Supplementary Material for "Placebo Statements in List Experiments (Evidence from a Face-to-Face Survey in Singapore)"

March 31, 2020

1 Survey details

All data were collected in person by a multi-ethnic team of enumerators comprised of university students from the National University of Singapore and Yale-NUS College. Enumerators were trained specifically for this project. Data collection took place between March 2 and April 18, 2019, typically on weekends (between 10am and 6pm) and weekday evenings (between 6pm and 8pm). We used a probability proportional to size strategy to approximate a representative sample. Clusters of 3 to 4 blocks were randomly selected using data provided by the government agency responsible for public housing (Housing and Development Board).[1] As in Wong (2013), we include only public housing residents in our study, primarily due to private housing being inaccessible to enumerators. This captures over 80% of Singapore's population. A balance table is included below (see Table 1 in Section 3),[2] which shows that our sample closely represents the public housing resident population on available observables.

One enumerator was responsible for an entire block. Enumerators knocked on all doors of selected blocks. When a household agreed to take the survey, the enumerator would describe the project and leave the questionnaire and a personal information sheet (i.e., consent form) with them, informing respondents that they would be back one hour later to collect the questionnaire. The questionnaire was two pages long (one sheet of paper), and took most respondents five to ten minutes to complete. All but two questions were close-ended (only "additional comments" and "please name your district" were open-ended). In 9% of cases, the respondent asked the enumerator to read out the questions and record their responses, which the enumerators always agreed to do. This occurred most frequently with elderly respondents. The survey was available in English, Chinese, and Malay. 82.9% of respondents completed the English version, while 15.8% and 1.3% chose the Chinese and Malay versions, respectively. That is, 23.4% of the Chinese respondents in our sample chose Chinese over English, whereas 9% of Malay respondents preferred to use the Malay version.

In total, enumerators knocked on 8,829 doors. Conditional on making contact with someone at the apartment (which happened 41% of the times), 38.56% of households

[1] The data is available online here: https://services2.hdb.gov.sg/web/fi10/emap.html

[2] This project was part of a series of projects on Singapore political attitudes and social cohesion. For more details on the survey methodology and the scope of the other projects linked to this one, we refer the reader to [Citation removed for anonymity].

agreed to take the survey. This is significantly higher than previous response rates for the Singapore Institute of Policy Studies, which uses phone calls.[3] In total, 1,278 responses to the list experiment were collected.

1.1 Details of the list experiment

The list experiment appeared at the end of the survey. All respondents received the same instructions: "Look at the following statements below. Can you tell us how many statements are true for you? **Please don't tick individual statements, just tell us the total number"** [Emphasis in the original].

Four neutral statements followed:

- I like durian very much.
- I have had a haircut within the last four days.
- I have eaten at a hawker centre at least once in the last two days.
- I have gone to a local "Meet the People" session at least once in the last year.

These statements were chosen following the generally accepted criteria for list experiments: they fit naturally into the context of the survey, they are uncorrelated (both with one another and with other broader socio-economic characteristics), and they are resistant to ceiling effects. The durian and hawker center statements are common enough to prevent floor effects.

The 4-item control group received these four statements only. The 5-item placebo group received a list which also included a statement that was designed to be false for all respondents: "I have been invited to have dinner with PM Lee at Sri Temasek next week."[4]

2 Supplementary Figures

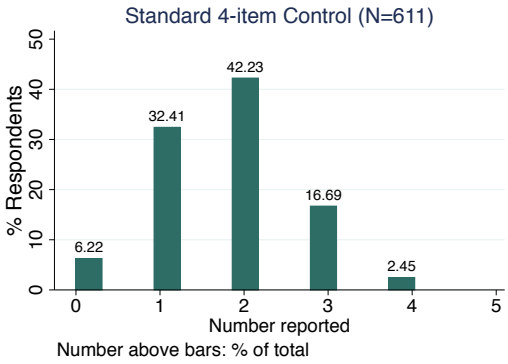
Figure 1 shows the frequency distribution of reported true statements across the standard and placebo control groups. Notable is that few respondents indicate '0' or '5' statements in the placebo group, which suggests that the presence of a clearly false placebo statement does not induce respondents to indicate extreme counts. The clearest inflation in the

[3]For instance, 24.6% in the Institute of Policy Studies "Post-Election Survey 2015" IPS (2015).

