

Hicks, John; Bondi, Arianna; Gareau-Paquette, Thomas; Patterson, Scott

Working Paper

Reproducing Do Policymakers Listen to Experts? Evidence from a National Survey of Local and State Policymakers (Lee 2022)

I4R Discussion Paper Series, No. 186

Provided in Cooperation with:

The Institute for Replication (I4R)

Suggested Citation: Hicks, John; Bondi, Arianna; Gareau-Paquette, Thomas; Patterson, Scott (2024) : Reproducing Do Policymakers Listen to Experts? Evidence from a National Survey of Local and State Policymakers (Lee 2022), I4R Discussion Paper Series, No. 186, Institute for Replication (I4R), s.l.

This Version is available at:

<https://hdl.handle.net/10419/307157>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



No. 186

I4R DISCUSSION PAPER SERIES

Reproducing *Do Policymakers Listen to Experts? Evidence from a National Survey of Local and State Policymakers* (Lee 2022)

John Hicks

Arianna Bondi

Thomas Gareau-Paquette

Scott Patterson

December 2024

I4R DISCUSSION PAPER SERIES

I4R DP No. 186

Reproducing *Do Policymakers Listen to Experts?* Evidence from a National Survey of Local and State Policymakers (Lee 2022)

**John Hicks¹, Arianna Bondi¹, Thomas Gareau-Paquette¹,
Scott Patterson¹**

¹McGill University, Montreal/Canada

DECEMBER 2024

Any opinions in this paper are those of the author(s) and not those of the Institute for Replication (I4R). Research published in this series may include views on policy, but I4R takes no institutional policy positions.

I4R Discussion Papers are research papers of the Institute for Replication which are widely circulated to promote replications and meta-scientific work in the social sciences. Provided in cooperation with EconStor, a service of the [ZBW – Leibniz Information Centre for Economics](#), and [RWI – Leibniz Institute for Economic Research](#), I4R Discussion Papers are among others listed in RePEc (see IDEAS, EconPapers). Complete list of all I4R DPs - downloadable for free at the I4R website.

I4R Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Editors

Abel Brodeur
University of Ottawa

Anna Dreber
Stockholm School of Economics

Jörg Ankel-Peters
RWI – Leibniz Institute for Economic Research

Reproducing *Do Policymakers Listen to Experts? Evidence
from a National Survey of Local and State Policymakers*
(Lee 2022)*

April 17, 2024

John Hicks	Arianna Bondi	Thomas Gareau-Paquette	Scott Patterson
<i>McGill University</i>	<i>McGill University</i>	<i>McGill University</i>	<i>McGill University</i>

Abstract

Lee (2022) evaluates whether elected officials update their policy positions based on expert evidence. His cross-subject and within-subject designs run in the American local and state policy-making context both confirm the capacity for politicians to update their beliefs in response to expert evidence, cutting across party-lines and regardless of the valence of the information provided. Our replication finds that the study as published is nearly perfectly replicable following Lee's publicly-available code, with some minor departures that merit revision. Following replication, we offer suggestions for improved clarity of the code, greater transparency with the data accessibility, and clarification of the minor inconsistencies identified between the code and the published work.

1 Introduction

Lee (2022) argues that elected officials in the American context update their beliefs in response to being presented with information from experts. His argument rests on the idea that not only may

*John Hicks is a Master's student in political science at McGill University. Arianna Bondi is a PhD student in political science at McGill University. Thomas Gareau-Paquette is Master's student in political science at McGill University. Scott Patterson is a PhD candidate in political science at McGill University

these policymakers have malleable personal views that can be shifted by the introduction of evidence, but they also bear responsibility for the condition of those they represent. Thus, they shift position in response to evidence in the interest of their constituents, indicating a dual basis for belief modification: personal conviction influenced by expert evidence and the pragmatic considerations of representing their constituents' welfare.

