Psychological Mechanisms for Eliciting Preferences and Beliefs

Evan Friedman   Suanna Oh   Duncan Webb

PSE   PSE   Princeton

KAEA-ASSA Academic Session III

January 2025

Challenges with eliciting preferences and beliefs

- A central challenge in social science research is being able to accurately elicit subjects' choices and beliefs.

  - Without incentives to give thoughtful answers, subjects may put little effort into answering questions

  - Proportion of inattentive respondents can be high (1-30%) (Curran 2016)

- Inattentive respondents yield data of a markedly poorer quality.

  - Measurement error could obscure meaningful correlations or experimental effects (Meade & Craig 2012; Maniaci & Rogge 2014)

Incentivizing survey responses

- Some types of questions are more straightforward to incentivize:
    - Preferences over objects that can be "paid out"
    - Beliefs about observable states
    - Choices with objectively correct answers

- We refer to such questions as being *verifiable*.
    - However, unclear whether paying out these questions always helps (Danz et al. 2022, Enke et al. 2023).

- What of *unverifiable* questions?
    - Preferences over hypothetical, abstract, or prohibitively expensive objects
    - Beliefs over hypothetical or far-future events
    - Choices with subjectively correct answers

Challenge of improving survey response quality

- **Peer prediction methods**
  (Prelec 2004; Cvitanic et al. 2019; Hussam, Rigol, and Roth 2020)
  - ▶ Tend to assume highly sophisticated Bayesian respondents
  - ▶ Often difficult for respondents to understand; relies on "intimidation method"

- **Attention checks**
  - ▶ Cost of losing data, and selective sample (Berinsky et al 2014)

- **Ex-post corrections**
  - ▶ e.g., duplicate questions, self-knowledge estimator (Falk, Neuber, and Strack 2021)
  - ▶ Potentially large increases in survey time, further depleting attention
  - ▶ Few improvements on the "intensive margin" of attention (Jacquemet et al. 2024)

- **Increasing total payments**
  - ▶ Increasing show-up fees improves quantity but not quality of work (Mason & Watts 2009)

This paper

We evaluate two novel methods for incentivizing survey answers that can be applied to both *verifiable* and *unverifiable* questions:

1. **"Restatement Method" (RM)**
   - Relies on the idea that thoughtful answers are easier to reproduce at a later time

2. **"Bonus" incentive (B)**
   - Relies on gift exchange effects from a surprise bonus

- Both methods are easy to comprehend, involve only small increases in survey time, and can be applied to arbitrary question types.

- We test the efficacy of these methods against **benchmarks**:
   - No incentive (NI: control), Conventional incentive (CI)

Incentives

Types of incentives

<1> Restatement (RM)

- "At the end of the experiment, we will randomly choose three questions from this section and ask you to tell us the same answers you gave previously."
- "By correctly selecting the same answers you gave previously, you will earn more money."

<2> Restatement - points only (RMP)

- Similar to the above, but subjects are awarded 1 point per question restated, which has no monetary value
- Allows us to see if RM can work "for free" due to intrinsic desire to perform well

Why the Restatement Method could work

- Consider any question that concerns:
  - stable underlying preferences or beliefs
  - long-term memories
  - reproducible cognitive processes

- If you put effort into answering well initially, you may easily reconstruct your answer *without* memorizing it

- However, may not work well for certain types of questions
  - e.g., questions with focal answers, social desirability bias, or difficult solutions
  - We avoid this type of questions in this study

Potential reactions to the method

- One obvious concern is that RM may lead to "gaming" behavior, e.g., chooing salient answers that are easy to restate or recording answers

  - However, this may conflict with their intrinsic desire to answer surveys
  - RMP may be less subject to the concern
  - Important to empirically test the methods

- As the first investigation, we give a fighting chance for the methods to work by using strategies to reduce gaming

  - Tell subjects that options may be worded differently with the same meaning and the order of options may be shuffled.
  - Text-copying is disabled

