

The SOM Institute's Notes on Survey Methodology – 2023:2

The impact sampling strategies have on experimental treatment effects

Alexandra Garcia Nilsson

The SOM Institute, University of Gothenburg
alexandra.garcia.nilsson@gu.se

Sophie Cassel

The SOM Institute, University of Gothenburg
sophie.cassel@gu.se

ABSTRACT

This study tested the generalizability of the Swedish Citizen Panel (SCP) by administering three well-known psychological experiments to both probability-based and self-selected samples of the panel. Results from the experiments were compared both in samples prestratified on sex, age, and education and samples not prestratified. Cohen's d was used to compare effect sizes between the SCP samples, the original findings, and other replications of the psychological experiments. Across all SCP samples (i.e., the probability and self-selected samples), none of the estimated effect sizes differed. Furthermore, the estimated effect sizes did not differ between prestratifying the samples or not. In line with other similar replication studies, the effects estimated in the SCP were smaller than the original published result. These results indicate that self-selected samples produce similar results as probability-recruited and can validly be used for psychological experiments.

INTRODUCTION

A self-selected sample is cheaper and easier to recruit and maintain than a probability-recruited one. However, the generalizability of results from studies using self-selected samples could be questioned due to respondents not being randomly selected. In psychological research on cognitive processes, sample quality is thought to be of less importance and self-selected and probability-based samples are thought to generate equivalent treatment effects.

The aim of the present study was to test whether participants recruited through self-selection generated equivalent treatment effects as participants recruited through probability-based sampling. To assess this, three cognitive bias experiments were administered to four different samples of the Swedish Citizen Panel (SCP). Two samples of 2,500 randomly selected participants each were drawn in a prestratified manner, one within the probability-recruited and one within the self-selected portion of the SCP. The other two samples of 2,500 participants each were drawn using random selection within the probability recruited and the self-selected portion of the SCP but without using prestratification. The estimated treatment effects among the four SCP samples were compared to the original effect sizes of the experiments as well as to replications within ManyLabs (ML) studies (Richard et al., 2014; Klein et al., 2018) presented below.

THE EXPERIMENTS

Three experiments were chosen from the replications in ML1 (Richard et al., 2014) and ML2 (Klein et al., 2018). The experiments were selected by the criteria of being successfully replicated in ML1 or ML2 and being easy to administer. All three of the experiments compared decision-making strategies.

The participants were, independently for each experiment, randomized into one of two experimental groups. The participants imagined themselves being in a situation and were presented with different information regarding the situation depending on the assigned experiment group. Thereafter they answered questions related to the presented situation.

The experiments were:

- Experiment 1 - *Why people are reluctant to tempt fate* (Risen & Gilovich, 2008), replicated in ML2 (Klein et al., 2018)
- Experiment 2 - *The framing of decisions and the psychology of choice* (Tversky & Kahneman, 1981), replicated in ML2 (Klein et al., 2018)
- Experiment 3 - *Gain versus loss framing* (Tversky & Kahneman, 1981), replicated in ML1 (Richard et al., 2014)

EXPERIMENT 1 - WHY PEOPLE ARE RELUCTANT TO TEMPT FATE

In the first experiment, participants imagined themselves being in a large lecture hall where the professor asked a question that no one volunteered to answer. Since no one volunteered, the professor told the class that a student would be chosen at random to answer the question.

- Experiment group 1: the participants were told they had done the reading for the lecture in the scenario and was sure about the answer to the question.
- Experiment group 2: the participants were told they had not done the reading for the lecture in the scenario and did not know the answer to the question.

Then, participants answered a question on the probability of being chosen by the lecturer to answer the question (see Appendix 1 for the wording of the scenarios and question).

The hypothesis was that if someone thought they had tempted fate they would estimate the probability of a bad thing happening as higher than if they had not tempted fate. In this case, the participants imagining themselves not having read the material were expected to estimate the probability of being asked the question as higher than the participants imagining themselves having read the material.

