Ethics of randomized field experiments:
Evidence from a randomized survey experiment
Online Appendix

Contents

Appendix A. Survey Data Collection in Detail

As mentioned in the paper, the first survey was designed by the author and implemented in Japan by the survey company INTAGE Research Inc. in March 2017. This company maintains a panel of respondents and undertakes online surveys. In my study, potential respondents in the panel were randomly selected with weights to create a representative Japanese sample in terms of residential area, gender, and age group. Those who joined the survey were paid if they fully completed the survey, although the author did not share information about the exact pay for this survey.

Individuals in panel are called *cue monitors*. The number of cue monitors reside in Japan as of January 2017 was 1.41 million. In general, they are paid in "cue monitor points" for participating online surveys. Those points can be exchanged for Amazon gift cards, electronic money, or vouchers at a rate of JPY 1.00 for one point. The points per survey depend on the survey.

The survey requests were sent from 2 p.m. on Friday (March 10, 2017) to 6,698 candidates. From the set number of respondents, only valid respondents based on INTAGE Research's determination standard were left, resulting in 2,107 respondents, which were 31.5% of the number of candidates who received the request. The first survey finished at 9 a.m. on March 13, 2017.

Appendix B. Descriptions Used in the First Survey

B1. Description of Fryer et al. (2015)

Study on a preschool
Recent findings show that the care and education one receives in early childhood affect one's academic achievement and lifetime earnings in adulthood. Following these findings, Professor X established a preschool in a low-income area.
Overview of the preschool:
  – *The preschool is free of charge.*
  – *This preschool uses a curriculum called "Tools of the Mind" to foster patience and social skills.*
  – *Inside the preschool is similar to a small "town," where one can experience various types of jobs.*
  – *Children of this preschool are surveyed periodically.*
  – *Followup surveys are planned for every few years following graduation.*
Professor X called for applicants to this preschool.
Overview of admissions:
  – *Parents and children, for a total of 140 families, applied for admission.*
  – *Only 70 children selected based on a lottery were admitted.*
  – *The remaining 70 children were not able to enroll in the preschool.*
  – *However, the children who were not able to enroll, as well as their parents, are regularly invited to parties held on holidays.*
After the preschool was opened, Professor X invited the children who were enrolled and their parents—as well as the children who were not able to enroll and their parents—to regularly held parties and surveyed them. The surveys were periodically conducted for over 10 years, even after the children entered primary school. Finally, Professor X conducted a study comparing children who attended the preschool with those who were not able to enroll. Note that the parents of the 140 children who became subjects of the study received an explanation regarding them being the subjects of the study, and they gave their consent.

B2. Description of Thornton (2008)

Study on HIV testing
The AIDS epidemic is a serious issue in the developing world, such as African countries. One of the reasons for this is that there are people who have sexual intercourse with multiple, unspecified partners while unaware of their HIV-positive status.
To prevent the AIDS epidemic, Professor X conducted a campaign in a developing country. The campaign offered a chance of earning reward equivalent to JPY 20 to those who accepted HIV testing and learned the results. In this country, JPY 20 is the amount one can earn doing agricultural work in one day.
The content of the campaign was as follows:
  – *HIV testing was free, and anyone could participate.*
  – *Of 3,000 adults in a certain region, only 1,500 who were randomly selected by a computer were involved in the campaign.*
  – *The remaining 1,500 adults could not take part.*
  – *Among the eligible adults, only those who underwent HIV testing and went to testing centers to learn about the results were awarded an amount of money equivalent to JPY 20.*
After the campaign, Professor X surveyed those who were—as well as those who were not— selected for participation in the campaign about whether they had received HIV testing and had gone to testing centers to be notified of the results. Finally, Professor X conducted a study comparing those who were and were not selected for their decision to learn HIV results. Note that the 3,000 adults gave their consent following an explanation that the survey results on their HIV testing status would be used for a certain research objective. However, they were not told that a comparison was made based on the presence/absence of a chance in reward.

B3. Description of Landry et al. (2006)

Study on charitable giving
Researchers sometimes try to obtain donations from citizens to conduct studies that contribute to society. To obtain more donations, Professor X came up with the idea of "offering donors a chance to win a prize in a raffle."
To examine whether this idea actually increases the amount of donations, Professor X conducted a fundraising project for "Natural Hazard Mitigation Research" in an area with 4,800 households.
Overview of the project:
  – *Donations were collected through door-to-door visits for all households.*
  – *Only 2,400 households who were randomly selected by a computer were asked to donate with a raffle in which one among all donors could win JPY 100,000 (households with a prize).*
  – *The remaining 2,400 households were asked to donate without the chance of winning in the raffle (households without a prize).*
  – *One of the "households with a prize" that actually donated money won JPY 100,000 in the raffle.*
  – *Collected donations were actually used for "Natural Hazard Mitigation Research."*
After the fundraising activities, Professor X counted all the donations from both the "households with a prize" and "households without a prize" and then compared the two groups. Note that the 4,800 households that were solicited for donations were not informed of their involvement in the study.

B4. Description of Allcott (2011)

Study on electricity conservation
Professor X prepared reports that present effective strategies to reduce the electricity bill of house-
holds. These "Home Energy Reports" contain data on average household electricity usage and
tips to conserve it. To examine the effectiveness of these reports in terms of reducing electricity
consumption, a project was run in an area populated by 40,000 households.
Overview of the project:
  – *The "Home Energy Reports" were mailed to only 20,000 households that were randomly se-*
    lected by a computer.
  – *The reports were mailed once a month for three months (3 times total).*
  – *The remaining 20,000 households did not receive the reports at all.*
Four months after the manuals started to be mailed out, an electric utility company compared usage
between households who were mailed the manuals three times and those to whom the manuals were
not mailed. Note that the 40,000 households were not informed of their involvement in the study.

B5. Description of Hanna et al. (2016)

Study on smoke from cooking stoves
In many developing countries, each household has a stove for cooking. Note that those stoves emit
smoke during cooking. This has become a social problem since women and children become sick by
inhaling smoke.
To tackle this problem, Professor X carried out a project in a developing country, whereby "improved
cooking stoves" were constructed free of charge in an area with 1,600 households.
Overview of the project:
  – *Stove construction was carried out over two periods over 5 years.*
  – *For the first three years, stoves were built for only 800 households selected by a lottery.*
  – *The remaining 800 households awaited their turn.*
At the end of the third year, Professor X observed the health status of both for those whom "improved
cooking stoves" were constructed and for those whom these stoves had not been built. Finally,
Professor X compared the health status of the two groups. Note that the 1,600 households were
informed that they were involved in a study. However, they were not informed that they were
compared based on the presence/absence of the improved stove.