In "Do Policy Makers Listen to Experts? Evidence from a National Survey of Local and State Policy Makers," Lee (2022) investigates this intersection of expert opinion and policymaker responsiveness via a national survey of local and state policy makers that presented policymakers with expert information on three issues — needle exchanges, GMO bans, and rent-control ordinances. These issues were selected for **1)** their relevance to the remit of the policymakers' legislative competences, **2)** their political contestation, despite the scientific consensus regarding their effectiveness and **3)** their varying degrees of congruence with the platforms of the Democratic and Republican parties. The survey began with an overview of the debates surrounding each issue. A subset of the participants (i.e., the treatment group) received additional detailed information on the scientific community's consensus for each issue. All respondents were then asked to reassess their personal position on the issue and to estimate the degree of consensus among experts.

Lee (*ibid.*) puts forward and examines two hypotheses:

1. *Direction Motivation Hypothesis*: Policy makers are less likely to update beliefs in the direction of evidence that goes against the party position;
2. *Accuracy Motivation Hypothesis*: Policy makers update their beliefs in the direction of the expert irrespective of the evidence's relationship to the party position.

To evaluate these hypotheses, the author employs a *cross-subject design*, which enables examination of differences in beliefs updating based on party affiliations, alongside a *within-subject design*. The latter involves probing control group respondents a second time, after presenting them with additional information, to measure the improvement in their accuracy in understanding scientific consensus (excluding those that are already fully accurate in their assessment). The necessity for these two approaches stems from the requirement in within-subject testing to maintain a pre-post design, ensuring the control group remains unaffected by any intervention prior to their subsequent evaluation. This separation is crucial for the integrity of the experimental design.

Lee (*ibid.*)'s design features a number of other particularities to ensure clean testing of the treatment effect, including randomization of issue order and symmetry in the placement of correct responses on the expert consensus question, among others.

Lee finds that exposure to expert information increases belief accuracy between 8 and 16 percentage

points and congruence with the expert-endorsed policy preference by between 4 and 16 percentage points, regardless of partisan affiliation. While he argues this as evidence against the notion that directional biases drive policy makers' position formation, he adds substantial caveats pertaining to the external validity of these findings.

Lee acknowledges that real-world policymaking, fraught with diverse pressures and information sources, might dilute the direct impact of expert advice. Yet, Lee also contends that any limitations regarding the study's external validity (such as the understated nature of the information treatment and potential discrepancies in survey responses versus real-world motivations) would more likely underestimate rather than exaggerate the observed effect of expert information on policymaker opinions. This suggests a conservative estimate of the impact of expert evidence on policy formation, reinforcing the significance of these findings despite these acknowledged limitations.

Replication data for the study is available via the Harvard Dataverse. As all replication data and code was made available in R, we used this code for our replication and did not translate it for any other software. We requested access to the original, unprocessed data, as it was evident that some measure of pre-processing had occurred before the dataset was submitted to the Dataverse. However, this request for data and clarification of coding decisions was not answered by the author as of April 2024.

2 Computational Reproducibility

2.1 Successful Reproduction

The vast majority of the provided R script was reproducible without issue. This included the remaining pre-processing of data, as well as the generation of figures. Figures 1, 2, 3, and 4 from the article all replicate properly, with the exception of a non-substantive, stylistic disparity noted in the following section. From the Appendix, Figures A3, A5, A7, A8, A9, A10, A11, A14, A15, A16, A17, A20, A23 were also successfully replicated without any significant issues.

2.2 Minor Discrepancies

In certain cases, there were minor, non-substantive disparities either between the R script and the published output, or between the R script and the replication files.

Regarding the former, our initial confusion arose from how the replication data arranged the categorical x-axis in plots, which differed from the paper's presentation. On initial review, the estimates and confidence intervals in Figure 4, along with similar figures in the Appendix, seemed to deviate significantly from those reported in the publication. However, upon closer inspection it became obvious

that contrary to paper, the categories for congruence on the axis were arranged from least to most congruent in the replicated figures. We contend that such formatting is actually more intuitive than the order in the paper. As such, these findings were, in fact, consistent. Therefore, we recommend explicitly defining the order of these categories in the code to prevent any confusion that might arise from the default ordering, ensuring clarity and replicability.