In the original study, the participants who imagined themselves not having read the material estimated the risk of being selected to answer the question as higher than those imagining having read the material ($d=0.39$, 95% CI [0.03, 0.75]) (Risen & Gilovich,

2008). The replication in ML2 had effects in the same direction as the original, even though the effect sizes were smaller. The original study was carried out on students, so the ML2 replication also made separate analyses for students ($d=0.22$, 95% CI[0.17, 0.28]) and non-students ($d=0.18$, 95% CI[0.14, 0.22]) (Klein et al., 2018). Furthermore, other replication studies have revealed results of failure to replicate the tempt fate-effect, and once the effect sizes even were in the opposite direction than the hypothesized (Mathur et al., 2020; Frank & Mathur, 2022).

EXPERIMENT 2 - THE FRAMING OF DECISIONS AND THE PSYCHOLOGY OF CHOICE

In the second experiment, participants imagined themselves being in a store buying two items. One item was significantly more expensive than the other. The cashier informed that one of the items was on sale in another store, a driving distance away.

- Experiment group 1: the cheaper item was on sale.
- Experiment group 2: the more expensive item was on sale.

A crucial part of the experiment was that the sale was in absolute value (in crowns, dollars, or other local currency) and the amount of the sale was the same for the two experimental groups. The proportional sale as a percentage of the original price of the item on sale would differ depending on whether the cheaper or more expensive item was on sale (see Appendix 2 for the wording of the scenarios and question).

The hypothesis was that if the sale as a proportion of the original price was higher, the customer would be more willing to make the effort of going to the other store, even though the amount of the sale was the same in the two situations.

In the ML2 replication, the effect was in the same direction as in the original, but the effect sizes were substantially lower. Original: $d=1.08$, 95% CI[0.71, 1.45] (Tversky & Kahneman, 1981) and in ML2: $d=0.40$, 95% CI[0.35, 0.45] (Klein et al., 2018). This experiment has been replicated many times and was, together with the third experiment, part of the research which was rewarded with a Prize in Economic Sciences in Memory of Alfred Nobel.

EXPERIMENT 3 - GAIN VERSUS LOSS FRAMING

In the third experiment, participants imagined a country in preparation for an epidemic disease which expected to kill 600 people. There were two alternative programs the country could enroll in, one with a certain outcome and one with probabilistic outcome. The expected amount of people dead/alive was the same in the two programs. The experiment groups were presented with the same two programs, but each framed differently.

- Experiment group 1: Policies were presented in terms of people saved.
- Experiment group 2: Policies were presented in terms of people dying.

The hypothesis was that people would be more willing to take a risk if the situation was framed in negative terms (dying), that is, to be more willing to choose the probabilistic program when presented with the information in negative terms (see Appendix 3 for the wording of the scenarios and questions).

The directions of the effects were in line with the hypothesis, both in the original study and in the ML1 replication. The effect sizes were again much lower in the replication ($d=0.62$, 95% CI[0.52, 0.71]) (Richard et al., 2014) than in the original ($d=1.13$, 95% CI[0.89, 1.37]) (Tversky & Kahneman, 1981).

This experiment has been replicated successfully many times. For example, in a study which in similar way as in this present study used experiments to evaluate Amazon's Mechanical Turk panel (Berinsky et al., 2012).

METHODS AND MATERIALS

SAMPLE

The experiments was administered to four samples of the SCP. Two samples were prestratified by sex (male, female), age (18–34, 35–49, 50–85 years), and education (low/middle education: less than 3 years of post-secondary education, high education: 3 or more years of post-secondary education), one with self-selected participants and the other of probability-recruited participants. Each stratified sample consisted of 2,500 participants,

1,382 participants in the stratified self-selected, and 1,326 participants in the probability-recruited sample completed the questionnaire where the experiments were embedded.

In addition, two samples without stratification were invited to complete the experiment, these samples were drawn using random selection within the probability recruited and the self-selected portion of the SCP. In these two samples, 2,500 participants in each were invited and 1,590 in the self-selected sample respective 1,543 in the probability-recruited sample completed it. For details on the demographic composition of responding participants see Appendix 4.