B6. Description of Hosono and Aoyagi (2018)

Study on recyclable waste sorting
An increasing amount of waste has become a social problem. To tackle this problem, it has been
suggested to decrease the amount of waste by sorting and recycling it.
Professor X carried out a project using stamps and gifts in a poor region populated by 500 households
in a developing country to increase household waste sorting.
Overview of the project:
  – *Households can get a stamp on their card if recyclable waste is sorted upon disposal.*
  – *Households were gifted with laundry detergent if they gathered a certain number of stamps.*
  – *Due to budgetary reasons, cards for collecting stamps were distributed to only 250 households*
    that were randomly selected by a computer.
  – *The remaining 250 households could not participate in collecting stamps.*

Two months after the stamp collection began, Professor X surveyed household waste disposed of by those who received the stamp card and who did not. Then, Professor X examined whether they had sorted recyclable waste from other waste. Furthermore, Professor X measured the weight of waste by type (cans, plastic, etc.) and compared the weight between the two groups. Note that the 500 households received an explanation regarding their involvement and were informed that their household waste was weighed. However, they were not informed that they were compared based on whether they had the stamp card.

Appendix C. Remarks on the Modifications Made to the Original Experiments

C1. Modifications made to Fryer et al. (2015)

There are at least two remarks on the modifications made to the original experiments of Fryer et al. (2015). First, while the project includes two treatments, I focused on the preschool and made the number of groups two to make the survey simple and short. Note that, according to Gneezy and List (2013) and Cappelen et al. (2020), children in the preschool treatment are further randomized to either the *Literacy Express* curriculum or to the *Tools of the Mind* curriculum. Again, for simplification, I focused on only one of them (the latter). Second, I made the group size 70 following information provided in Gneezy and List (2013). This figure is similar to the size of the analytical sample of Fryer et al. (2015) and Cappelen et al. (2020). For an implementer of the CHECC project, I anonymized and framed it as "Professor X."

C2. Modifications made to Thornton (2008)

In the present study, I focus on the behavior of learning HIV status to simplify the description of Thornton (2008). Note that, however, Thornton (2008) also studies other behaviors, such as the purchase of condoms. In addition, the experiment of Thornton (2008) created random variation in the distance to the HIV results center. This design is also abstracted from the description used in the present study. Moreover, while there were more than two variations in incentives provided in Thornton (2008), I reduce these variations to two groups—a control and a treatment—for simplification.[1]

C3. Modifications made to Landry et al. (2006)

See Section 3.2 of the paper.

C4. Modifications made to Allcott (2011)

Allcott (2011) pools observations from 17 experiments that include approximately 600,000 households in total. However, in the present study, I described the sample size as "*40,000 households*," which is approximately equivalent to the average sample size of the 17 experiments. This size of the experiment is to some extent comparable with the HER experiment conducted in Japan in 2015 (Jyukankyo Research Institute Inc., 2016).

---

[1] In Thornton (2008), subjects were given randomly assigned vouchers between zero and three dollars, redeemable upon obtaining their test results at a nearby center.

C5. Modifications made to Hanna et al. (2016)

While Hanna et al. (2016) collected data on approximately 2,500 households and randomly divided them into three groups using a public lottery, in the present study, I simplified my description to two groups of 1,600 households.

C6. Modifications made to Hosono and Aoyagi (2018)

In Hosono and Aoyagi (2018), three treatments are evaluated to encourage the sorting of recyclable waste (e.g., plastics and aluminum) from other garbage: free distribution of buckets to store recyclables, face-to-face persuasive communication, and in-kind incentives. In the present study, I focus on in-kind incentive treatment and control groups. Households in the treatment group can obtain a stamp on their card if they dispose of recyclable waste separately from other garbage, and they can obtain laundry detergent if they collect ten stamps.

Appendix D. Descriptions Used in the Second Survey

D1. Description of Treatment 2 in the survey for Fryer et al. (2015)

Study on a preschool
Recent findings show that the care and education one receives in early childhood affect one's academic achievement and lifetime earnings in adulthood. Following these findings, Professor X established a preschool in a low-income area.
Overview of the preschool:
  – The preschool is free of charge.
  – This preschool uses a curriculum called "Tools of the Mind" to foster patience and social skills.
  – Inside the preschool is similar to a small "town," where one can experience various types of jobs.
  – Children of this preschool are surveyed periodically.
  – Followup surveys are planned for every few years following graduation.
Overview of the selection:
  – Parents and their children from 140 families living in the area are defined as the research subjects.
  – Only 70 children selected based on a lottery were admitted to enroll in the preschool.
  – The remaining 70 children were not able to enroll in the preschool.
  – However, the children who were not able to enroll, as well as their parents, were regularly invited to parties held on holidays.
After the preschool was opened, Professor X invited the children who were enrolled and their parents—as well as the children who were not able to enroll and their parents—to regularly held parties and surveyed them. The surveys were periodically conducted for over 10 years, even after the children entered primary school. Finally, Professor X conducted a study comparing children who attended the preschool with those who were not able to enroll. Note that the parents of the 140 children who became subjects of the study received an explanation regarding them being the subjects of the study, and they gave their consent.

D2. Description of Treatment 3 in the survey for Fryer et al. (2015)

Study on a preschool

Recent findings show that the care and education one receives in early childhood affect one's academic achievement and lifetime earnings in adulthood. Following these findings, Professor X established a preschool in a low-income area.

Overview of the preschool:
– The preschool is free of charge.
– This preschool uses a curriculum called "Tools of the Mind" to foster patience and social skills.
– Inside the preschool is similar to a small "town," where one can experience various types of jobs.
– Children of this preschool were surveyed periodically.
– Followup surveys were planned for every few years following graduation.

Professor X called for applicants to this preschool.

Overview of admissions:
– Parents and their children in 140 families applied for admission.
– Only 70 children selected based on a lottery were admitted.
– The remaining 70 children were not able to enroll in the preschool.

After the preschool was opened, Professor X conducted periodical surveys for the children who were enrolled and their parents as well as the children who were not able to enroll and their parents. The surveys were periodically conducted for over 10 years, even after the children entered primary school. Finally, Professor X conducted a study comparing children who attended the preschool with those who were not able to enroll. Note that the parents of the 140 children who became subjects of the study received an explanation regarding them being the subjects of the study, and they gave their consent.

D3. Description of Treatment 1 in the survey for Landry et al. (2006)

Study on charitable giving

Researchers sometimes try to obtain donations from citizens to conduct studies that contribute to society. To obtain more donations, Professor X came up with the idea of "offering donors a chance to win a prize in a raffle."

