitive Strategy:

1. If $v = 0$, then there are two possible cases:

$$\begin{array}{llll} \text{Case 1} & \beta_I = \beta_t = 0 & \beta_{-I} \in \{0,1\} & \text{w.p. } 1-q \\ \text{Case 2} & \beta_I = 1 & \beta_t = \beta_{-I} = 0 & \text{w.p. } q(1-q) \end{array}$$

   Therefore,

$$\begin{aligned} \Pr(v=0) &= \frac{1-q}{1+q}, \\ E(\beta_{ATE}|v=0) &= \frac{(1-q)(q/2)+q(1-q)/2}{(1-q)(1+q)} = \frac{q}{1+q}. \end{aligned}$$

2. If $v = 1$, then

$$\beta_I = 1 \quad \beta_t = \beta_{-I} = 1 \quad \text{w.p. } q^2$$

   Therefore,

$$\begin{aligned} \Pr(v=1) &= q^2, \\ E(\beta_{ATE}|v=1) &= 1. \end{aligned}$$

It is easy to check that also in this case for any given k, Evaluator's strategy is weakly monotone in experimental evidence v since $E(\beta_{ATE}|v=1) = 1 > E(\beta_{ATE}|v=0) = q/(1+q)$ and is strictly so in evidence for $k \in (q/(1-q), 1)$.
We now show that, when Evaluator plays a Responsive Strategy, the Counter-intuitive strategy is not an equilibrium strategy, as we can find a deviation that makes Researcher better off. In particular, Researcher would want to deviate when his private information is $\beta_I = 0$. The probability distribution of the possible outcomes in this case is:

	Counter-intuitive Strategy: $t = I$ w.p.	Deviation: $t = -I$ w.p.
$v = 0$	1	$1 - q$
$v = 1$	0	q

Since the distribution of outcomes under the deviation first-order stochastically dominates the one under the Counter-intuitive Strategy and Evaluator uses a responsive strategy, which is increasing in evidence, Researcher's expected payoff from the deviation would be greater than the one from the Counter-intuitive Strategy. Therefore, he would want to deviate.

Definition 7 *A Non-manipulative Strategy is one where Researcher's choice is independent of his private information.*

Since we are considering only pure strategies, a Non-manipulative Strategy is one where Researcher always chooses a specific experiment site, regardless of the private information.

Again, Evaluator's strategy is weakly monotone in experimental evidence v since $E(\beta_{ATE}|v = 1) = \frac{1+q}{2} > E(\beta_{ATE}|v=0) = \frac{q}{2}$, strictly when $k \in (\frac{q}{2}, \frac{1+q}{2})$.

It is easy to show, that for the same argument as for the Counter-intuitive Strategy, Researcher would want to deviate when his private information is $\beta_I = 0$. The same argument would also apply to the case Researcher chooses the experiment site randomly.

7.3 Welfare Analysis

The graphs in Figure 4 are constructed using the information in Table 1. Similarly the graphs in Figures 5-6 are constructed using the Researcher's and Evaluator's strategies

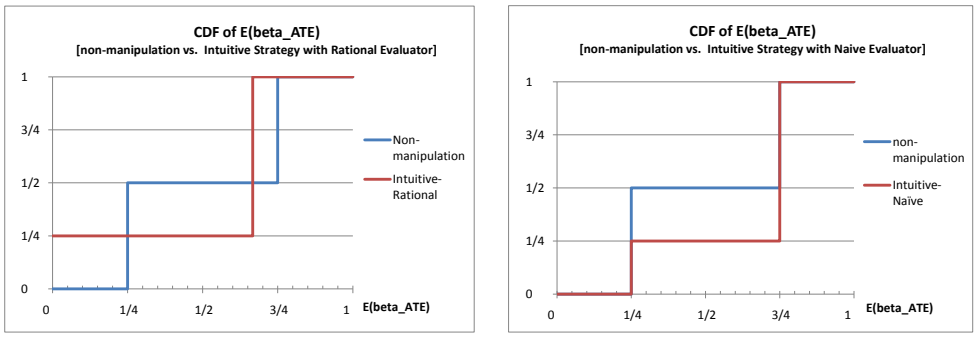

Figure 5: The distribution of the expected value of average treatment effect, β_{ATE}.

in the different situations and the information contained in Table 1 about the probability of evidence and on the posterior beliefs in the different situations.