PROCEDURE

Information and the questions of the experiments were translated into Swedish. The order of the experiments and the experimental groups were randomized for each participant. In Experiment 2 – the framing of decisions and the psychology of choice – the consumer items were the same as in the ML2 replication and the prices were exchanged to Swedish crowns.

Effect sizes for a self-selected sample were compared with effect sizes for a probability-recruited sample, for both stratified subsamples and not stratified samples. The effect sizes were also related to the original version of the experiments and the replications from ML1 or ML2.

Cohen's d , a standardized measurement to compare effect sizes, was used to measure the effect sizes and 95 percent confidence intervals were used to compare the effect sizes.

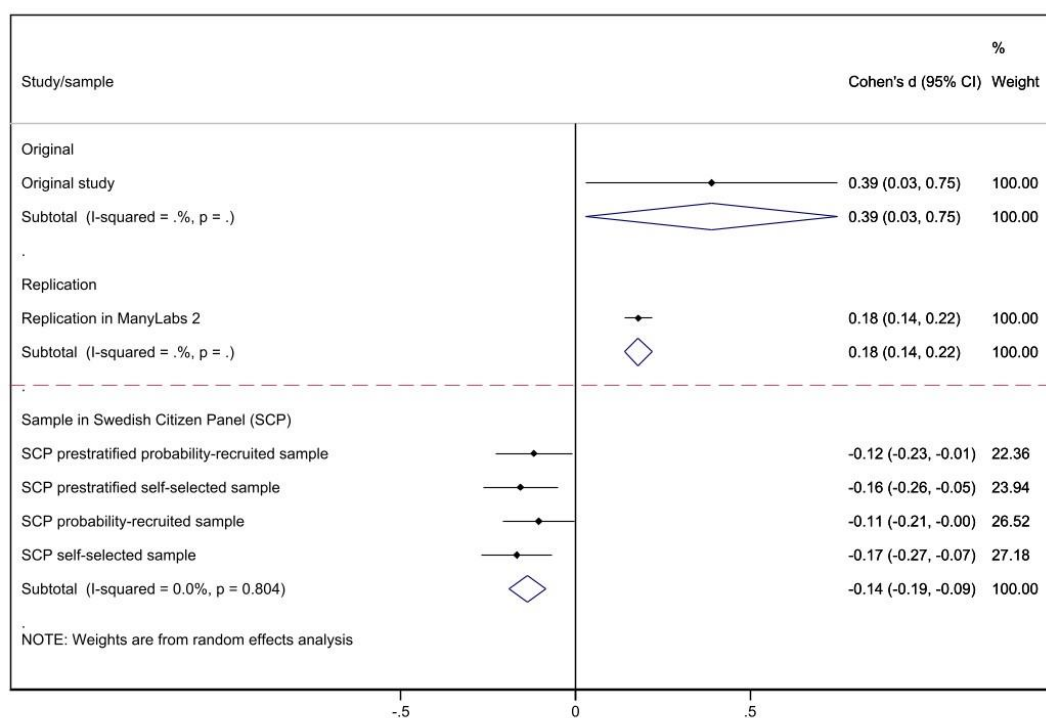
RESULTS

EXPERIMENT 1 - WHY PEOPLE ARE RELUCTANT TO TEMPT FATE

The effect sizes did not differ between the four samples of the SCP (prestratified probability-recruited sample: $d=-0.12$, 95% CI[-0.23, -0.01]; prestratified self-selected sample: $d=-0.16$, 95% CI [-0.26, -0.05]; probability-recruited sample: $d=-0.11$, 95% CI[-0.21, -0.00]; self-selected sample: $d=-0.17$, 95% CI[-0.27, -0.07]).

Unexpectedly, the effects sizes were significantly negative across all four SCP samples, contradicting the original and replicated results. That is, participants were less likely to believe they were going to be chosen when they had not prepared for the lecture. The average effects sizes for samples of the SCP was $d=-0.14$, 95% CI $[-0.19, -0.09]$, compared to $d=0.39$, 95% CI $[0.03, 0.75]$ in the original study and $d=0.18$, 95% CI $[0.14, 0.22]$ for the ML2 replication.