To examine whether this idea actually increases the amount of donations, Professor X conducted a fundraising project for "Natural Hazard Mitigation Research" in an area with 4,800 households.

Overview of the project:
– Donations were collected through door-to-door visits for all households.
– One year later, donations were again collected through door-to-door visits.
– In the door-to-door visits in the first year, the widely used practice of solicitation was used (no prize phase).
– In the visits in the second year, households were asked to donate with a raffle in which one donor could win JPY 100,000 (prize phase).
– One of the households that actually donated in the "prize phase" won JPY 100,000 in the raffle.
– Collected donations were actually used for "Natural Hazard Mitigation Research."

After the fundraising activities, Professor X counted all the donations in both the "no prize phase" and "prize phase" and then compared the two phases. Note that the 4,800 households that were solicited for donations were not informed of their involvement in the study.

D4. Description of Treatment 2 in the survey for Landry et al. (2006)

Study on charitable giving
Researchers sometimes try to obtain donations from citizens to conduct studies that contribute to society. To obtain more donations, Professor X came up with the idea of "telling donors the result of solicitation in neighboring town."
To examine whether this idea actually increases the amount of donations, Professor X conducted a fundraising project for "Natural Hazard Mitigation Research" in an area with 4,800 households.
Overview of the project:
  - *Donations were collected through door-to-door visits for all households.*
  - *Only 2,400 households who were randomly selected by a computer were asked to donate through a flyer that states the following: "In the neighboring town, 80% of the households donated" (households with a message).*
  - *The remaining 2,400 households were asked to donate using a flyer without this message (households without a message).*
  - *Collected donations are actually used for "Natural Hazard Mitigation Research."*
After the fundraising activities, Professor X counted all the donations from both the "households with a message" and "households without a message" groups and then compared them. Note that the 4,800 households that were solicited for donations were not informed of their involvement in the study.

D5. Description of Treatment 3 in the survey for Landry et al. (2006)

Study on recyclable waste sorting
The increasing cost of solid waste management has become a social problem. To tackle this problem, it has been suggested to separate food waste from garbage and recycle it to decrease the amount of waste. To increase sorting food waste at home, Professor X came up with the idea of "offering recyclers a chance to win a prize in a raffle."
To examine whether this idea actually increases the amount of food waste sorted, Professor X conducted a recycling campaign project in an area with 4,800 households.
Overview of the project:
  - *In cooperation with the city government, the municipal collection of food waste separately from garbage was begun.*
  - *Sorting of food waste was solicited through door-to-door visits for all households.*
  - *Only 2,400 households who were randomly selected by a computer were asked to sort with a raffle in which one among all recyclers could win JPY 100,000 (households with a prize).*
  - *The remaining 2,400 households were asked to sort without the chance of winning in the raffle (households without a prize).*
  - *One of the "households with a prize" that actually sorted food waste won JPY 100,000 in the raffle.*
  - *Collected food waste was composted and used by farmers in the area.*
After the campaign, Professor X measured the amount of sorted waste for both the "households with a prize" and "households without a prize" groups and then compared them. Note that the 4,800 households that were involved in the sorting campaign were not informed of their involvement in the study.

Appendix E. Additional Figures

保育園の研究

幼児期にどのような保育・教育を受けたかが、成人になってからの学力や生涯年収に影響を
与えるという研究成果の報告が増えています。そこで、ある貧困地域で研究者Xさんが保育園
を立ち上げました。

保育園の概要:
− 保育料は無料です。
− この保育園では「心の道具箱」というカリキュラムを使って、人付き合いや辛抱することを育
みます。
− 保育園の中は小さな「街」のようになっていて、様々な仕事を体験できます。
− 園児は定期的な調査の対象となります。
− さらに、卒園後も数年ごとに追跡調査が行われる予定です。

そして、この保育園の入園児の募集をしました。

募集の様子:
− 140組の親子が入園を希望して応募しました。
− 抽選で当選した70名の幼児だけが入園を許可されました。
− 残りの70名の幼児は入園できませんでした。
− ただし、入園できなかった幼児とその親は、祝日に開かれるイベントに定期的に招待される
ことになりました。

保育園の開園後、定期的に開かれるイベントに園児70名とその親、そして、落選した70名の
幼児とその親が招待され、その場で調査が行われました。この調査は幼児の小学校入学後も
、10年以上に渡って定期的に行われました。そして、研究者Xさんが立ち上げた保育園に通っ
た幼児と入園できなかった幼児を比較する研究を行いました。
なお、対象となった140名の幼児の親は、自分たちの子供が研究対象となることの説明を受け
、承諾していました。