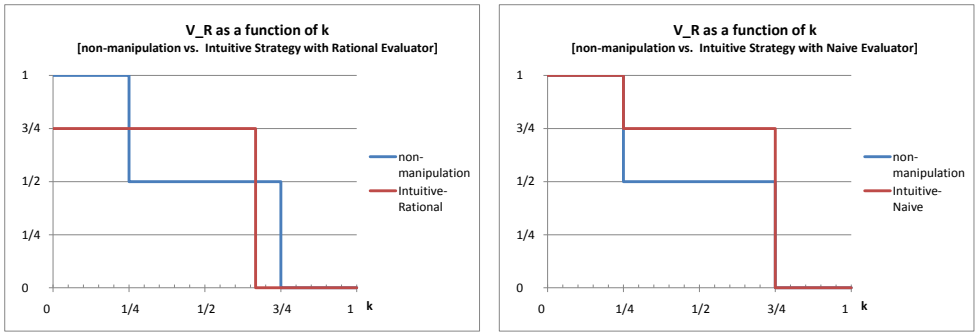

Figure 6: Receiver's expected payoff, V_R.

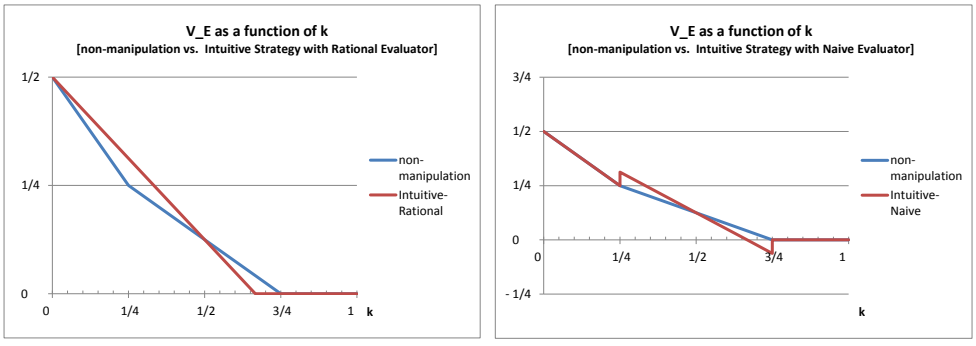

Figure 7: Evaluator's expected payoff, V_E.

Let us analyze first Researcher's (expected) payoff under No Manipulation, Manip-

ulation with rational Evaluator, and Manipulation with a naïve Evaluator:

No Manipulation		Manipulation (rational Evaluator)		Manipulation (naïve Evaluator)	
$V_R = 1$	If $k \leq \frac{1}{4}$			$V_R = 1$	If $k \leq \frac{1}{4}$
$V_R = 1/2$	If $k \in (\frac{1}{4}, \frac{3}{4}]$	$V_R = 3/4$	If $k \leq \frac{2}{3}$	$V_R = 3/4$	If $k \in (\frac{1}{4}, \frac{3}{4}]$
$V_R = 0$	If $k > \frac{3}{4}$	$V_R = 0$	If $k > \frac{2}{3}$	$V_R = 0$	If $k > \frac{3}{4}$

For example, under No Manipulation (the left part of the table), the Evaluator posterior beliefs are $E(\beta|0) = 1/4$ and $E(\beta|1) = 3/4$. Therefore she always accepts, independent of evidence, for $k \leq 1/4$, which gives a payoff 1 to Researcher. Similarly, she accepts only after good evidence for $k \in (1/4,\ 3/4]$, which gives Researcher a payoff of 1 with probability $1/2$ ($v = 1$) and 0 with probability $1/2$ ($v = 0$). Evaluator always rejects for higher k, which gives Researcher a 0 payoff.

When Researcher uses the Intuitive Strategy and faces a naïve Evaluator (the right part of the table), Evaluator has the same posterior beliefs as under No Manipulation and therefore uses the same strategy: she always accepts for $k \leq 1/4$, accepts after positive evidence for $k \in (1/4,\ 3/4]$, and always rejects for $k > 3/4$. However, now Researcher's expected payoffs must be calculated using the probability of observing positive or negative evidence induced by the Intuitive Strategy. So what changes with respect to No Manipulation is that for intermediate k, Evaluator accepts with probability $3/4$ ($v = 1$ under Manipulation).

When Researcher uses the Intuitive Strategy and faces a rational Evaluator (middle part of the table), Evaluator posterior beliefs are $E(\beta|0) = 0$ and $E(\beta|1) = 2/3$. Therefore, she now always rejects after negative evidence, independent of k, and accepts after positive evidence only for $k \leq 2/3$. It follows that Researcher payoff is 1 with probability $3/4$ (the probability of $v = 1$ under manipulation) and 0 with probability $3/4$ (the probability of $v = 0$ under manipulation). For higher k Researcher's expected payoff is zero.