Figure 1. Effect of not having read the literature on the perceived probability of being asked the question.



Note. Standardized effect sizes by Cohen's d . Bars indicate a 95% confidence interval. The published results appear above the red line and the SCP results appear below the red line.

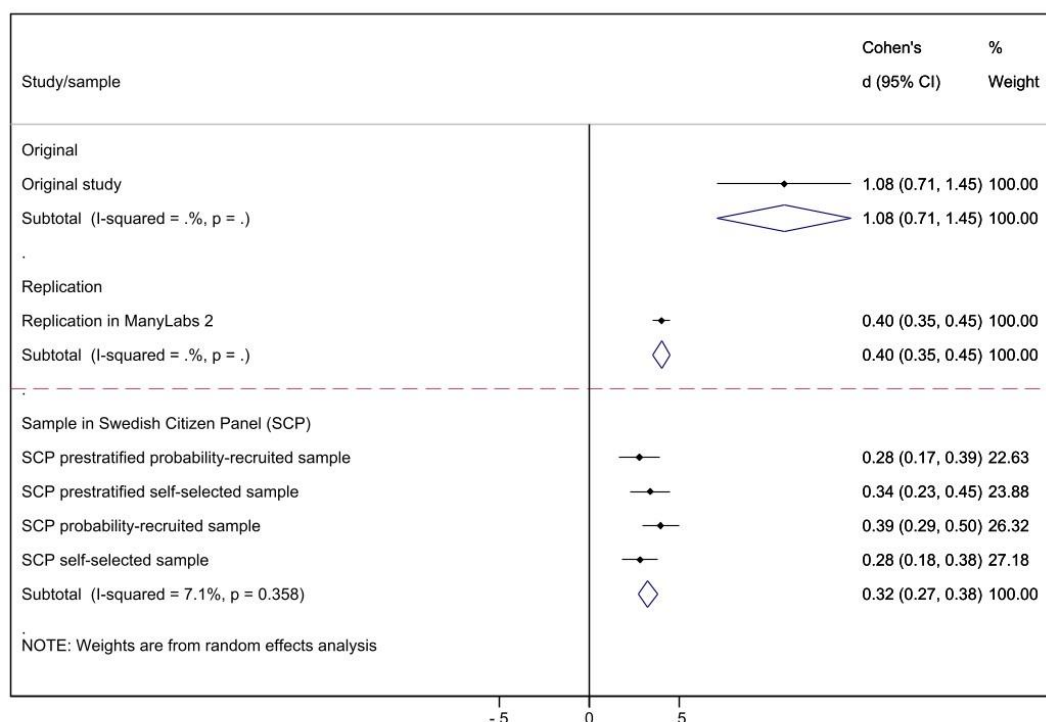
EXPERIMENT 2 - THE FRAMING OF DECISIONS AND THE PSYCHOLOGY OF CHOICE

Across all four samples of the SCP, participants who were told the cheaper item was on sale were more likely to travel to the other store. Furthermore, the effect sizes did not differ between the four SCP samples (prestratified probability-recruited sample: $d=0.28$, 95% CI $[0.17, 0.39]$; prestratified self-selected sample: $d=0.34$, 95% CI $[0.23, 0.45]$; probability-

recruited sample: $d=0.39$, 95% CI[0.29, 0.50]; self-selected sample: $d=0.28$, 95% CI[0.18, 0.38]).

The average effect sizes of the SCP samples ($d=0.32$, 95% CI[0.27, 0.38]) were in the same direction as the published results but smaller. For comparison, effect sizes in the original study was $d=1.08$, 95% CI[0.71, 1.45] respectively $d=0.4$, 95% CI[0.35, 0.45] for the ML 2 replication.

Figure 2. Effect of the base cost of the discounted item on willingness to travel.



Note. Effect sizes measured by Cohen's d . Bars indicate a 95% confidence interval. The published results appear above the red line and the SCP results appear below the red line.