Types of incentives

<3> <u>Bonus</u> incentive (B)

- ▶ Before the main section, tell subjects that they will receive a bonus of £3

- Keep total payment fixed, but some part is a surprise bonus
  - ▶ Everyone selects into the experiment at the same advertised base pay of £4
  - ▶ No incentive group is immediately told about higher total pay of £7 before consent

Why the Bonus Method could work

- A number of studies show that workers increase effort in response to employer's gifts, such as a surprise bonus

  (Gneezy and List, 2006; Kube, Marechal, and Puppe, 2012, DellaVigna et al. 2022)

- No study on whether this could be used to increase efforts on surveys

Benchmarks

<4> <u>Conventional</u> incentive

- ▶ Payments based on 3 randomly selected questions.
- ▶ £1 per correct answer; preference-based questions are paid out

<5> <u>No incentive</u> (Control)

- ▶ Subjects are given a higher show-up fee of £7 at the start of the survey
- ▶ Expected payoff is similar across all treatment arms

- **Language** in all treatments kept as similar as possible
  - ▶ "We use this procedure because we would like you to pay attention and put effort into answering the questions."
  - ▶ Addresses the concern that restatement method induces experimenter demand

Experimental design

Online survey - 4 stages & 5 treatments

(1) Background	(2) Main stage	(3) Spillover	(4) Review
	RM		
	RMP		
	B		
	CI		
	NI		

- **Background stage** includes demographic questions and an attention check question
- At the beginning of **Main stage**, different incentives are provided

Measuring quality of responses I

1. **Correctness** - increased share of correct responses on questions with objectively correct answers
   - Raven's matrices
   - First-order stochastically dominating lotteries
   - Simple math problems described in words
   - Basic cognitive reflection tests (CRT's)
   - Dot counting, etc.

Examples of correctness questions

A lottery ticket pays an amount of money that depends on chance. For example, lottery ticket X below pays £3 with probability 45%, £1 with probability 35%, and £6 with probability 20%.

	45%	35%	20%
Lottery X	£3	£1	£6

Below are 5 lottery tickets: A, B, C, D and E. Each row represents the payments for each ticket.

Please choose your preferred lottery ticket:

	45%	35%	20%
Lottery A	£1	£7	£4
Lottery B	£8	£4	£1
Lottery C	£4	£1	£7
Lottery D	£3	£8	£1
Lottery E	£1	£4	£8

Measuring quality of responses II

1. **Correctness** - increased share of correct responses on questions with objectively correct answers

2. **Consistency** - increased consistency when answering two (obfuscated) duplicate questions based on beliefs
   - The restatement method is potentially more useful for unverifiable questions
   - Internal consistency of answers has been used as a measure of survey quality (Meade & Craig 2012)
   - If beliefs are stable over the survey horizon, then any variation across multiple elicitations is a measurement error.
   - Questions are cross-randomized to be **verfiable or unverifiable**

Example of consistency questions

What proportion of total income in the U.S. do you think is earned by the 20% of people who earn the most (before tax and redistribution)?

0-1%	○
1-19%	○
19-36%	○
36-52%	○
52-68%	○
68-84%	○
84-99%	○
99-100%	○

What proportion of total income in the U.S. do you think is earned by the 80% of people who earn the least (before tax and redistribution)?

99-100%	○
81-99%	○
64-81%	○
48-64%	○
32-48%	○
16-32%	○
1-16%	○
0-1%	○

- For unverifiable version, "What proportion of income in the U.S. do you think **will go** to [*the top 20% of earners vs. the bottom 80% of earners*] **in 2032**?"

Additional effort measures

- **Spillover stage** involves unincentivized correctness or consistency questions
    - ▶ **Spillover**: How do answer qualities change once incentives are removed?

- **Response time**: does response time increase?