Q5　この研究は倫理的に問題があると感じますか。

〈回答は1つ〉

| 大いに 問題がある | やや 問題がある | どちらとも 言えない | ほぼ 問題がない | 全く 問題がない |

Figure A1: Screenshot of the survey

Notes: This figure shows a screenshot of the survey. Fryer, Levitt, and List (2015) is shown (in Japanese).
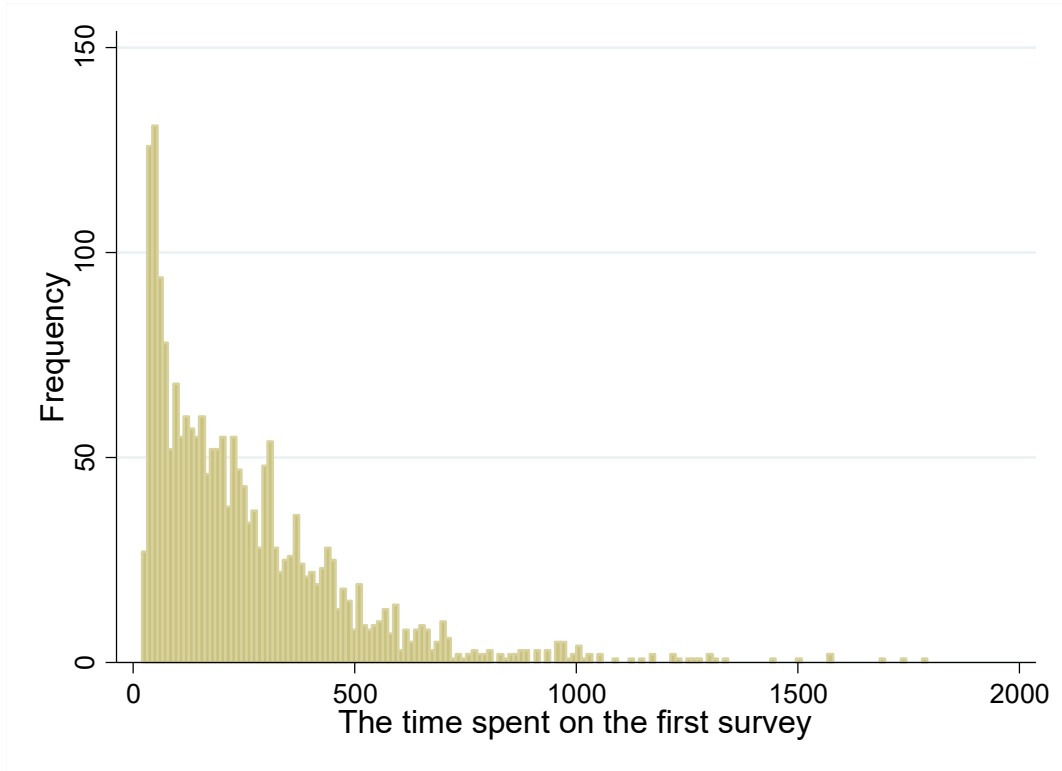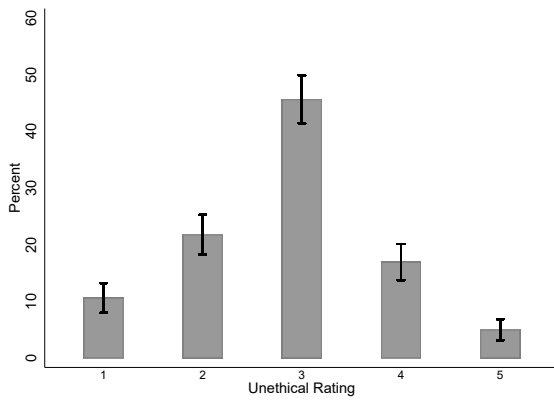
Figure A2: Distribution of time spent on the first survey

Notes: This figure shows the distribution of the time spent on the first survey. The vertical axis shows the density. The horizontal axis shows the time in seconds. This figure only shows the distribution shorter than 1,800 seconds, while the maximum value was 184,468 seconds.

Panel A. Control

Panel B. Treatment 1: No informed consent
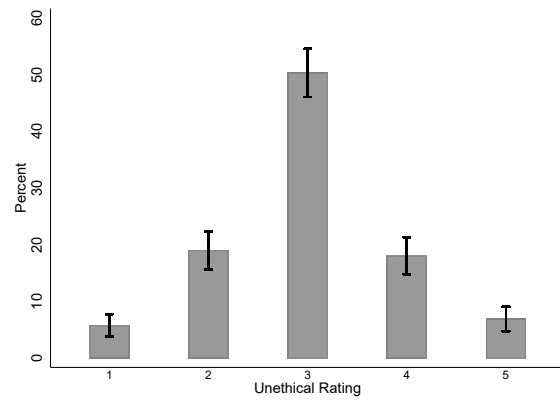
Panel C. Treatment 2: Sampling by a researcher
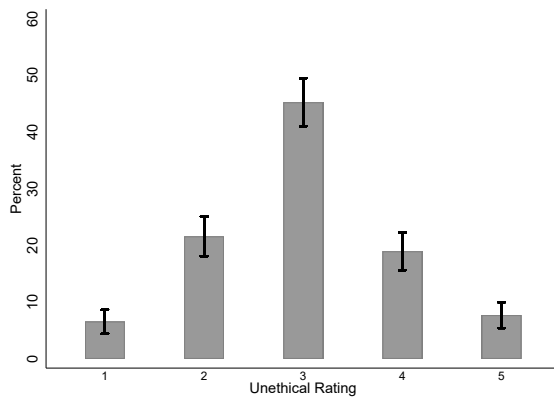
Panel D. Treatment 3: No holiday party
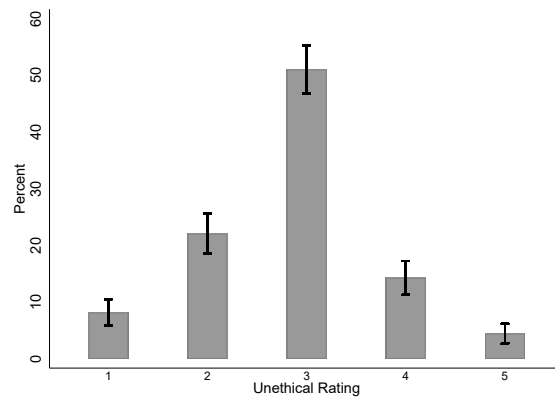
| 1: There is no ethical issue at all | 2: There is almost no ethical issue |
| 3: Neutral | 4: There is a slight ethical issue | 5: There is a major ethical issue |

Figure A3: Response to the randomized survey (Fryer, Levitt, and List, 2015)

Notes: This figure shows the distribution (percentages) of the survey response to the question "Do you recognize any ethical issues in this study?" for the four descriptions based on Fryer, Levitt, and List (2015). The vertical bars and caps are 95 % confidence intervals.

Panel A. Control
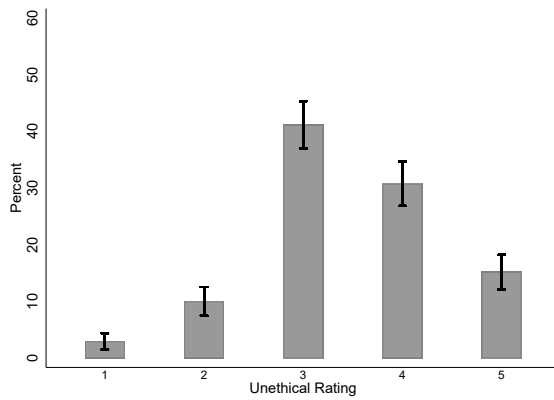
Panel B. Treatment 1: Before–after study
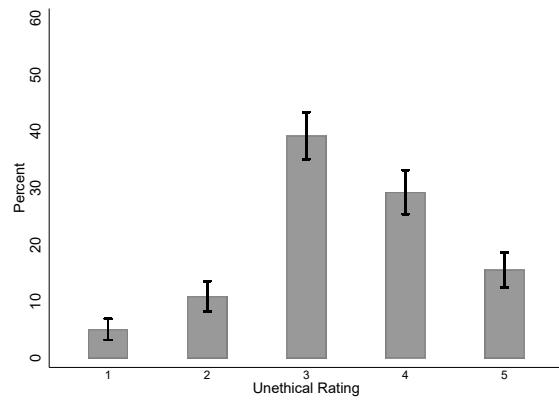
Panel C. Treatment 2: Social comparison message
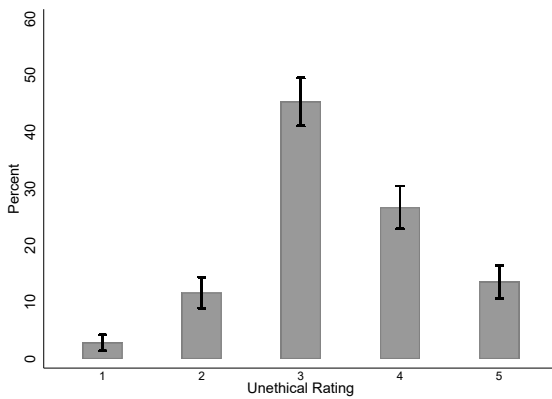
Panel D. Treatment 3: Waste sorting
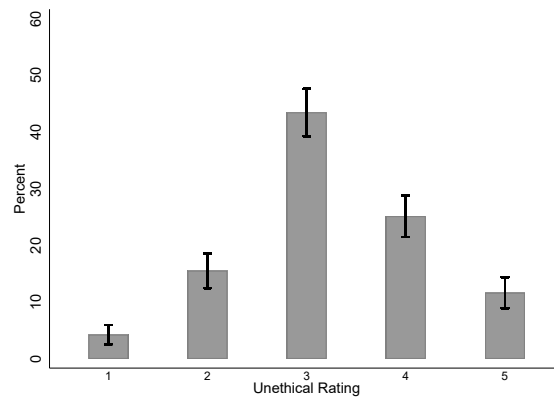
1: There is no ethical issue at all	2: There is almost no ethical issue
3: Neutral 4: There is a slight ethical issue	5: There is a major ethical issue