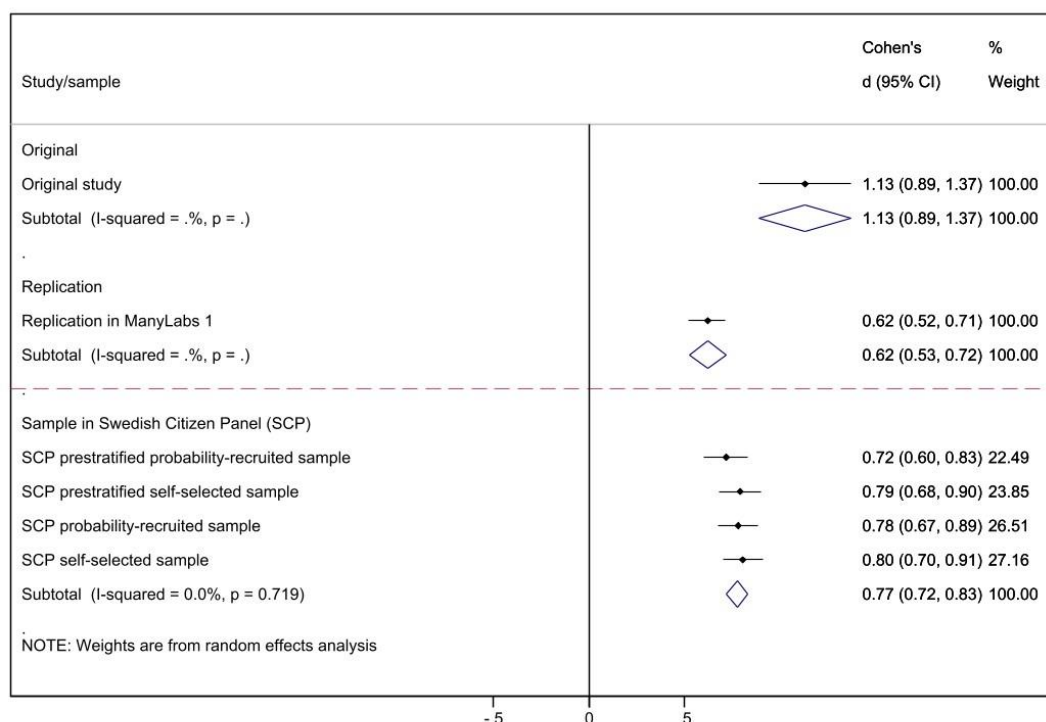
EXPERIMENT 3 - GAIN VERSUS LOSS FRAMING

As expected, across all four SCP samples, participants were more likely to favour the risky (probabilistic) policy when the policies were framed in terms of losses (people dying) compared to when they were framed as gains (people saved). Furthermore, the effect sizes did not differ between the four different SCP samples (prestratified probability-recruited

sample: $d=0.72$, 95% CI[0.60, 0.83]; prestratified self-selected sample: $d=0.79$, 95% CI[0.68, 0.90]; probability-recruited sample: $d=0.78$, 95% CI[0.67, 0.89]; self-selected sample: $d=0.80$, 95% CI[0.70, 0.91]).

The average effect size of the SCP samples ($d=0.77$, 95% CI[0.72, 0.83]), as well as the effect sizes for each SCP sub-sample, were all between the effect sizes found in the original study ($d=1.13$, 95% CI[0.89, 1.37]) and the ML1 replication ($d=0.62$, 95% CI[0.52, 0.71]).

Figure 3. The effect of framing a situation in positive terms on the willingness to take a risk.



Note. Effect sizes measured by Cohen's d . Bars indicate a 95% confidence interval. The published results appear above the red line and the SCP results appear below the red line.

CONCLUSION

The results of the present study produced no significant difference in effect sizes between different samples from the SCP. Such a finding is comforting for those who administer psychological experiments involving cognitive processes among self-selected samples and among self-selected samples in the SCP in particular. Whether a sample is prestratified or

not did not matter for these types of experiments, at least in terms of generating similar effect sizes.