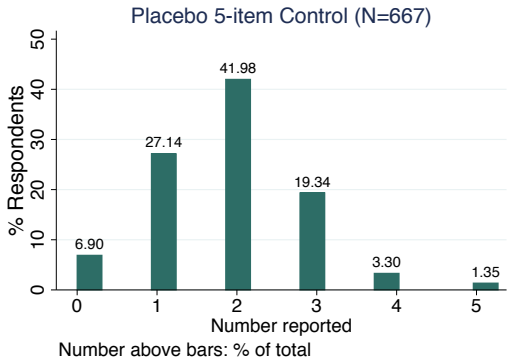
[4]Lee Hsien Loong is the Primer Minister of Singapore. Sri Temasek is the Prime Minister's official residence.

placebo group is with '3' reported true statements, while fewer respondents indicate '1' statement to be true. This may be partially a reflection of the perceived midpoint shifting to '3', as anticipated with satisficing (Krosnick, 1999; Kramon and Weghorst, 2012).

Figure 1: Distribution of reported true statements for each group



(a) Frequencies: Standard control group



(b) Frequencies: Placebo control group

3 Supplementary Tables

3.1 Balance table

Table 1 presents the descriptive statistics for both the standard and placebo control groups for respondents that are permanent residents and citizens, which comprise 92.73% of our total sample. We report only the resident statistics in the balance table, because official data on temporary residents is unavailable. While results reported in the main text reflect the full sample, the findings are robust when temporary residents are excluded.

3.2 Robustness checks

We check for the possibility that the placebo statement could trigger respondents to agree to another statement. Specifically, the Prime Minister placebo item could make some respondents think about politics, and thus more likely to indicate that they attended a "Meet the People" session. In other words, if the placebo item is related to another statement, then the observed inflation could result from a framing effect, rather than from the $J/J+1$ structure.

Table 1: Descriptive statistics: balance table

	Singapore	Control	
		Standard (4-item)	Placebo (5-item)
Age†			
% 61y.o.+	16.43	18.31	16.94
% Below 40y.o.	44.08	41.83	44.14
Ethnicity			
% Chinese	73.5	72.87	69.19
% Malay	15.6	11.88	13.87
% Indian	8.9	10.99	11.77
% Female	51.2	54.61	52.89
Type of property			
% in 1-2 Bedroom	7.37	7.80	5.79
% in 3 Bedroom	22.75	28.90	28.46
% in 4 Bedroom	40.25	34.57	39.71
% in 5+ Bedroom	29.50	28.72	26.04
% at least one car	32.8	28.37	29.10
% Employed	70.23	69.45	70.94
% Civil servants†	13.6	18.18	16.26
Education†			
% $\leq$ Secondary	36.44	36.94	36.99
% $\geq$ College	32.38	32.61	33.88
% Pol. knowledge	NA	37.59	31.03
# Observations		611	667

Unless otherwise stated, Singapore statistics are taken from the "HDB Sample Household Survey 2013", published by the Housing and Development Board (HDB (2014)). These statistics refer to the *resident population* (that is, full citizens and permanent residents)) in public housing. Superscript † denotes that, due to lack of available data, the statistic reported refers to Singapore as a whole. Age: Data taken from the Singapore 2010 census. "% 61y.o.+" and "% 18-40y.o" refer to the percentages with respect to the adult population (defined as 18y.o. or older). Type of property: Statistics Singapore (2016). Civil servants: Statistics Singapore (2015). Education: Statistics Singapore (2015). NA: official data not available (including for income).

It is not possible to deconstruct responses to the list experiment to directly test for this potential contamination. We can, however, assess whether such contamination takes place by means of the following question in the survey: "Do you participate in activities from neighborhood committees or organizations, for example Residents' Committees or Neighborhood Committees?" (The possible answers are "once per week or more", "a few times per month", "few times per year", "once a year or less", or "never"). Participation in such events is likely to be highly correlated with participation in local meetings with the district representative, which are organized by the same groups.