Figure A4: Response to the randomized survey (Landry et al., 2006)

Notes: This figure shows the distribution (percentages) of the survey response to the question "Do you recognize any ethical issues in this study?" for the four descriptions based on Landry et al. (2006). The vertical bars and caps are 95 % confidence intervals.

Figure A3 shows the distribution of the responses to the survey on Fryer et al. (2015). Panel A shows the responses in a control group, where the distribution is quite similar to that in the first survey (see Panel A in Figure 1). This result shows that I succeeded in replicating the first survey one year later with different samples from the same country. Panels B, C, and D show the responses in the treatment 1, 2, and 3 groups, respectively. All the treatments decrease the recognition of no ethical issue while increasing neutral responses.

Figure A4 shows the distribution of the responses to the survey on Landry et al. (2006). Again, the response in the control group (Panel A) is similar to the response in the first survey (Panel C in Figure 1), meaning that the result is replicated. Treatments 1 (Panel B) and 2 (Panel C) are similar to the control, while treatment 3 (Panel D) slightly shifts the distribution to the left.[2]

---

[2] Note that the distribution of the responses to the waste-sorting version of Landry et al. (2006)(Figure A4 Panel D) is closer to that of Hosono and Aoyagi (2018)(Figure 1 Panel F) rather than that of Landry et al. (2006) in the first survey (Figure 1 Panel C).

Appendix F. Additional Tables

Table A1: Survey response time and characteristics (OLS)

	(1)	(2)
Dependent variable: Time spent on the first survey (seconds)		
Female	-655.751	-1022.648
	(713.732)	(648.026)
Age	21.759	8.542
	(21.392)	(22.982)
Married	-379.050	-547.440
	(425.457)	(419.803)
Living with children	1200.081**	1251.696**
	(585.969)	(568.889)
Full-time employee	660.288	716.329
	(648.012)	(627.491)
Part-time employee	1553.957	2540.860*
	(1048.779)	(1325.376)
Temporary/contract employee	-318.174	232.391
	(524.304)	(561.940)
Self-employed	-154.851	679.092
	(585.592)	(621.967)
Housewife/househusband	1393.794	1736.837
	(1106.674)	(1069.023)
Unemployed/retired	-550.859	13.776
	(414.779)	(579.761)
Household income		0.309
(10 thousand yen)		(1.142)
Constant	-1087.996	-514.549
	(1019.103)	(737.741)
Province dummy variables	Yes	Yes
Observations	2107	1645
Pseudo-R^2/ R^2	0.033	0.051

Notes: This table reports the estimates from linear regression analyses on the association between time spent for the first survey and characteristics of respondents. The median of the dependent variable is 205 seconds (3.4 minutes), and the average is 1,374 seconds (23 minutes). The coefficients are reported. Robust standard errors are in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table A2: Summary statistics of the randomized survey experiments by group (Fryer et al., 2015)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	C	T1	T2	T3	P-value		
					C vs T1	C vs T2	C vs T3
Female	0.488	0.486	0.516	0.471	0.951	0.359	0.585
	(0.500)	(0.500)	(0.500)	(0.500)			
Age	46.153	47.030	46.835	47.664	0.286	0.408	0.062
	(13.415)	(13.528)	(13.532)	(13.156)			
Married	0.603	0.615	0.612	0.633	0.708	0.782	0.323
	(0.490)	(0.487)	(0.488)	(0.483)			
Living with children	0.378	0.384	0.371	0.369	0.851	0.825	0.765
	(0.485)	(0.487)	(0.484)	(0.483)			
Household income	553.589	551.651	537.681	564.353	0.915	0.376	0.558
(10 thousand yen)	(265.456)	(261.062)	(252.768)	(268.040)			
Full-time employee	0.225	0.229	0.242	0.223	0.884	0.519	0.916
	(0.418)	(0.421)	(0.429)	(0.416)			
Part-time employee	0.130	0.151	0.129	0.148	0.335	0.965	0.393
	(0.337)	(0.358)	(0.336)	(0.356)			
Temporary/contract employee	0.061	0.047	0.058	0.072	0.281	0.821	0.475
	(0.240)	(0.211)	(0.234)	(0.259)			
Self-employed	0.039	0.069	0.049	0.056	0.031	0.441	0.202
	(0.194)	(0.254)	(0.216)	(0.229)			
Housewife/househusband	0.175	0.181	0.225	0.178	0.811	0.041	0.895
	(0.380)	(0.385)	(0.418)	(0.383)			
Unemployed/retired	0.115	0.110	0.094	0.111	0.772	0.248	0.831
	(0.320)	(0.313)	(0.292)	(0.315)			
Living in Tokyo	0.132	0.115	0.126	0.128	0.405	0.751	0.838
	(0.339)	(0.320)	(0.332)	(0.334)			
Living in Osaka	0.060	0.067	0.064	0.076	0.617	0.776	0.283
	(0.237)	(0.250)	(0.245)	(0.265)			

Notes: This table reports the means for each of the groups for Fryer et al. (2015) with standard deviations in parentheses. Columns 5–7 report p-values for the differences. The number of observations is 2,146, except for Household income (the number of observations is 1,681).