Furthermore, the effect sizes from the SCP samples were in line with the results of both the original experiment and the compared ML replication for the second and third experiments. In general, effect sizes from the SCP samples were more similar to the effect sizes generated in the replication studies than they were compared to the original studies.

However, in the first experiment, the effects acquired within the SCP samples deviated from the ones in the original and ML2 replication studies. But, other replications of this experiment (Mathur et al., 2020; Frank & Mathur, 2022) have either failed to produce any effects or found similar results as the SCP samples. Future research should continue to investigate the tempt fate-hypothesis across different contexts and culture to assess during which boundaries the effect may exist.

Finally, we urge future studies to assess the replicability and effect sizes of non-psychological experiments, such as political science and sociological experiments, in order to investigate if self-selected samples differ from probability-based samples in terms of estimated effect sizes.

REFERENCES

- Berinsky, A., Huber, G., & Lenz, G. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*, 20(3), 351-368. <https://doi:10.1093/pan/mpr057>
- Frank, M. C., & Mathur, M. B. (2022, June 6). Replication of Risen & Gilovich (2008, JPSP). Retrieved from <https://osf.io/sg3su/>
- Klein, R.A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., ... , Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*. 1(4), 443-490. <https://doi.org/10.1177/2515245918810225>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., ... , Nosek, B. A., (2014). Many Labs 1: Investigating Variation in Replicability. *Social Psychology*, 45(3), 142-152. <https://doi.org/10.1027/1864-9335/a000178>
- Risen, J.L. & Gilovich, T. (2008) Why People Are Reluctant to Tempt Fate. *Journal of Personality and Social Psychology*, 95(2), 293-307. <https://doi.org/10.1037/0022-3514.95.2.293>
- Mathur, M. B., Bart-Plange, D. J., Aczel, B., Bernstein, M. H., Ciunci, A. M., Ebersole, C. R., Falcão, F., Ashbaugh, K., Hilliard R. A., Jern A., Kellier, J., Kessinger, G., Kolb, V.S., Kovacs, M., Ambrosio Lage, C., Langford, E.V., Lins, S., Manfredi, D., Meyet, V., ..., Frank, M. C. (2020). Many Labs 5: Registered multisite replication of the tempting-fate effects in Risen and Gilovich (2008). *Advances in Methods and Practices in Psychological Science*, 3(3), 394-404. <https://doi.org/10.1177/2515245918785165>

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453 – 458. <https://doi.org/10.1126/science.7455683>

APPENDIX 1. INFORMATION AND QUESTION FOR EXPERIMENT 1

Text for experiment group 1:

Läs följande scenario noggrant.

Föreställ dig att du sitter i en större föreläsningssal med några hundra andra studenter. Du sitter ungefär i mittsektionen av salen, lite mer än halvvägs bak. Föreställ dig att föreläsaren ställer en fråga om kurslitteraturen som ligger till grund till föreläsningen, men ingen av studenterna räcker upp handen för att svara. Du har läst kursmaterialet och känner dig säker på att föreläsaren skulle gilla ditt svar, men du föredrar att inte lämna svar frivilligt i stora klasser. Klassen sitter i tystnad i två minuter innan föreläsaren förklarar att om ingen ställer upp frivilligt så väljer han någon istället.

Text for experiment group 2:

Läs följande scenario noggrant.

Föreställ dig att du sitter i en större föreläsningssal med några hundra andra studenter. Du sitter ungefär i mittsektionen av salen, lite mer än halvvägs bak. Föreställ dig att föreläsaren ställer en fråga om kurslitteraturen som ligger till grund till föreläsningen, men ingen av studenterna räcker upp handen för att svara. Du har inte läst kursmaterialet och känner dig säker på att du inte skulle kunna svara på frågan. Klassen sitter i tystnad i två minuter innan föreläsaren förklarar att om ingen ställer upp frivilligt så väljer han någon istället.

Question, answered by both experiment groups:

Hur troligt tror du att det är att föreläsaren skulle välja ut dig för att svara på frågan?
Markera ditt svar genom att välja lämpligt nummer nedan.