- **Review stage** involves re-stating their answers on 3 randomly selected questions as well as debrief questions
    - ▶ **Restatement success**: are respondents able to restate their answers better?
    - ▶ **Self-reports**: do respondents report increasing attention and effort?

Subjects

- 2,428 subjects across 8 treatments (299-307 each)

- US participants recruited on Prolific
  - A stronger test of the mechanisms as Prolific participants tend to have better answer quality
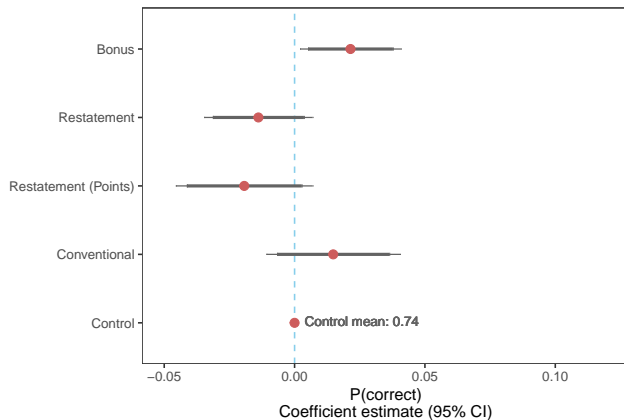
Main results

Impact on consistency
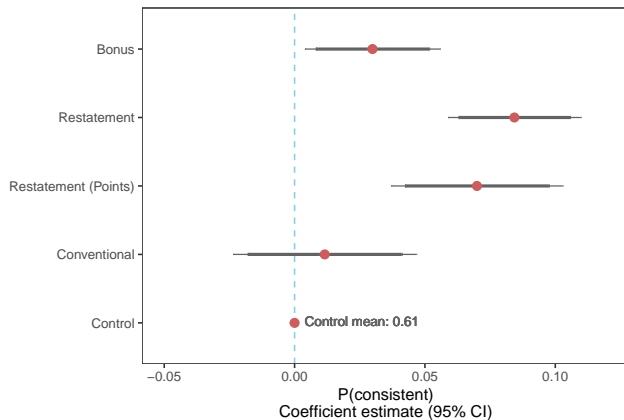
- RM increases consistency by 8.4 pp (13.7%, $p < 0.001$)
- RMP also increases consistency by 7.0 pp (11.5%, $p < 0.001$)
- B has a smaller but statistically significant positive effect (3.0 pp)

Impact on correctness

- RM and RMP do not have statatistically significant effect on correctness (if anything, slightly negative)
- B has a small positive effect on correctness (stat. sig. with LASSO controls)

Impact on consistency and correctness

	Consistent Answer (=1)			Correct Answer (=1)		
	(1)	(2)	(3)	(4)	(5)	(6)
Bonus	0.026*	0.026*	0.030**	0.006	0.006	0.021**
	(0.014) [0.062]	(0.014) [0.062]	(0.013) [0.024]	(0.012) [0.585]	(0.012) [0.586]	(0.010) [0.030]
Restatement	0.086***	0.086***	0.084***	-0.020	-0.020	-0.014
	(0.014) [<0.001]	(0.014) [<0.001]	(0.013) [<0.001]	(0.013) [0.110]	(0.013) [0.110]	(0.011) [0.191]
Restatement (Points)	0.054***	0.054***	0.070***	-0.020	-0.020	-0.019
	(0.017) [0.001]	(0.017) [0.001]	(0.017) [<0.001]	(0.016) [0.189]	(0.016) [0.189]	(0.013) [0.149]
Conventional	0.000	0.000	0.012	0.011	0.011	0.015
	(0.018) [0.997]	(0.018) [0.997]	(0.018) [0.520]	(0.014) [0.437]	(0.014) [0.437]	(0.013) [0.257]
Num. observations	9712	9712	9712	33992	33992	33992
Num. participants	2428	2428	2428	2428	2428	2428
Dep. var. control mean	0.61	0.61	0.61	0.74	0.74	0.74
R-squared	0.01	0.22	0.25	0.00	0.07	0.14
Question FEs		X	X		X	X
LASSO controls			X			X
p-val: Bonus = Conv.	0.14	0.14	0.14	0.74	0.74	0.74
p-val: Restate. = Bonus	0.00***	0.00***	0.00***	0.03**	0.03**	0.03**
p-val: Restate. = Conv.	0.00***	0.00***	0.00***	0.04**	0.04**	0.04**
p-val: Restate. = Restate (Points)	0.06*	0.06*	0.06*	0.99	0.99	0.99
p-val: Restate. (pooled) = 0	0.00***	0.00***	0.00***	0.07*	0.07*	0.07*

