Online Appendix

How Deliberation Happens

# Studies And Cases

The data used for the analysis in the main text draws on nineteen deliberative forum cases that were conducted as part of twelve studies—where multiple cases within a study share the same topic and survey instrument. Two studies (Uppsala Speaks and Human Genome Editing) also include a control group, where participants were surveyed at the same pre- and post-deliberation time point as the other participants, without participating in the deliberative forum or receiving any information about the topic. These control cases are not included in the MLM analysis.

## Cases and Sample Size

Table A.1.1 shows 19 separate cases used in the MLM analysis, along with two control groups. Detailed information about the study that cases are conducted within can be found in Table A.1.2.

Cases within a study share the same overall design, except where specified, and survey instrument. Case 7 is the same design as case 6, except participants were recruited from land management professionals and decision makers, whereas case 6 involved randomly selected lay citizens; Case 11 is the same design as case 10, except participants were recruited from patient advocacy groups, rather than being lay citizens; Cases 14-16 follow the same design, except that breakout discussion groups were allocated according to difference in position for 14, and similarity in position for 15 (see Table A.1.1).

Table A.1.1: Studies and Cases

|  |  |  |  |
| --- | --- | --- | --- |
| Studies | Cases | Data points (n†) | Participants (N‡) |
| 1. Uppsala Speaks | *01 Control* | *20* | *20* |
| 1 Group Briefing | 22 | 23 |
| 2 Group Building Plus | 26 | 27 |
| 2. Far North Queensland Citizens’ Jury | 3 FNQCJ | 12 | 12 |
| 3. Australian Citizens’ Parliament | 4. ACP | 52 | 152 |
| .4 Sydney Climate Change Adaption | 5 Sydney CC | 21 | 24 |
| 5. ForestERA | 6 Citizen | 12 | 12 |
| 7 Stakeholder | 12 | 86 |
| 6. Biobanking, Regulation | 8 UBC | 19 | 20 |
| 9 Mayo | 18 | 20 |
| 10 WA Citizens | 10 | 15 |
| 11 WA Stakeholders | 16 | 26 |
| 7. Fremantle Bridge 21st Century | 12 Fremantle Bridge | 41 | 165 |
| 8.Climate Change and the Public Sphere | 13 CCPS | 34 | 34 |
| 9.Energy Futures Study | 14 WA | 22 | 23 |
| 15 NSW | 14 | 18 |
| 16 Vic | 16 | 19 |
| 10.Valsamoggia Council Amalgamation | 17 Valsamoggia | 16 | 16 |
| 11.Great Barrier Reef Futures | 18 GBR Futures | 7 | 7 |
| 12. Human Genome Editing CJ | *019 AusCJ Control* | *19* | *19* |
| 19 AusCJ | 17 | 23 |
| Total | (Excluding Control Groups) | 387 | 722 |

Note: † n = number of participants included in analysis; ‡ N = number of forum participants.