Table 2 summarizes the expected observable implications of a hypothetical contamination effect. If the "dinner with PM" item indeed triggers a positive response to the "local meeting" item, then we should see that the average for the 5-item control increases only among those for whom attending local meetings would be counted as false. On the other

hand, such contamination effect would not be observable among those who participate in local meetings often. This is because for them, the "local meeting" item would already count as "+1", so "dinner with PM" would not further increase their average response (if we consider contamination effects strictly as described above).

In other words, we should find that the average for the 4-item and the 5-item list experiments is the same among those who often participate, whereas we should find the average of the 5-item control to be strictly larger than for the 4-item control for those who rarely participate in local events. This difference, denoted by ε in Table 2, would capture the hypothetical contamination effect.

Table 2: Summary of expected results if *contamination effects* across items existed

Participation in local meetings	Average 4-item (standard control)	Average 5-item (placebo control)	Difference
Never/rarely participate	x	$x + \varepsilon$	ε
Often participate	$x + 1$	$x + 1$	0

Note: The 4-item average for the second group is $x + 1$ because they are expected to count as true the item "I have gone to a local *Meet the People* session at least once in the last year", whereas this item is expected to be counted as false for the first group.

We effectively check this by dividing the sample into those who "never" or "once a year or less" participate in local activities, and those who participate "a few times a month" or more. Table 3 shows the results. The difference between the average response in the 4-item vs. the 5-item lists is larger for those who do not actively participate in local activities. With that said, we should be cautious in noting that the sample size for the latter group is relatively small (91 observations in total).

Table 3: Results by level of participation in local activities

Participation in local meetings	Average 4-item (standard control)	Average 5-item (placebo control)	Difference
Never/rarely participate	1.72 ($N = 544$)	1.87 ($N = 606$)	0.15
Often participate	2.29 ($N = 48$)	2.12 ($N = 43$)	-0.17

Respondents sorted into either subsample depending on their answer to the following question:"Do you participate in activities from neighborhood committees or organizations, for example Residents' Committees or Neighborhood Committees?". Those who answer "never" or "once a year or less" are in the "Never/rarely participate" group. Those who answer "once per week or more" or "a few times per month" are in the "Often participate" group. Those who did not answer or answered "a few times per year" are in neither group.

Whereas a contamination effect is one possible interpretation for this, we believe that, in fact, these results reflect similar patterns as those observed in Table 1 in the main

manuscript: that is, respondents who are socially engaged (i.e., who participate more often in local/social activities) are less likely to mechanically inflate answers, in a similar fashion to those who are politically knowledgeable or the higher educated. We suggest that future research looks into this issue by using two different placebo statements in the same study.

Finally, Table 4 shows the results of the robustness check for the linear regression model presented in Table 2 in the main document. Here, we use the non-linear specification suggested by Imai (2011) and Blair and Imai (2012).[5] We estimate:[6]

$$LIST_i = f(X_i, \beta) + PLACEBO_i \times g(X_i, \gamma) + \varepsilon_i, \tag{1}$$

where f and g are non-linear functions that represent the regression models for the conditional expectations of the control and placebo items given the covariates. We estimate the model using the **R** package `'list'`. Table 4 below shows the results. They provide further support for the notion that educational attainment predicts vulnerability to mechanical inflation.

[5] We note that Ahlquist (2017) raises substantial concerns about this estimator.
[6] Expression (5) in Imai (2011), expression (4) in Blair and Imai (2012).

Table 4: Estimated coefficients from the item count technique regression models where the extra statement in the treatment group is designed to be false for all respondents (placebo). Maximum likelihood estimation of logistic models using the `'list'` package in **R** (Blair and Imai (2012)). See expression (1) for details on the specification.