Notes: Standard errors clustered at the participant level.

Robustness checks

- Results appear robust to using alternate outcomes:
    - Anderson index (Z)
    - Intensive margin (0-1)

- Results appear robust to using alternate specifications:
    - Randomization inference
    - Controlling for baseline response time
    - Dropping those failing attention checks (13%)

- Results vary little by:
    - (For consistency:) Q verifiability, position of Q pair, or distance between Q pair
    - (For correctness:) Specific Q type

Impact in the spillover section

- Given no further incentive, only the RMs' effect on consistency persists.

|	Consistent Answer (=1)	Correct Answer (=1)
	(1)	(2)
Bonus	0.001	0.018
	(0.019) [0.956]	(0.017) [0.311]
Conventional	0.003	0.022
	(0.024) [0.900]	(0.021) [0.294]
Restatement	0.076***	0.007
	(0.020) [<0.001]	(0.018) [0.699]
Restatement (Points)	0.093***	0.000
	(0.024) [<0.001]	(0.022) [0.983]
Num. observations	2428	7284
Num. participants	2428	2428
Dep. var. control mean	0.83	0.71
R-squared	0.01	0.03
p-val: Bonus = Conv.	0.94	0.82
p-val: Restate. = Bonus	0.00***	0.53
p-val: Restate. = Conv.	0.00***	0.46
p-val: Restate. = Restate. (Points)	0.48	0.74
p-val: Restate. (pooled) = 0	0.00***	0.79

Impact on other effort measures

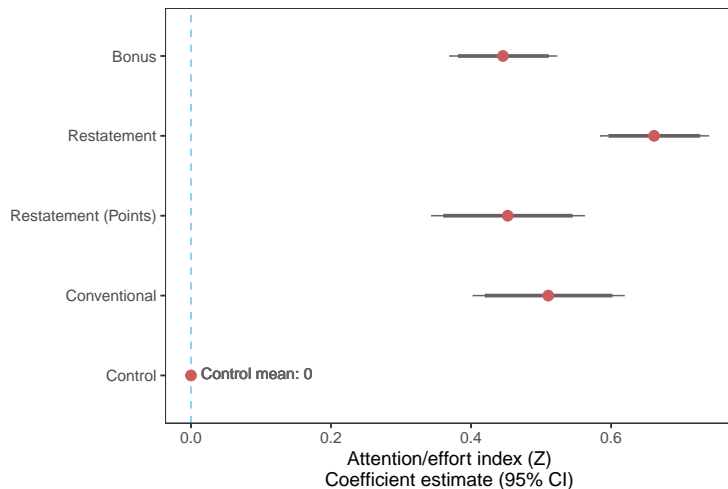# Are people spending longer time answering? Yes
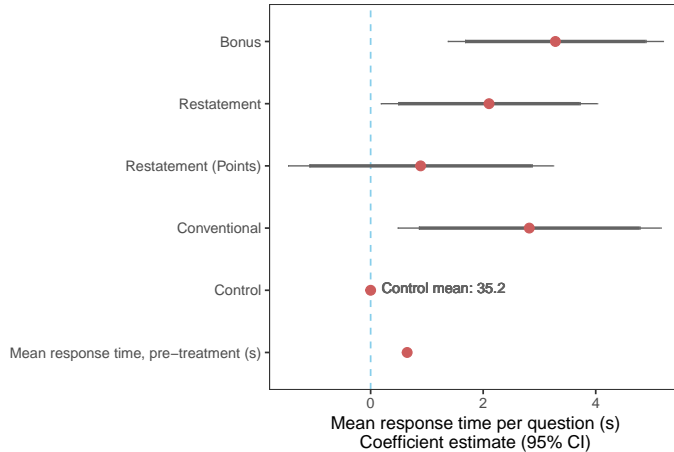
- Response time increases for all treatments in the main stage.