Table A.1.2: Study Information

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Study | Description | Participedia Link | Year | References |
| 1. Uppsala Speaks | Three-day minipublic developing recommendations for policies dealing with the influx of internal EU migrants engaging in street begging. | <https://participedia.net/case/5923> | 2016 | Jennstål (2018; 2019) |
| 2. Far North Queensland Citizens’ Jury | Four Day minipublic concerning the future of the Bloomfield Track | <https://participedia.net/case/38> | 2000 | Niemeyer (2004) |
| 3. Australian Citizens’ Parliament | Three-day minipublic to consider reform to Australia’s political system | <https://participedia.net/case/27> | 2009 | Carson et al. (2013) Hartz-Karp et al. (2010) |
| 4. Sydney Climate Change Adaption | Three and a half day minipublic concerned with climate change adaptation strategies for local government | <https://participedia.net/case/5970> | 2013 | Schlosberg et al. (2017) |
| 5. ForestERA | Three-day deliberative process considering land management options in northern New Mexico, intermediated using GIS mapping techniques | <https://participedia.net/case/5968> | 2007 | Hobson (2009) |
| 6. Biobanking, Regulation | Four day (over two weekends) deliberative process concerned with regulation of collection and testing of genetic data | <https://participedia.net/case/5> <https://participedia.net/case/5975> | 2007/8 | Secko et al. (2009) |
| 7. Fremantle Bridge 21st Century | Single day minipublic on options for the replacement of a heritage bridge between Perth and Fremantle | <https://participedia.net/case/4429> | 2006 | Niemeyer et al. (2013) |
| 8.Climate Change and the Public Sphere | Three-day minipublic concerned with climate change adaptation policy. | <https://participedia.net/case/5924> | 2010 | Hobson and Niemeyer (2011) |
| 9.Energy Futures Study | Three-day minipublic considering technological options for energy generation. | <https://participedia.net/case/5928> | 2005 | Ashworth et al. (2010) |
| 10.Valsamoggia Council Amalgamation | Two-day minipublic concerning a proposed amalgamation among four local councils in the Bologna region of Italy. | <https://participedia.net/case/5937> | 2013 | Felicetti et al. (2016) |
| 11.Great Barrier Reef Futures | Two-day minipublic run as a pilot study at James Cook University on management of the GBR | <https://participedia.net/case/8017> | 2019 | In Preparation |
| 12.Human Genome Editing CJ | Three-day minipublic on regulating human genome editing technologies | <https://participedia.net/case/7660> | 2021 | In Preparation |

Note: \*More detailed information for each case can be found via the links provided.

# Survey Instrument

To measure DRI it is necessary to design instruments to gather information covering relevant considerations and policy options specific to each case. As noted in the main text, our data set covers the totality of such cases where appropriate data is available. It is generally not possible to use data from pre/post deliberation questionnaires designed for other purposes.

## Developing Surveys

Although DRI is a stand-alone method, the approach used in Q methodology provides a template for generation of the “considerations” statements for a DRI survey. Q methodology involves sampling statements to capture a larger “concourse”—the volume of communication applicable to a situation (Stephenson 1986). This is consistent with our position that deliberation ideally involves all individuals incorporating all such considerations relevant to the issue at hand into their reasoning, which in turn should inform their preferences for particular policies or actions (Niemeyer 2020). In Q methodology, propositional statements can be sampled from existing public discourse, generally using a sampling frame specific to the issue at hand (Brown 1970).

The consideration surveys follow this approach for all our studies. But the specific mechanism for collecting statements varies. For example, for the FNQCJ case, statements were collected via focus groups prior to the deliberative forum. In contrast, for the AusCJ case, an anticipatory case where comparatively little public discourse exists, a more extensive process of statement collection was necessary, involving expert interviews, literature search, and media search.

### Study Survey Instruments

The complete set of survey items used for all studies can be obtained from <https://doi.org/10.7910/DVN/6OKNOY>. Alternatively, Table B.1.1 provides links to downloadable surveys for data for each case that have been uploaded at *Participedia*. A few examples of consideration statements and the full set of policy preference options for the FNQCJ study are provided in the main text.

Table B.1.1: Study Survey Instruments

|  |  |
| --- | --- |
| Study | Survey Instrument Download† |
| 1. Uppsala Speaks | https://tinyurl.com/kfe6esx |
| 2. FNQCJ | https://tinyurl.com/yhz9bxuh |
| 3. Australian Citizens’ Parliament | https://tinyurl.com/yckbtf8a |
| 4. Sydney Climate Change Adaption | https://tinyurl.com/2euc326f |
| 5. ForestERA | https://tinyurl.com/35xzx5zn |
| 6. Biobanking, Regulation | https://tinyurl.com/3cvkmwu5 |
| 7. Fremantle Bridge | https://tinyurl.com/3cvkmwu5 |
| 8. Climate change and the public sphere (CCPS) | https://tinyurl.com/4f5ne2jx |
| 9. Energy Futures Study | https://tinyurl.com/yc22knuk |
| 10.Valsamoggia Council Amalgamation | https://tinyurl.com/2p95sypy |
| 11. Great Barrier Reef Futures | https://tinyurl.com/4j3t3t8m |
| 12. Human Genome Editing CJ | https://tinyurl.com/ms7cdcp7 |

Notes: †Links valid 7 February 2022.