With respect to the latter, there were sparse instances of variable miscoding. For example, the R script repeatedly references a *bias* variable, which is not present in the dataset. Thus, we recoded the dataset's variable *partisan_bias* to fit the code, assuming the two to be the same. Relatedly, we had to alter the code to create a *bias_num* for the replication of Figure 4, which originally drew errors as existing code drew from a non-numeric `if-else` condition that could not run properly.

These departures from the published work and inconsistencies between the code and dataset are in no way suggestive of data manipulation. They are merely indicative of improper cleaning of the code prior to provision of the replication data. Materials in the publication were likely produced in a different session from when the replication data was assembled. It would have been ideal to ensure that prior to providing these materials the replication was exact. Nonetheless, they do not cast any doubt on the study, nor did they impose significant delays in running the replication. On a similar note, we notice that the order in which figures and table are generated in the R script does not match the order of presentation in the paper and the Appendix. While this does not constitute a significant impediment to replication, it does make the process more time-consuming. Hence, re-ordering the replication files to match the output of the paper would ease the replication tasks. This exercise could be undertaken while addressing the gaps presented in the next section of this report.

We also note that the replication file begins with installation of the `doStata` R package, which, during our first attempt at replication, was not installable and appeared to have no documentation available online. Subsequently, we discovered that the package could be installed from a GitHub repository not obviously associated with the original source (danmckinleythompson/doStata). This suggests that the package may have migrated to this repository during our replication efforts, serving as a cautionary note for future replication efforts about the potential pitfalls of depending solely on CRAN for package availability. However, the package's intended use in the study needs to be clarified. Therefore, it appears to be an unnecessary addition.

2.3 Reproducibility Failures

There were some cases of substantive differences between the paper output and the output from replication. These inconsistencies do not undermine the entire study, but we present them as reason for greater attentiveness and response by the author.

Multiple tables from the Appendix do not match their counterparts in the replication material. Although the replicated tables resemble those in the Appendix in terms of labels and structure, there are discrepancies in the actual figures. This includes tables A7, A14, A21, A22, A23, and A24, all of which demonstrated departures from the published Appendix.

For example, when replicated, the code we assume to match with table A7 has multiple inaccuracies. Terminology such as `local area` is used instead of `sub-county`, and `district` appears, a term absent from the paper's Appendix. Furthermore, one term is omitted, and the other is renamed in the replicated tables. Beyond these terminological differences, the numerical values also showed inconsistencies. For instance, the `Rent Control-State` variable was reported as 0.91 in the publication but appeared as 1.00 in the replication. Similarly, the values for `state` in the Appendix matched those labeled as `district` in the replication dataset, further indicating a misalignment between the original and replicated data.

Table A22 nearly replicates as in the paper, however its sample counts and proportions are off. This divergence suggests potential issues in the data processing, either during the pre-processing phase or within the main code, leading to differences from the analysis employed in the original paper.

A caveat we must note with all of these reproducibility failures is that they largely pertain more to effect sizes than to significance or direction, with the general patterns remaining consistent in the data. They therefore do not fundamentally compromise the arguments made in the study or the robustness/descriptive checks used to increase confidence in the study. They merely call into question the extent to which readers should be convinced of these checks' ability to fully allay concerns of confounders or counter-arguments to that which is proposed.

Some tables could not be replicated due to issues with the provided script. For instance, the code in the script pertaining to "Ceiling Effects Tables" drew unsolvable errors as the `condition` variable used in their creation does not exist within the provided dataset. Consequently, these tables couldn't be assessed for consistency with the published findings. The code output for one of these tables appears to match table A19, yet it is unclear precisely what these tables should match within the Appendix.