Are people more likely to restate the same answers? Yes
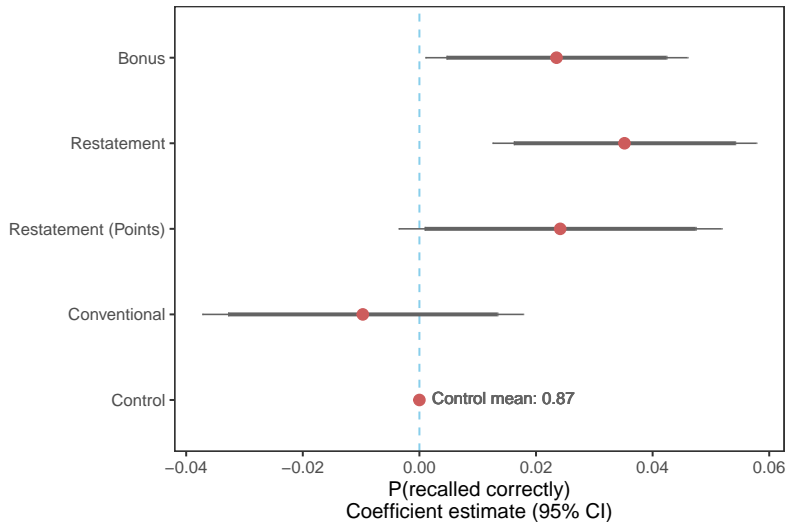
- B and RM increase restatement rates (2.4 - 3.5 pp, $p < 0.015$).

Do participants report increasing attention and effort? Yes

Attention/effort index (Z)
Coefficient estimate (95% CI)

Summary so far

- Overall these results seem to indicate that subjects increase attention and effort in response to the incentives.

- Yet, CI fails to generate any impact on correctness or consistency.

- B has a small positive effect on both measures of answer quality.
  - Is there any evidence of reciprocity?

- RM has no effect on correctness, but has a large effect on consistency.
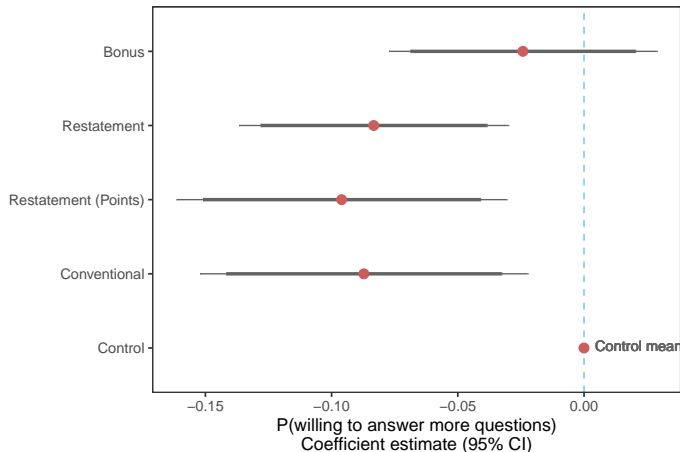 - Is answer quality actually improving?