Table A3: Summary statistics of the randomized survey experiments by group (Landry et al., 2006)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	C	T1	T2	T3	P-value		
					C vs T1	C vs T2	C vs T3
Female	0.481	0.482	0.515	0.482	0.982	0.272	0.975
	(0.500)	(0.500)	(0.500)	(0.500)			
Age	47.019	47.028	46.254	47.375	0.991	0.353	0.661
	(13.411)	(13.373)	(13.532)	(13.335)			
Married	0.602	0.632	0.607	0.621	0.306	0.860	0.517
	(0.490)	(0.483)	(0.489)	(0.486)			
Living with children	0.354	0.405	0.353	0.390	0.082	0.991	0.217
	(0.479)	(0.491)	(0.478)	(0.488)			
Household income	553.800	545.192	543.112	565.366	0.633	0.554	0.526
(10 thousand yen)	(263.917)	(256.813)	(260.631)	(266.316)			
Full-time employee	0.219	0.216	0.246	0.238	0.913	0.283	0.436
	(0.414)	(0.412)	(0.431)	(0.427)			
Part-time employee	0.139	0.135	0.145	0.140	0.856	0.784	0.940
	(0.346)	(0.342)	(0.352)	(0.348)			
Temporary/contract employee	0.067	0.068	0.047	0.057	0.954	0.165	0.524
	(0.250)	(0.251)	(0.212)	(0.233)			
Self-employed	0.061	0.069	0.038	0.044	0.582	0.076	0.218
	(0.240)	(0.254)	(0.190)	(0.206)			
Housewife/househusband	0.156	0.184	0.224	0.196	0.217	0.004	0.081
	(0.363)	(0.388)	(0.417)	(0.397)			
Unemployed/retired	0.122	0.116	0.098	0.094	0.766	0.201	0.139
	(0.328)	(0.321)	(0.297)	(0.292)			
Living in Tokyo	0.135	0.105	0.135	0.126	0.130	0.994	0.643
	(0.342)	(0.307)	(0.342)	(0.332)			
Living in Osaka	0.063	0.062	0.073	0.068	0.943	0.502	0.719
	(0.243)	(0.241)	(0.261)	(0.253)			

Notes: This table reports the means for each of the groups for Landry et al. (2006) with standard deviations in parentheses. Columns 5–7 report p-values for the differences. The number of observations is 2,146, except for Household income (the number of observations is 1,681).

Table A4: Comparisons of the six studies for respondents who spent a long time on the survey (coefficients)

	Ordered logit		OLS	
	(1)	(2)	(3)	(4)
Fryer et al. (2015)	-0.453***	-0.443***	-0.248***	-0.241***
	(0.095)	(0.095)	(0.053)	(0.052)
Thornton (2008)	0.055	0.054	0.045	0.045
	(0.092)	(0.092)	(0.051)	(0.051)
Landry et al. (2006)	0.720***	0.729***	0.411***	0.412***
	(0.093)	(0.092)	(0.052)	(0.051)
Allcott (2011)	-0.292***	-0.294***	-0.145**	-0.150***
	(0.102)	(0.101)	(0.057)	(0.056)
Hanna et al. (2016)	0.072	0.079	0.045	0.049
	(0.116)	(0.116)	(0.065)	(0.064)
Order (1–6)	-0.075***	-0.076***	-0.039***	-0.040***
	(0.013)	(0.014)	(0.008)	(0.007)
Female		0.328***		0.184***
		(0.078)		(0.043)
Age		0.013***		0.007***
		(0.003)		(0.001)
Constant			3.224***	2.787***
			(0.051)	(0.092)
Control *implementers*	Yes	Yes	Yes	Yes
Other control variables	No	Yes	No	Yes
Observations	5670	5670	5670	5670
Pseudo-R^2/ R^2	0.015	0.022	0.043	0.063

Notes: This table reports the estimates from the same regression analyses as Table 3 but dropping samples where time spent on the survey is in the bottom 10% (shorter than 49 seconds). Five studies are compared to Hosono and Aoyagi (2018). The coefficients are reported. Standard errors, clustered at the respondent level, are in parentheses. Columns 2 and 4 include other variables in Table 2 as well as *Time spent on the survey*. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table A5: Results of the randomized survey experiment with interactions over time (Fryer, Levitt, and List, 2015)

	(1)	(2)	(3)	(4)
	Ordered logit		OLS	
T1: Deleting the informed consent statement	0.019	0.017	0.053	0.052
	(0.157)	(0.158)	(0.072)	(0.072)
	[0.886]	[0.899]	[0.466]	[0.481]
T2: Samples are selected rather than self-selection	-0.036	-0.040	0.023	0.019
	(0.154)	(0.155)	(0.073)	(0.073)
	[0.778]	[0.752]	[0.760]	[0.796]
T3: Deleting holiday parties statement	-0.207*	-0.207*	-0.077	-0.079
	(0.152)	(0.152)	(0.067)	(0.067)
	[0.092]	[0.094]	[0.251]	[0.241]
Long time	-0.938***	-0.939***	-0.397***	-0.397***
	(0.162)	(0.162)	(0.084)	(0.084)
T1 × Long time	0.644***	0.647***	0.242**	0.242**
	(0.227)	(0.227)	(0.116)	(0.116)
	[0.006]	[0.006]	[0.036]	[0.037]
T2 × Long time	0.633***	0.637***	0.254**	0.256**
	(0.230)	(0.230)	(0.121)	(0.121)
	[0.008]	[0.006]	[0.036]	[0.034]
T3 × Long time	0.385*	0.386*	0.147	0.148
	(0.226)	(0.226)	(0.116)	(0.116)
	[0.098]	[0.096]	[0.204]	[0.201]
Order (1/2)		-0.195**		-0.077*
		(0.080)		(0.041)
Constant			3.042***	3.157***
			(0.052)	(0.079)
Multiple-Hypothesis Testing	0.031	0.032	0.162	0.153
Observations	2146	2146	2146	2146
Pseudo-R^2/ R^2	0.011	0.012	0.025	0.026

Notes: This table reports the estimates from regression analyses where the dummy variable of time spent on the survey being longer than the median is incorporated into the analysis of Table 4. The coefficients are reported. Standard errors are in parentheses in columns 1 and 2. Robust standard errors are in parentheses in columns 3 and 4. The randomization-*t* p-values in brackets. Inference in each column is based on a randomization inference procedure of Young (2019). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The row of *Multiple-Hypothesis Testing* reports the randomization-*t* p-values for multiple-hypothesis testing computed based on a randomization inference procedure of Young (2019), which applies the procedure of Westfall and Young (1993). It tests the null hypothesis that all treatment effects in each equation (each column) are zero.