Let us now analyze Evaluator's (expected) payoffs under No Manipulation, Manipulation with rational Evaluator, and Manipulation with Naïve Evaluator:

No Manipulation		Manipulation (rational Evaluator)		Manipulation (naïve Evaluator)	
$V_E = \frac{1}{2} - k$	If $k \leq \frac{1}{4}$			$V_E = \frac{1}{2} - k$	If $k \leq \frac{1}{4}$
$V_E = \frac{1}{2}(\frac{3}{4} - k)$	If $k \in (\frac{1}{4}, \frac{3}{4}]$	$V_E = \frac{3}{4}(\frac{2}{3} - k)$	If $k \leq \frac{2}{3}$	$V_E = \frac{3}{4}(\frac{2}{3} - k)$	If $k \in (\frac{1}{4}, \frac{3}{4}]$
$V_E = 0$	If $k > \frac{3}{4}$	$V_E = 0$	If $k > \frac{2}{3}$	$V_E = 0$	If $k > \frac{3}{4}$

Consider the No Manipulation environment (the left part of the table). Evaluator always accepts when $k \leq 1/4$, so she always pays the cost k, and obtains the ex-ante expected value of the project $E(\beta) = 1/2$. When $k \in (1/4, 3/4]$, Evaluator accepts only after positive evidence, which happens with probability $1/2$. Conditional on this, she pays the cost k and obtains $E(\beta|1) = 3/4$. For higher k, she always rejects so his payoff is zero.

Consider the situation with Manipulation and rational Evaluator (the middle part of the table). Evaluator accepts if and only if $k \leq 2/3$ and evidence is positive. Her expected payoff is therefore zero for $k > 2/3$. For $k \leq 2/3$, conditional on positive evidence, which happens with probability $3/4$ under manipulation, she pays the cost k and obtains $E(\beta|1) = 2/3$.

Consider now the situation with Manipulation and naïve Evaluator (the right part of the table). Evaluator uses the same acceptance strategy as in the left part of the table, so there are three cases to consider: $k \leq 1/4$, $k \in (1/4, 3/4]$, and $k > 3/4$. When k is small or large, she always accepts or always rejects, respectively, like under No Manipulation. Her payoffs do not depend on the evidence and are therefore the same as under No Manipulation. For intermediate k, Evaluator accepts conditional on positive evidence, which happens with probability $3/4$ under Manipulation, so she pays the cost k and obtains $E(\beta|1) = 2/3$.

8 Appendix B: Additional Results

See Figure 8.

9 Appendix C: Instructions

Welcome

Welcome to this experiment on economic decision making. There will be two parts in today's experiment, each consisting of 30 rounds. Your earnings will depend on