Potential mechanisms

Evidence of reciprocity under Bonus
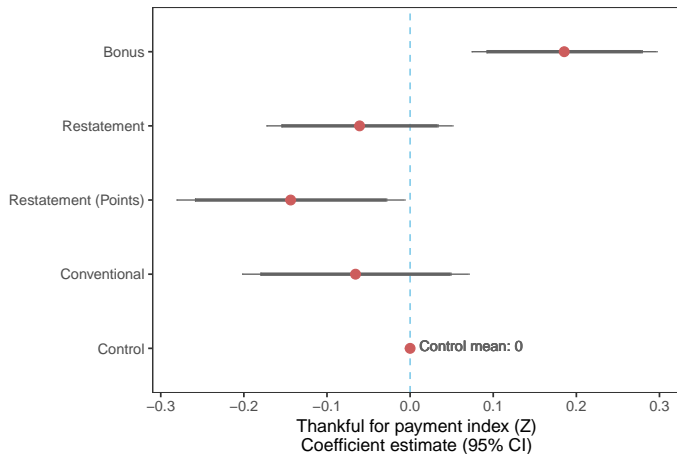
- Only B increases feeling thankful for the payment
- Notably, little evidence of reciprocal feelings under RM

Evidence of reciprocity under Bonus

- More B and NI subjects agree to answering extra questions as a favor. Note that some NI subjects were told about additional payments right before this question.

Are people gaming under RM?

Gaming Strategy	Measure	Restatement	Restatement (Points)	p-value of diff.
Choosing easy-to-remember answers	% questions	3.7%	2.4%	0.16
	% participants	14.5%	11.0%	0.14
Searching for answers online	% questions	0.8%	0.5%	0.52
	% participants	4.3%	3.3%	0.47
Writing answers down	% questions	2.4%	1.3%	0.15
	% participants	5.8%	3.3%	0.10

- While some people report choosing easy-to-remember answers or using other strategies, the share of affected answers appears small
  - In line with the findings that overall correctness does not go down
- Also some people report memorizing answers as a way to increase restatement rates.
- However, the increase in consistency is not driven by the subjects who report gaming or memorizing.
  - Correctness indeed decreases more for these subjects.

Supplementary experiment on RM

- Even without gaming, RM might still make subjects more likely to remember and recognize answers, increasing consistency

  - We devise a stronger test of whether RM increases answer quality based on correlation of answers

- Two treatments: restatement method-points (RMP) and no incentives (NI)

  - Approximately 450 subjects per treatment

- Instead of duplicate questions (that may be recognized as such), pairs of distinct questions that are expected to be correlated

  - e.g. "How much of a threat is global climate change to our country?" and "How important is making choices that help reduce the effects of global climate change for being a good member of society?"

  - Taken from Pew with correlations > 0.3

Results

| | Correct Answer ($=1$) | | Correlation pair: Answer 1 | |
	(1)	(2)	(3)	(4)
Restatement (Points)	-0.011	-0.011	0.007	0.007
	(0.015) [0.445]	(0.015) [0.445]	(0.019) [0.698]	(0.019) [0.699]
Answer 2			0.484***	0.484***
			(0.016) [<0.001]	(0.016) [<0.001]
Answer 2 $\times$ Restatement (Points)			-0.010	-0.010
			(0.023) [0.656]	(0.023) [0.669]
Num. observations	12698	12698	10884	10884
Num. participants	907	907	907	907
R-squared	0.00	0.11	0.23	0.23
Question FEs		X		
Question-pair FEs				X

Conclusion

- RM increases subject effort and internal consistency of answers.
  - But much of this effort seems to be channeled toward remembering answers.

- Bonus has small positive effects on both consistency and correctness.
  - Reciprocity appears to be one important channel.

- Overall, our work shows psychological forces can increase effort on surveys (without relying on additional payments).
  - Suggests the methods' potential as well as the potential of other psychological mechanisms.

Thank you for your attention!

Please send questions/comments to: suanna.oh@psemail.eu