There was also code to generate multiple figures and tables that seemingly have no place in either the paper or the Appendix. This includes the very last line of code, which generates the output displayed in Table 1 below:

Table 1: Extra Table Produced in Replication Code

	issue	party_bin	mean
1	Needle Exchange	Democrat	-0.77
2	Needle Exchange	Republican	-0.57
3	Needle Exchange		-0.34
4	GMO Ban	Democrat	0.50
5	GMO Ban	Republican	0.42
6	GMO Ban		0.19
7	Rent Control	Democrat	0.81
8	Rent Control	Republican	0.74
9	Rent Control		0.37

This table does not exist within the paper or Appendix, and the lack of clean variable relabelling suggests it might be a residual piece of code from an earlier phase of analysis that was inadvertently left in the final version. Inclusion of such unexplained elements in the script should either be clarified with accompanying notes for the benefit of future researchers or removed to prevent confusion regarding the study’s replicability, particularly if deemed superfluous.

3 Data Processing Stage

Though — as detailed in the above section — Lee (2022)’s code demonstrates a high degree of replicability, we must note that the replication code does have a questionable element in the stage of the data processing. The dataset provided to researchers is in a semi-processed form, with certain variables modified or partially aggregated. There is no clear indication that such a decision was made to preserve respondent anonymity, for example, so it would be both more transparent and more intelligible to those replicating to have the raw (anonymized) data provided, instead of a semi-processed dataset, which in its existing state is less navigable and limits the capacity of researchers to verify the accuracy of the codes and design claims. Of course, a pre-processed dataset may streamline replication and be desirable to researchers for particularly computationally intensive projects. However, in such cases, it would be ideal to have full documentation of the processing that went into that dataset, for full transparency, even if not necessary for the analysis itself. As noted in the introductory section, we made an effort to contact the author in hopes of either clarifying the stage of data processing or obtaining the full unprocessed data; however, this request was not answered.

4 Conclusion

To conclude, we reiterate that the vast majority of Lee’s work stands out for its replicability. Beyond its significant contributions to the field, Lee’s work is to be applauded for its thoroughness, which is evident not only in the extensive checks included in the paper and its Appendix, but also in the straightforward, comprehensive code made available to researchers. While there are minor discrepancies between the publication and the provided code, the documentation is largely complete. Nonetheless, the lack of access to the original, unprocessed dataset poses challenges for conducting in-depth analyses and evaluating the study’s robustness more thoroughly. The existing discrepancies between the paper and the replication material, while not detracting from the publication’s overall validity — as they pertain more to effect sizes than to significance or direction, with the general patterns remaining consistent — call for further clarification. Should these differences be attributable to simple clerical error or specific stages of data processing that are not accessible for external review, addressing them would solidify confidence in the study’s findings.

5 Summary of R coding notes

Table 2: Consolidated Replication Status and Notes

Element Name	Replication Status	Replication Notes
fig_pref_control_by_party [Fig 1]	Successful	Removed + sign because ggsave can’t accept this and draws error
fig_effects.png [Fig 2]	Successful	Removed + sign because ggsave doesn’t accept it
fig_effects_by_party [Fig 3]	Successful	None
fig_updating_by_prior [Fig 4]	Successful	Figure replicates successfully, but the x-axis order is different. Paper x-axis order is from somewhat to most, replication is least to most.
Continued on next page		