	Constrained model				Unconstrained model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Placebo item								
Education	-0.119*	-0.174**	-0.075	-0.119	-0.016	-0.227**	0.125	-0.298**
	(0.062)	(0.068)	(0.089)	(0.091)	(0.093)	(0.099)	(0.139)	(0.119)
61+ years old			-0.153	0.241			0.522	-1.994*
			(0.975)	(2.111)			(1.144)	(1.112)
Hhd. Income			-0.125	-0.103			-0.188	-0.055
			(0.116)	(0.104)			(0.134)	(0.101)
Panel B: Control item					$h_0(y; x, \psi_0)$			
Education	0.012	0.018**	0.011	0.013	0.009	0.014*	0.007	0.014
	(0.008)	(0.008)	(0.010)	(0.010)	(0.007)	(0.008)	(0.010)	(0.010)
61+ years old			0.043	-0.025			0.046	0.120
			(0.113)	(0.127)			(0.096)	(0.097)
Hhd. Income			0.006	-0.001			0.005	-0.004
			(0.009)	(0.009)			(0.008)	(0.009)
					$h_1(y; x, \psi_1)$			
Education					-7.615	0.055	-0.476**	0.173*
					NaN	(0.040)	(0.243)	(0.090)
61+ years old							-2.039	2.169**
							(2.216)	(1.044)
Hhd. Income							0.249	0.027
							(0.196)	(0.069)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Log-likelihood	-1708.2	-1686.6	-1565.5	-1545.1	-1702.3	-1684.5	-1560.2	-1540.2
Observations	1256	1254	1158	1156	1256	1254	1158	1156

*** Significant at 1% level; **at 5% level; *at 10% level.

Standard errors in parenthesis. Dependent variable: number stated in the list experiment. Education: years of schooling. 61+ years old: dummy for being 61 years of age or older. Hhd. Income: monthly household income, in thousands of Singapore dollars. Controls: gender, ethnicity, socioeconomic status. To proxy for the latter, in (2) we use apartment size. In specifications (4), (6), and (8) we use car ownership instead of apartment size because the algorithm does not converge when using apartment size. Constrained model: control group parameters are constrained to be equal. See `http://imai.princeton.edu/software/list.html` and `http://imai.princeton.edu/talk/files/UCI11.pdf` for further details on the R package `'list'`. Webpages last accessed April 19, 2019. R package`'list'` version 8.3. The NaN in column (5) arises due to convergence issues (R version 3.3.3 (2017-03-06); RStudio Version 1.1.463).

3.3 Replication study using the data from Ahlquist et al. (2014)

This section is devoted to analyzing whether our general conclusions are robust to data from a list experiment set in an entirely different context. We use the data from Ahlquist et al. (2014), which is available online in *Harvard Dataverse*.[7] Their study is based on an online sample from the United States, conducted in 2013. The sample size is 3,000: 1,472 are in the 4-item control group, whereas the remaining 1,528 are in the placebo 5-item control group. Table 5 summarizes the results. For an easier comparison, we keep the same age, education and income partitions.[8]

Table 5: Results using data from Ahlquist et al. (2014). Mean number of reported true statements, by group. Number of observations are presented in parentheses.

		Control		
		Standard (4–item)	Placebo (5–item)	p-value
Whole sample		0.90 (1472)	0.96 (1528)	0.0412
Education	None or primary	1.07 (73)	1.20 (71)	0.2725
	Secondary	0.87 (816)	0.98 (858)	0.0181
	College or above	0.92 (583)	0.91 (599)	0.5354
Household income†	$<$ \$50K per year	0.84 (675)	0.86 (727)	0.3451
	$\geq$ \$50K per year	0.99 (587)	1.08 (577)	0.0499
Age	61+	1.00 (361)	1.21 (369)	0.0025
	$18-60$	0.87 (1111)	0.88 (1159)	0.3470

$p-$values for one-sided test of difference in means between the 4-item control and the 5-item placebo groups. E.g., 0.0181 is the p-value resulting of testing the difference in means between secondary school degree 4-item group and secondary school degree 5-item group. 'Secondary education': has a high school degree. 'College or above' includes two-year graduate degrees. †: Household median income in 2013 was \$51,939 according to the US Census Bureau – details can be found here: https://www.census.gov/library/publications/2014/demo/p60-249.html.