Table A6: Results of the randomized survey experiment with interactions with time (Landry et al., 2006)

	(1)	(2)	(3)	(4)
	Ordered logit		OLS	
T1: Before–after study without control	0.015	0.017	0.006	0.007
	(0.153)	(0.153)	(0.072)	(0.072)
	[0.909]	[0.901]	[0.939]	[0.921]
T2: Treatment is a message rather than a raffle	-0.143	-0.145	-0.067	-0.069
	(0.152)	(0.152)	(0.069)	(0.069)
	[0.266]	[0.259]	[0.338]	[0.320]
T3: Promoting waste sorting rather than donations	-0.317**	-0.325**	-0.158**	-0.164**
	(0.154)	(0.154)	(0.072)	(0.071)
	[0.014]	[0.011]	[0.027]	[0.021]
Long time	0.762***	0.751***	0.346***	0.340***
	(0.157)	(0.157)	(0.081)	(0.081)
T1 × Long time	-0.136	-0.128	-0.101	-0.097
	(0.224)	(0.225)	(0.121)	(0.121)
	[0.547]	[0.570]	[0.402]	[0.421]
T2 × Long time	-0.029	-0.015	-0.015	-0.006
	(0.222)	(0.222)	(0.115)	(0.115)
	[0.896]	[0.947]	[0.895]	[0.954]
T3 × Long time	-0.149	-0.138	-0.082	-0.076
	(0.223)	(0.223)	(0.117)	(0.117)
	[0.506]	[0.534]	[0.474]	[0.508]
Order (1/2)		-0.125		-0.085**
		(0.079)		(0.042)
Constant			3.271***	3.403***
			(0.050)	(0.081)
Multiple-Hypothesis Testing	0.072	0.057	0.126	0.100
Observations	2146	2146	2146	2146
Pseudo-R^2/ R^2	0.015	0.015	0.029	0.031

Notes: This table reports the estimates from regression analyses where the dummy variable of time spent on the survey being longer than the median is incorporated into the analysis of Table 5. The coefficients are reported. Standard errors are in parentheses in columns 1 and 2. Robust standard errors are in parentheses in columns 3 and 4. The randomization-t p-values in brackets. Inference in each column is based on a randomization inference procedure of Young (2019). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively. The row of *Multiple-Hypothesis Testing* reports the randomization-t p-values for multiple-hypothesis testing computed based on a randomization inference procedure of Young (2019), which applies the procedure of Westfall and Young (1993). It tests the null hypothesis that all treatment effects in each equation (each column) are zero.

Appendix G. Additional Analysis on the Effect of an Implementer

In the literature, it is considered that practical ethical issues of field experiments are likely associated with a question of whether the researchers who are designing programs should be regulated as researchers or as implementers (for more details, see Glennerster and Powers, 2016).[3] While the previous studies discuss this issue from the normative perspective, I empirically examine respondents' recognition of ethical issues by the type of implementers in this section in Online Appendix.

Do respondents recognize fewer ethical issues if the experiment is run by an implementer other than researchers? To examine this question, in the three studies examined in the first survey, I randomly made small changes in the descriptions. For the respondents who are assigned Allcott (2011), Hanna et al. (2016), or Hosono and Aoyagi (2018), a randomly assigned half of them are shown a description that mentions the implementer of the program being someone other than "Professor X."

In the experiments studied in Allcott (2011), a company called OPOWER was the implementer of the program. Thus, a randomly assigned half of the respondents is shown descriptions similar to those in Appendix B4, but "Professor X" is replaced with "a company." In the experiments studied in Hanna et al. (2016), the program was not implemented by the authors but by an NGO. Thus, a randomly assigned half of the respondents is shown descriptions similar to those in Appendix A4, but "Professor X" is replaced with "a nonprofit organization." In the experiments studied in Hosono and Aoyagi (2018), the program was implemented by the JICA, as was already mentioned in Section 3. Thus, "Professor X" is replaced with "an international development agency."

This design of randomized survey experiments allows me to estimate a causal impact of changing the implementer of the experiment to a nonresearcher. I conduct regression analyses using the subsample of the dataset used in Section 3.

Table A7 presents the summary statistics for the respondents in the randomized survey experiments to verify the randomizations. From this table, I conclude that two groups in each survey experiment are very similar.

---

[3] In contrast to the present study, Barnett and Camfield (2016) discuss specific ethical issues that arise in the randomized evaluation of programs by nonresearchers.

Table A7: Summary statistics of the randomized survey experiments by group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Allcott (2011)			Hanna et al. (2016)			Hosono and Aoyagi (2018)		
	Prof. X	Company	*P*-value (1) vs (2)	Prof. X	NPO	*P*-value (4) vs (5)	Prof. X	IDA	*P*-value (7) vs (8)
Female	0.482	0.459	0.309	0.480	0.483	0.899	0.482	0.466	0.486
	(0.500)	(0.499)		(0.500)	(0.500)		(0.500)	(0.499)	
Age	46.607	47.411	0.209	46.701	46.367	0.601	46.633	47.120	0.447
	(14.065)	(14.024)		(14.062)	(14.068)		(14.034)	(14.375)	
Married	0.610	0.598	0.596	0.608	0.614	0.817	0.611	0.589	0.337
	(0.488)	(0.491)		(0.488)	(0.487)		(0.488)	(0.492)	
Living with children	0.379	0.381	0.932	0.378	0.398	0.360	0.378	0.388	0.670
	(0.485)	(0.486)		(0.485)	(0.490)		(0.485)	(0.488)	
Household income	534.991	538.564	0.781	534.529	543.525	0.480	534.022	549.268	0.235
(10 thousand yen)	(249.133)	(249.175)		(249.475)	(245.305)		(248.779)	(252.662)	
Full-time employee	0.251	0.229	0.265	0.247	0.269	0.261	0.252	0.215	0.061
	(0.433)	(0.420)		(0.431)	(0.444)		(0.434)	(0.411)	
Part-time employee	0.125	0.107	0.211	0.124	0.125	0.935	0.125	0.108	0.260
	(0.331)	(0.309)		(0.329)	(0.331)		(0.331)	(0.311)	
Temporary/contract employee	0.051	0.059	0.429	0.052	0.049	0.787	0.052	0.053	0.871
	(0.220)	(0.236)		(0.222)	(0.217)		(0.221)	(0.225)	
Self-employed	0.055	0.063	0.444	0.056	0.047	0.391	0.055	0.059	0.722
	(0.228)	(0.243)		(0.230)	(0.213)		(0.228)	(0.236)	
Living in Tokyo	0.126	0.112	0.368	0.125	0.117	0.591	0.125	0.118	0.614
	(0.332)	(0.316)		(0.331)	(0.322)		(0.331)	(0.323)	
Living in Osaka	0.072	0.069	0.774	0.071	0.081	0.363	0.071	0.080	0.447
	(0.258)	(0.253)		(0.256)	(0.274)		(0.257)	(0.271)	

Notes: Standard deviations are in parentheses. Columns 3, 6 and 9 report *p*-values for the differences.