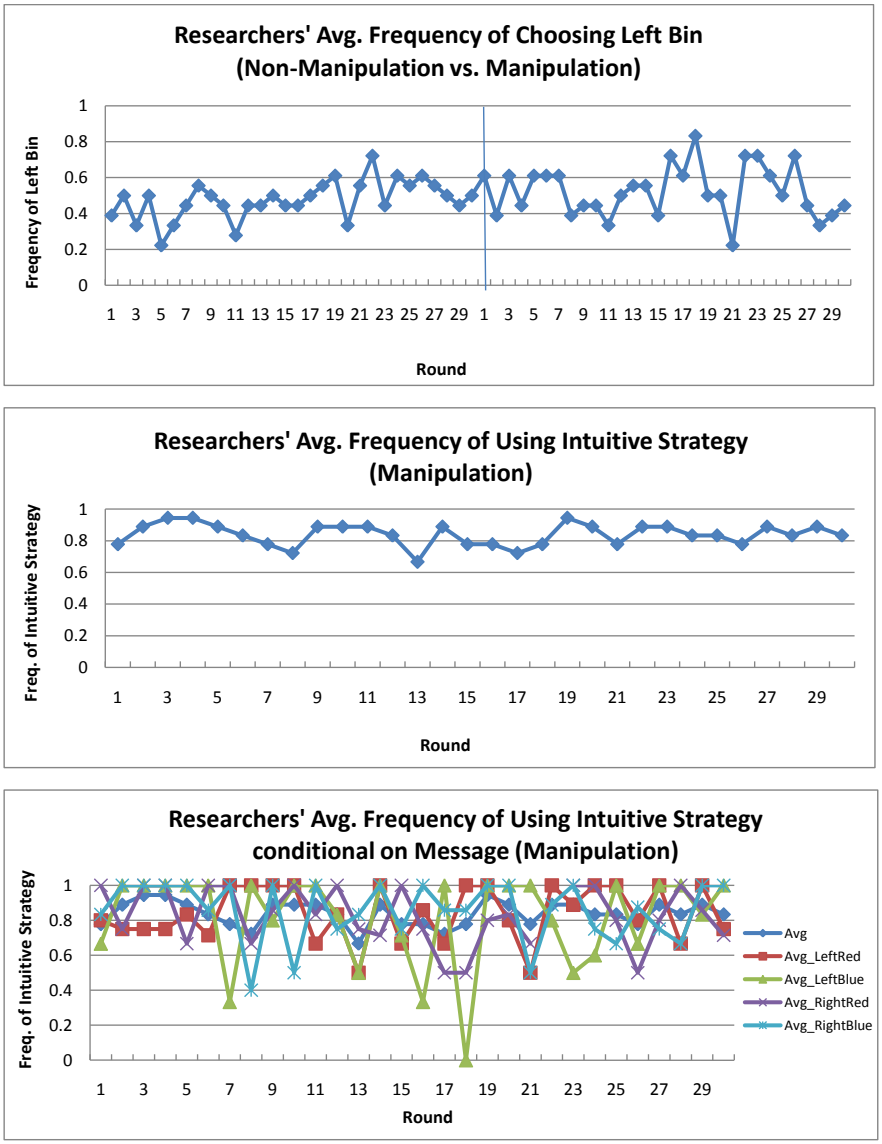

Figure 8: Researchers' Choices over Time

your own decisions, other participants' decisions and some random events which will be explained later. Before reading the details of the experiment, as general guidelines, please

- read the instructions carefully as they explain how you earn money from the decisions that you make;
- do not talk with other participants. In particular, do not discuss your decisions or your results with anyone at any time during the experiment;
- silence your mobile device during the experiment.

Part One Instructions

General Information

At the beginning of the experiment, half of the participants in the room will be randomly assigned as Player A and the other half as Player B. Your role will appear on the screen and will remain the same throughout the experiment. At the beginning of each round you will be randomly paired with another person who is assigned to the other role from your own. That is, if you are a Player A (Player B), in each round you will be randomly paired with a Player B (Player A) with all possible pairings being equally likely.

Specifics

In each round, you and your matched player will play a game as follows. There are two bins, LEFT BIN and RIGHT BIN. In each bin there are 50 balls of a same colour, RED or BLUE. The computer will randomly draw the colour of the balls in each bin at the beginning of each round, with 50% chance being RED and 50% chance being BLUE. The colour of the balls in each bin is drawn independently, that is, the colour of the balls in the LEFT BIN will not affect the colour of the balls in the RIGHT BIN, and vice versa. Therefore, although all the balls in the same bin are always in the same colour, balls in different bins can be in different colours. There are in total four possible outcomes for the colour of the balls in the two bins. For your convenience, the four outcomes and the total number of RED balls in each case is provided in the following table.

LEFT BIN	RIGHT BIN	Total RED balls in the two bins	Value of the project
RED	RED	100	100 points
RED	BLUE	50	50 points
BLUE	BLUE	0	0 points
BLUE	RED	50	50 points

The two bins together represent a project. Each RED ball has a value of 1 point and each BLUE ball has no value. Therefore, the value of the project is equal to the total number of RED balls in the two bins.

In each round, Player B will be assigned by the computer a K value, which is his endowed income for the round. The value of K can be equal to 10, 40, or 70 and it may vary across rounds and across players. Each Player B's value of K for the round will be shown on the screen of the matched Player A and Player B before they make decisions.

Right after the computer randomly draws the colour of the balls in each bin and prior to Player A and Player B making decisions, neither player will observe the colour of the balls in the two bins. Before Player B makes a decision, Player A will have to choose one bin, LEFT BIN or RIGHT BIN. Then the computer will reveal the colour of the balls in the chosen bin to the matched Player A and Player B. After that, Player B will have to make a decision on whether to IMPLEMENT or NOT IMPLEMENT the project by clicking one of the two buttons.

If Player B chooses IMPLEMENT, Player A will receive 100 points, and Player B will forgo his endowed income and his earnings for the round will be equal to the value of the project. Alternatively, if Player B chooses NOT IMPLEMENT, Player A will receive 0 points, and Player B will receive his endowed income K points.