Table 5 shows support for our conclusions when using this new dataset. First, on average, there is an increase in the mean response rate of 0.07 points with just with the inclusion of the necessarily false placebo statement. Furthermore, there is support for our finding that low educational attainment and older ages are associated with a propensity to increase item counts in the presence of the placebo statement. On initial glance, the results for income appear reversed: those with higher income levels are more likely to

[7] https://dataverse.harvard.edu/dataverse/.

[8] Note that income is divided into 'below' and 'above' US 2013 median income. There is no information on political knowledge in Ahlquist et al. (2014).

inflate their reported number in the presence of a placebo statement. A closer look, however, reveals that this is due to the fact that mechanical inflation is prevalent also among those whose household income is around the median. When we divide the dataset into those above and below the 67th percentile, the picture approximates that of our study:[9] respondents in households below that cutoff point increase their mean response by 0.07 points (p-value: 0.046), whereas those in households above that cutoff point do not increase it at all. In fact, they decrease it by 0.04 points (p-value: 0.28). This suggests that, in essence, income plays a minor role in this study: mechanical inflation appears to be driven mostly by age and formal education levels.

4 Power Calculations and meta-analysis

Columns (1)–(3) in Table 6 summarize the findings in this paper, Kiewiet de Jong and Nickerson (2014) (referred to below as KJ-N), Ahlquist et al. (2014) and Holbrook and Krosnick (2010). Columns (4)–(5) summarize the power of each test assuming that the effect size is as observed (column 4), and again assuming a hypothetical 10% effect size (column 5). Both columns assume a one-sided test, with α set at 0.05. All computations are carried out using the `sampsi` command in STATA, and are detailed below.

This table suggests that inflation is more likely than not and about the size of many reported studies: those studies closer to the standard 80% power threshold are also the ones which most strongly suggest the presence of mechanical inflation effects. Furthermore, a meta-analysis combining all five studies suggests an average effect size of 7 percentage points (8 percentage points if studies are weighted by the number of observations).

4.1 Computational details

For column (4) in Table 6

```
** Assume effect size is what is observed
sampsi 1.767594 1.890555, n1(611) n2(667) sd1(.8861395) sd2(.9917153) a(0.05) onesided //
Singapore
sampsi 0.69 0.77, n1(301) n2(300) sd1(.89) sd2(.89) a(0.05) onesided // Uruguay
sampsi 1.90 1.91, n1(336) n2(336) sd1(1.19) sd2(1.19) a(0.05) onesided // Honduras
sampsi 0.899 0.961, n1(1478) n2(1528) sd1(0.899) sd2(1.04) a(0.05) onesided // USA--Ahlquist
et al.
** For Holbrook & Krosnick's, first calibrate unreported standard deviations to match their
reported t-statistic of t(1,510)=1.40, footnote 20, page 53.
```

[9] The 67th percentile is at $80,000 a year. Data from https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hinc/hinc-01.2013.html.

Table 6: Summary of results

	Controls			Statistical power	
	Standard (4-item)	Placebo (5-item)	Difference (p-value)	when effect size is... ...as observed	...10%
	(1)	(2)	(3)	(4)	(5)
Singapore	1.77 (0.89) N=611	1.89 (0.99) N=667	0.12*** (0.0100)	0.76	0.60
KJ-N (Uruguay)	0.69 (0.89) N=301	0.77 (0.89) N=300	0.08 (0.1810)	0.29	0.39
KJ-N (Honduras)	1.90 (1.19) N=336	1.91 (1.19) N=336	0.01 (0.4970)	0.06	0.29
Ahlquist et al. (2014) (USA)	0.90 (0.90) N=1472	0.96 (1.04) N=1528	0.06** (0.0412)	0.54	0.88
Holbrook and Krosnick (2010) (USA)	1.77 (1.2) N=769	1.86 (1.3) N=743	0.09* (0.0809)	0.40	0.46