Table A8: Results of the randomized survey experiment on the effect of changing the implementer of the experiment

	Allcott (2011)		Hanna et al. (2016)		Hosono and Aoyagi (2018)	
	(1) Ologit	(2) OLS	(3) Ologit	(4) OLS	(5) Ologit	(6) OLS
Company	0.097 (0.113) [0.399]	0.044 (0.061) [0.485]				
Nonprofit organization			-0.226** (0.113) [0.046]	-0.112* (0.060) [0.064]		
International development agency					-0.043 (0.114) [0.711]	-0.010 (0.057) [0.862]
Order (1-6)	-0.152*** (0.033)	-0.083*** (0.017)	-0.080** (0.033)	-0.039** (0.017)	-0.079** (0.034)	-0.038** (0.017)
Constant		3.238*** (0.076)		3.257*** (0.074)		3.212*** (0.072)
Number of Observations	1051	1051	1053	1053	1054	1054
Pseudo R-squared / R-squared	0.008	0.021	0.003	0.008	0.002	0.005

Notes: This table reports the estimates from regression analyses that use a subsample of the first survey. The dependent variable is the response to the question "Do you recognize any ethical issues in this study?" on a five-point scale (1–5). Columns 1 and 2 report the results for the survey on two types of descriptions of Allcott (2011). Randomly assigned respondents are shown a description similar to that in Appendix B4, but the implementer of the project is a "company" instead of Professor X. Columns 3 and 4 report the results on Hanna et al. (2016), where the randomly assigned half of the respondents are shown "NPO" instead of Professor X. Columns 5 and 6 report the results on Hosono and Aoyagi (2018), where the randomly assigned half of the respondents are shown "international development agency" instead of Professor X. The coefficients are reported. Standard errors are in parentheses in columns 1, 3 and 5. Robust standard errors are in parentheses in columns 2, 4, and 6. The randomization-t p-values are in brackets. Inference in each column is based on a randomization inference procedure of Young (2019). ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table A8 reports the estimation results. Changing the implementer of the program from a researcher to a company does not change the concerns (columns 1 and 2). Changing the implementer of the program to an international development agency does not change the concerns (columns 5 and 6). However, changing the implementer of the program to a nonprofit organization (NPO) significantly decreases the concerns (columns 3 and 4). Since the coefficient (and the standard error) of the OLS estimation result is -0.11 (0.06), the magnitude of this *implementer effect* is approximately half of the effect of changing an outcome variable from charitable giving to waste sorting (see Table 8). Note that the treatment is significant even after adjusting for multiple-hypothesis testing within ordered logit and OLS regression (the randomization-t p-value of 0.064). The result suggests that although the magnitude of the effect is not large, the implementer of the program can affect respondents' recognition of ethical issues.

Several interpretations are possible. Respondents may consider that the objective of the program and its random assignment is different for researchers and NPOs. People may consider that unlike NPOs, researchers may randomize an intervention just to extend knowledge but not to improve social welfare. If they feel this way, then they may rate the experiment conducted by researchers lower than that conducted by NPOs. Another interpretation can be that respondents

may trust NPOs more than social scientists. If the implementer of the program reliably explains how the findings obtained from the experiment can contribute to society, subjects and related individuals may be more likely to accept the experiment. In addition, the result implies that ethical concerns are lesser when a researcher evaluates a program implemented by NPOs rather than one implemented by herself/himself.

References

Allcott, Hunt (2011) "Social norms and energy conservation," *Journal of Public Economics*, Vol. 95, No. 9-10, pp. 1082–1095.

Barnett, Chris and Laura Camfield (2016) "Ethics in evaluation," *Journal of Development Effectiveness*, Vol. 8, No. 4, pp. 528–534.

Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden (2020) "The Effect of Early-Childhood Education on Social Preferences," *Journal of Political Economy*, Vol. Forthcoming, Forthcoming.

Fryer, Roland G, Steven D Levitt, and John A List (2015) "Parental incentives and early childhood achievement: A field experiment in Chicago heights," NBER Working Paper Series 21477, National Bureau of Economic Research, Available at `http://www.nber.org/papers/w21477`.

Glennerster, Rachel and Shawn Powers (2016) "Balancing risk and benefit: Ethical tradeoffs in running randomized evaluations," in DeMartino, George and Deirdre McCloskey eds. *Oxford Handbook of Professional Economic Ethics*, Oxford: Oxford University Press, Chap. 20, pp. 367–401.

Gneezy, Uri and John A. List (2013) *The Why Axis: Hidden Motives and the Undiscovered Economics of Everyday Life*: PublicAffairs.

Hanna, Rema, Esther Duflo, and Michael Greenstone (2016) "Up in smoke: The influence of household behavior on the long-run impact of improved cooking stoves," *American Economic Journal: Economic Policy*, Vol. 8, No. 1, pp. 80–114.

Hosono, Tomoyuki and Keitaro Aoyagi (2018) "Effectiveness of interventions to induce waste segregation by households: Evidence from a randomized controlled trial in Mozambique," *Journal of Material Cycles and Waste Management*, Vol. 20, No. 2, pp. 1143–1153.

Jyukankyo Research Institute Inc. (2016) "Improvement of infrastructure to promote rationalization of energy use in fiscal year 2015 (Survey on the effect of providing information on energy use on the promotion of changes in household energy conservation behavior)," Technical report, Minister of Economy, Trade and Industry (METI), Japan.

Landry, Craig E, Andreas Lange, John A List, Michael K Price, and Nicholas G Rupp (2006) "Toward an understanding of the economics of charity: Evidence from a field experiment," *Quarterly Journal of Economics*, Vol. 121, No. 2, pp. 747–782.

Thornton, Rebecca L (2008) "The demand for, and impact of, learning HIV status," *American Economic Review*, Vol. 98, No. 5, pp. 1829–63.

Young, Alwyn (2019) "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results," *Quarterly Journal of Economics*, Vol. 134, No. 2, pp. 557–598.