0										10
Inte alls										Extremt
troligt	1	2	3	4	5	6	7	8	9	troligt
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

APPENDIX 2. INFORMATION AND QUESTION FOR EXPERIMENT 2

Text and question for experiment group 1:

Läs följande scenario noggrant.

Föreställ dig att du är i en butik där du är på väg att köpa en porslinsvas för 2680 kronor och en prydnad att ha på väggen för 320 kronor. Säljaren i butiken informerar dig om att prydnaden du vill köpa finns på rea för 220 kronor i företagets andra butik som ligger 20 minuters bilresa bort.

Skulle du åka till den andra butiken?

- ☐ Ja, jag skulle åka till den andra butiken
- ☐ Nej, jag skulle inte åka till den andra butiken

Text and question for experiment group 2:

Läs följande scenario noggrant.

Föreställ dig att du är i en butik där du är på väg att köpa en porslinsvas för 320 kronor och en prydnad att ha på väggen för 2680 kronor. Säljaren i butiken informerar dig om att prydnaden du vill köpa finns på rea för 2580 kronor i företagets andra butik som ligger 20 minuters bilresa bort.

Skulle du åka till den andra butiken?

- ☐ Ja, jag skulle åka till den andra butiken
- ☐ Nej, jag skulle inte åka till den andra butiken

APPENDIX 3. INFORMATION AND QUESTION FOR EXPERIMENT 3

Text for experiment group 1:

Läs följande scenario noggrant.

Föreställ dig att ditt land förbereder sig för ett utbrott av en ovanlig sjukdom, som förväntas döda 600 människor. Två olika program för att bekämpa sjukdomen har föreslagits. Anta att de exakta vetenskapliga uppskattningarna av konsekvenserna av programmen är följande:

Om Program A genomförs kommer 200 personer att räddas.

Om Program B genomförs finns det 1/3 sannolikhet att 600 personer kommer att räddas och 2/3 sannolikhet att inga människor kommer att räddas.

Text for experiment group 2:

Läs följande scenario noggrant.

Föreställ dig att ditt land förbereder sig för ett utbrott av en ovanlig sjukdom, som förväntas döda 600 människor. Två olika program för att bekämpa sjukdomen har föreslagits. Anta att de exakta vetenskapliga uppskattningarna av konsekvenserna av programmen är följande:

Om Program A genomförs kommer 400 personer att dö.

Om Program B genomförs finns det 1/3 sannolikhet att ingen kommer att dö och 2/3 sannolikhet att 600 människor kommer att dö.

Question, answered by both experiment groups:

Vilket program skulle du välja?

☐ Program A

☐ Program B

APPENDIX 4. DESCRIPTIVE STATISTICS

Table 1. Demographic composition of responding participants.

	Self-recruited sample	Probability-recruited sample	Stratified self-recruited sample	Stratified probability-recruited sample
<i>Gender</i>				
Woman	38%	48%	52%	49%
Man	62%	52%	48%	51%
<i>Age</i>				
18-34	6%	10%	17%	17%
35-49	27%	20%	24%	22%
50-85	66%	70%	59%	61%
Not available	1%	-	-	-
<i>Education</i>				
Low/middle	38%	34%	73%	72%
High	62%	66%	27%	28%
<i>n</i>	1,590	1,543	1,382	1,326

Note. Proportion of respondents by gender, age and education in the different samples of the present study, presented as percentages.



The SOM Institute is an academic organization located at the University of Gothenburg, Sweden. Since 1986 the SOM Institute conduct annual cross-sectional surveys among the Swedish population with a focus on Society, Opinion, and Media, as well as administering the web panel called the Swedish Citizen Panel. The annual surveys and the web panel both function as infrastructures, enabling researchers and public organizations to effectively collect research and opinion data in collaboration with researchers at the SOM Institute.

In order to strengthen contemporary research on Swedish society, as well as to contribute to international methodological development, the SOM Institute frequently publish notes on methodological research. These are made available for scholars and the public alike at <https://www.gu.se/en/som-institute/publications/research-on-survey-methodology>

For more information, please contact us at: info@som.gu.se