Column (3): p-values refer to a one-sided t-test for difference in means. Columns (4)–(5): all details for power computations (using STATA) in Section 4.1. Results in columns (1)–(3) for Uruguay and Honduras are taken from Table 1 in Kiewiet de Jong and Nickerson (2014) (page 670). Note: standard deviation for the standard control group is not provided. Hence, we assume it is the same as for the placebo control group. Results for Ahlquist et al. (2014) are our own computations using the publicly available original dataset. Standard deviations for Holbrook and Krosnick (2010) are calibrated to match their reported means and t-statistic: t(1,510)=1.40, page 53, footnote 20.

```
ttesti 743 1.86 1.3 769 1.77 1.2 // 1.3 and 1.2 work
sampsi 1.77 1.86, n1(769) n2(743) sd1(1.2) sd2(1.3) a(0.05) onesided // USA-HK
```

For column (5) in Table 6

```
** Assume effect size is 10%
sampsi 1.767594 1.867594, n1(611) n2(667) sd1(.8861395) sd2(.9917153) a(0.05) onesided //
Singapore
sampsi 0.69 0.79, n1(301) n2(300) sd1(.89) sd2(.89) a(0.05) onesided // Uruguay
sampsi 1.90 2.00, n1(336) n2(336) sd1(1.19) sd2(1.19) a(0.05) onesided // Honduras
sampsi 0.899 0.999, n1(1478) n2(1528) sd1(0.899) sd2(1.04) a(0.05) onesided // USA--Ahlquist
et al.
sampsi 1.77 1.87, n1(769) n2(743) sd1(1.2) sd2(1.3) a(0.05) onesided // USA--HK
```

References

Ahlquist, John S., Kenneth R. Mayer, and Simon Jackman. 2014. "Alien Abduction and Voter Impersonation in the 2012 U.S. General Election: Evidence from a Survey List Experiment", *Election Law Journal,* **13 (4)**, 460-475.

Ahlquist, John S. 2017. "List Experiment Design, Non-Strategic Respondent Error, and Item Count Technique Estimators", forthcoming in *Political Analysis.*

Blair, Graeme, and Kosuke Imai. 2012. "Statistical Analysis of List Experiments", *Political Analysis,* **20 (1)**, 47-77.

Department of Statistics Singapore. 2015. "General Household Survey, 2015", *Department of Statistics, General Household Survey 2015, Singapore.* Available at `https://www.singstat.gov.sg/publications/publications-and-papers/GHS/ghs2015`, last accessed, January 11, 2017.

Department of Statistics Singapore. 2016. "Key Household Income Trends, 2016", *Department of Statistics, Singapore.* Available at `http://www.singstat.gov.sg/docs/default-source/default-document-library/publications/publications_and_papers/household_income_and_expenditure/pp-s23.pdf`, last accessed, March 11, 2017.

Holbrook, Allyson L, and Jon A. Krosnick. 2010. "Social Desirability in Voter Turnout Reports: Tests Using the Item Count Technique", *The Public Opinion Quarterly,* **74 (1)**, 37-67.

Housing and Development Board. 2014. "Public Housing in Singapore: Residents' Profile, Housing Satisfaction and Preferences (Sample Household Survey 2013)", *Housing and Development Board, Singapore.* Accessible at `http://www.hdb.gov.sg/cs/infoweb/monograph-1-29-dec-2014`, last accessed March 4, 2017.

Imai, Koisuke. 2011. "Multivariate Regression Analysis for the Item Count Technique", *Journal of the American Statistical Association,* **106 (494)**, 407-416.

Institute of Policy Studies. 2015. "POPS (8): IPS Post-Election Survey 2015". Available at `https://lkyspp.nus.edu.sg/ips/wp-content/uploads/sites/2/2015/10/POPS-8_GE2015_061115_web-Final.pdf` Last accessed, July 9, 2017.

Kiewiet de Jong, Chad P. and David W. Nickerson. 2014. "Artificial Inflation or Deflation? Assessing the Item Count Technique in Comparative Surveys", *Political Behavior,* **36**, 659-682.

Kramon, Eric, and Keith R. Weghorst. 2012. *Newsletter of the APSA Experimental Section,* **3(2)**, 14-24.

Krosnick, Jon A. 1999. "Survey Research", *Annual Review of Psychology,* **50**, 537-567.

Wong, Maisy. 2013. "Estimating Ethnic Preferences Using Ethnic Housing Quotas in Singapore", *Review of Economic Studies,* **80 (3)**, 1178-1214.