Table 2 – continued from previous page

Element Name	Replication Status	Replication Notes
fig_bias	Minor Discrepancy	Figure 4 initially doesn't replicate because <code>df\$bias_num</code> has all values of 0. This occurs because the variable used for <code>bias_num</code> is <code>ifelse</code> on a character variable, when <code>bias</code> (taken from the <code>partisan_bias</code> variable it should be) is numeric. After fixing this, replication isn't identical. Compared to figure A4 in the appendix, all Dem. columns are higher. Removed + sign.
fig_deference	Successful	Removed + sign. Figure replicates successfully, matching appendix A5.
fig_polarity	Successful	Removed + sign. Figure replicates successfully.
fig_levels	Successful	None
fig_levels_by_party	Successful	Removed + sign. Figure replicates successfully (matching appendix A7).
fig_levels_by_prior	Successful	Removed + sign. Figure replicates successfully to match A8.
fig_levels_by_staffer	Successful	Removed + sign. Figure replicates successfully to match A13.
fig_levels_by_level	Successful	Removed + sign. Figure replicates successfully to match figure A12.
fig_levels_by_familiarity	Successful	Removed + sign. Figure replicates successfully to match figure A9.
fig_levels_by_bias	Successful	Removed + sign. Figure replicates but doesn't match any figure in the paper/appendix. No figure has numerical x-axis.
fig_levels_by_deference	Successful	Removed + sign. Figure replicates successfully to match figure A11.
Continued on next page		

Table 2 – continued from previous page

Element Name	Replication Status	Replication Notes
fig_levels_by_age_bin	Successful	Figure replicates successfully to match Figure A15.
fig_levels_by_exp_bin	Successful	Removed + sign. Figure replicates successfully to match figure A16.
fig_levels_by_gender	Successful	Removed + sign. Figure replicates successfully to match figure A14.
fig_levels_by_method	Successful	Removed + sign. Figure replicates successfully to match figure A17.
Ceiling Effects Tables	Failure	Above code doesn't replicate (df\$condition doesn't exist).
tab_effects_beliefs_CP18	Successful	"out" doesn't work but .tex text matches Table A8.
tab_effects_preferences_CP18	Successful	"out" doesn't work but .tex text matches Table A9.
tab_effects_staffer	Successful	Out doesn't work, but .tex output matches Table A20.
tab_effects_urban	Minor Discrepancy	Replication doesn't quite match any table in the paper/appendix. It looks like it should match A22 but the sample counts and proportions are different.
tab_effects_college	Minor Discrepancy	Replication doesn't quite match any table in the paper/appendix. It looks like it should match A23 but the sample counts and proportions are different.
tab_effects_population	Minor Discrepancy	Replication doesn't quite match any table in the paper/appendix. It looks like it should match A24 but the sample counts and proportions are different.
Continued on next page		

Table 2 – continued from previous page

Element Name	Replication Status	Replication Notes
tab_effects_voteshare	Minor Discrepancy	Replication doesn't quite match any table in the paper/appendix. It looks like it should match A21 but the sample counts and proportions are different.
tab_effects_ideo	Successful	Table replicates successfully to match table A14.
tab_effects_belief_bin	Successful	Table replicates successfully to match Table A16.
tab_effects_preference_bin	Successful	Table replicates successfully to match Table A15.
tab_update_beliefs_CP18	Successful	Table replicates successfully to match table A10.
tab_update_preferences_CP18	Successful	Table replicates successfully to match Table A11.
tab_update_party	Successful	Table replicates successfully to match table A17.
tab_update_beliefs_CP18_with_weights	Successful	Table replicates successfully to match table A12.
tab_update_preferences_CP18_with_weights	Successful	Table replicates successfully to match table A13.
Familiarity Table 1	Minor Discrepancy	Replicates table A7 with inaccuracies. Term "sub-county" is used in appendix, not local area. Term district used in code, not in paper appendix. Appears one of district/local area is removed for appendix and the one left is renamed. Even so numbers don't match (ex: Rent Control-State is 0.91 in paper, 1.00 in this code).
Familiarity Table 2	Failure	Table doesn't match anything in the paper/appendix.

References

Lee, Nathan (2022). “Do Policy Makers Listen to Experts? Evidence from a National Survey of Local and State Policy Makers”. In: *American Political Science Review* 116.2, pp. 677–688.