At the end of each round, the computer will display the outcome of the round, including the K value of Player B, your choice, the colour of the balls in Player A's chosen bin, the total number of RED balls, the points you earn in the round. You click the "OK" button to proceed to the next round.

There are three practice rounds, where the objective is to get you familiar with the computer interface and the earnings calculation. Please note that the practice rounds are entirely for this purpose, and any earnings in the practice rounds will not contribute to your final payment at all. Once the practice rounds are over, the experimenter will announce "The official experiment begins now!" after which the official experiment starts. In the official experiment, there are in total 30 rounds for Part One.

A Brief Summary

First, your role as a Player A or a Player B will be randomly assigned at the beginning of the experiment. Your role will not change during the experiment.

Second, remember that after each round you will be matched randomly with a player whose role is different from yours. Therefore, the probability of you being matched with the same individual in two consecutive rounds is low.

Third, in each round the following events will happen in sequence for each pair of Player A and Player B:

1. The computer will randomly draw the colour of the balls in each bin, without informing either player of the colour;

2. Player B will be assigned a K value, which will be shown to both players;

3. Player A will choose one bin;

4. The colour of the balls in the bin chosen by Player A will be shown to both players;

5. Player B will choose whether or not to implement the project;

6. The results for the round will be displayed.

Earnings

You will receive $10 for showing up in the session. At the end of the experiment, the computer will randomly choose FOUR rounds, TWO out of 30 from part one and TWO out of 30 from part two, to determine your actual earnings. Each round has an equal probability to be chosen. Your earnings in each round are calculated in points, which will be converted to Canadian dollars at the exchange rate of 10 Points = 1 Dollar. Your final payment will be the summation of the earnings in the four randomly chosen rounds, plus the show-up fee. Please note that you will not be told which rounds are chosen before the end of the experiment, so you should make careful decisions in every round. You will be paid in cash, individually and privately, at the end of the experiment.

In the rare case, if your total payment is less than $15 including the show-up fee, you will receive $15 instead.

Questions?

Now is the time for questions. If you have any question, please raise your hand. Our experimenter will come to answer your question individually.

Part Two Instructions

From now on until the end of today's experiment, everything is the same as in the original instruction except the following part.

Before Player A makes the decision, the computer will randomly choose a bin, each bin being chosen with 50% chance. The computer will reveal the TRUE colour of the balls in that bin to Player A by sending a private message. The message will take the following form:

The balls in the LEFT/RIGHT BIN are RED/BLUE

The content of the message depends on the computer's choice between the two bins and the colour of the balls. Player B will not observe the content of the message.

Accordingly, in each round the following events will happen in sequence for each pair of Player A and Player B:

1. The computer will randomly draw the colour of the balls in each bin, without informing either player of the colour;

2. Player B will be assigned a K value, which will be shown to both players;

3. The computer will randomly choose one bin and inform Player A of the colour of the balls in that bin by sending a private message;

4. Player A will choose one bin;

5. The colour of the balls in the bin chosen by Player A will be shown to both players;

6. Player B will choose whether or not to implement the project;

7. The results for the round will be displayed.

There are 3 practice rounds and in total 30 official rounds in Part Two.

References

Allcott, H. (2015): "Site selection bias in program evaluation," *The Quarterly Journal of Economics*, 130(3), 1117–1165. 1

Au, P. H., and K. K. Li (2018): "Bayesian Persuasion and Reciprocity: Theory and Experiment," Discussion paper. 1

Blume, A., E. K. Lai, and W. Lim (2017): "Strategic Information Transmission:A Survey of Experiments and Theoretical Foundations," Discussion paper. 1

Brodeur, A., M. Lé, M. Sangnier, and Y. Zylberberg (2016): "Star Wars: The Empirics Strike Back," *American Economic Journal: Applied Economics*, 8(1), 1–32. 1

Chalmers, I. (1999): "Why transition from alternation to randomisation in clinical trials was made," *British Medical Journal*, 319(7221), 1372. 21

Chang, A. C., and P. Li (2017): "A Preanalysis Plan to Replicate Sixty Economics Research Papers That Worked Half of the Time," *American Economic Review*, 107(5), 60–64. 1

Chung, W., and R. Harbaugh (2016): "Biased Recommendations from Biased and Unbiased Experts," Discussion paper. 1

Crawford, V., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1452. 1

Di Tillio, A., M. Ottaviani, and P. N. Sørensen (2017a): "Persuasion bias in science: Can economics help?," *The Economic Journal*, 127(605), F266–F304. (document), 1, 7, 2, 2.2, 2.4, 5, 6, 21

——— (2017b): "Strategic sample selection," Discussion paper. 1

Fisher, R. A. (1925): *Statistical methods for research workers*. Oliver and Boyd. Edinburgh, Scotland. 21

——— (1926): "The arrangement of field experiments," *Journal of the Ministry of Agriculture of Great Britain*, 33, 503–513. 21

——— (1935): *The Design of Experiments*. Oliver and Boyd. Edinburgh, Scotland. 21

Fréchette, G., A. Lizzeri, and J. Perego (2017): "Rules and Commitment in Communication," Discussion paper, New York University. 1

Glaeser, E. L. (2008): "Researcher incentives and empirical methods," in *The foundations of positive and normative economics: A handbook*, ed. by A. Caplin, and A. Schotter, pp. 300–319. New York: Oxford University Press. 2

Hart, P. D. (1999): "A change in scientific approach: from alternation to randomised allocation in clinical trials in the 1940s," *Bmj*, 319(7209), 572–573. 21

Head, M. L., L. Holman, R. Lanfear, A. T. Kahn, and M. D. Jennions (2015): "The extent and consequences of p-hacking in science," *PLoS biology*, 13(3), 1–15. 1

Hoffmann, F., R. Inderst, and M. Ottaviani (2014): "Persuasion through selective disclosure: Implications for marketing, campaigning, and privacy regulation," Discussion paper, JW Goethe University Frankfurt, University College London, and Bocconi University. 4

Holt, C. A., and S. K. Laury (2002): "Risk Aversion and Incentive Effects," *American Economic Review*, 92(5), 1644–1655. 14

Imbens, G. W., and D. B. Rubin (2015): *Causal inference in statistics, social, and biomedical sciences*. New York: Cambridge University Press. 3

Jin, G. Z., M. Luca, and D. Martin (2015): "Is no news (perceived as) bad news? An experimental investigation of information disclosure," Discussion paper, National Bureau of Economic Research. 8

Jump, R. (2011): "A Star's Collapse," *Times Higher Education*, November 28. 1

Kamenica, E., and M. Gentzkow (2011): "Bayesian Persuasion," *American Economic Review*, 101(6), 2590–2615. 1, 7

Kandori, M. (1992): "Social Norms and Community Enforcement," *Review of Economic Studies*, 59(1), 63–80. 15

Kearns, C. E., L. A. Schmidt, and S. A. Glantz (2016): "Sugar industry and coronary heart disease research: A historical analysis of internal industry documents," *JAMA Internal Medicine*, 176(11), 1680–1685. 1

Kolata, G. (2018): "Harvard Calls for Retraction of Dozens of Studies by Noted Cardiac Researcher," *New York Times*, October 15. 1

Kolotilin, A., T. Mylovanov, A. Zapechelnyuk, and M. Li (2017): "Persuasion of a privately informed receiver," *Econometrica*, 85(6), 1949–1964. 7

Leamer, E. (1983): "Let's Take the Con Out of Econometrics," *American Economic Review*, 73(1), 31–43. 2

Min, D. (2017): "Screening for Experiments," Discussion paper, University of Arizona. 1

Neyman, J. S. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. (Translated and edited by D.M. Dabrowska and T. P. Speed, Statistical Science (1990), 5(4), 465-480.)," *Annals of Agricultural Sciences*, 10, 1–51. 1, 21

Nguyen, Q. (2017): "Bayesian Persuasion: Evidence from the Laboratory," Discussion paper, Utah State University. 1

O'Connor, A. (2016): "How the sugar industry shifted blame to fat," *New York Times*, September 12. 1

Open Science Collaboration, (2015): "Estimating the reproducibility of psychological science," *Science*, 349(6251), aac4716. 1

Rayo, L., and I. Segal (2010): "Optimal Information Disclosure," *Journal of Political Economy*, 118(5), 949 – 987. 7

Rosenberger, W. F., and J. M. Lachin (2015): *Randomization in clinical trials: Theory and practice*. New York: John Wiley & Sons. 3

Rubin, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of educational Psychology*, 66(5), 688–701. 1

Simonsohn, U., L. D. Nelson, and J. P. Simmons (2014): "P-curve: A key to the file-drawer," *Journal of Experimental Psychology: General*, 143(2), 534–547. 1

Yoder, N. (2016): "Designing Incentives for Academic Research," Discussion paper, University of Georgia. 1