## Generating Survey Data

### Surveying Consideration Opinions

Opinions regarding the set of considerations can be surveyed using a Likert response along a disagree-agree scale—usually either a 5-point or 7-point scale, but up to 11-point for some of our cases. However, although Likert rating produces reasonable results, we have found that obtaining opinion on considerations using quasi-ranking produces stronger results. Quasi-ranking is commonly used in Q methodology, where responses are sorted into a pre-defined distribution. The shape approximates a normal distribution across a series of (usually up to 11) categories (columns) from “most disagree” to “most agree.”

Generally, the difference between forced and free ranking of statements has few consequences. We tested the difference, obtaining both types of response by first asking participants to provide a Likert response to each consideration statement, prior to their quasi-ranking. The correlation between the resulting DRI scores using these collection methods is 0.93.

### Surveying Policy Preferences

Policy preferences are obtained using a ranked ordering of the available options (usually between 5 and 10) in the form of actions or policies.

# DRI Method: Calculations and Procedures

## Procedure for Calculating DRI

Here we expand on the procedure for calculating DRI. Figure C.1.1 is a more detailed version of Figure 1 in the main text. The figure plots intersubjective consistency for four sample individuals (A,B,C,D) from the FNQCJ case.

The DRI method involves aggregating geometric distances rather than using a least squares method (such as Lin’s Concordance) to ascertain the consistency relationship between considerations and preferences. This approach overcomes the problem of domain restriction (Meade 2010). The effect of domain restriction is illustrated by comparing the left-hand side of Figure C.1.1, which reports data from a subsample of actual pairs from the FNQCJ case (reported in Figure 1 in the main text) to the right-hand side of the figure, which illustrates a hypothetical “strong consensus” case where the same individuals increase their overall agreement regarding considerations and preferences but maintain the same level of intersubjective consistency. If we were to use a least squares method, the strong consensus case would result in a lower aggregate result, even though there has been no actual change in level of consistency. Hence our method involves aggregation and standardization of intersubjective consistency, measured as the distance from the 1:1 line, using the following steps:

1. Correlating pairs of individuals’ survey responses to considerations, and also (separately) pairs’ policy preferences using Spearman correlation, ρ.
2. Calculating intersubjective consistency for all combinations of pairs of individuals.
3. Calculating DRI for each individual by aggregating intersubjective consistency for all pairs including that individual (this data is used in the MLM analysis).
4. Aggregating individual DRI to produce DRI for the group (reported in Table 3 in the main text).

In step 1, the intersubjective correlations for both considerations () and preferences () are calculated for all sets of pairs (J = [AB, AC, AD, BC, BD, CD]) using Spearman’s rank correlation. For *n* = 4 individuals in the example, the total number of pairs ) can be calculated:

.

Spearman ρ is used because it is appropriate for correlating non-linear relationships and ordinal data such as Likert scale ranks (Lehman et al. 2013, 127). These calculations were performed using the R script (data step 1; see Dataverse repository <https://doi.org/10.7910/DVN/6OKNOY>).

Figure C.1.1 plots all six example pairs (i, j), based on their Spearman correlations (and ).

Figure C.1.1: Detailed Illustrative Intersubjective Consistency Plot



Step 2 calculates the distance of all pairs from the 1:1 line (intersubjective consistency), as shown in Figure C.1.1. The modal orthogonal distance (*d*) is obtained for a given pair (*i*, *j*) using . In the left hand side of the figure, the distances for all example pairs (*da,b*, *da,c*, *da,d*, *db,c*, *db,d*, *dc,d*) are indicated using dashed lines.

Step 3, calculating individual DRI (DRI*Ind*) for each member of the group (*i*), involves finding the average modal distance (, where: for all pairs that include individual *i*; (*Ji*; *j*…*mi*; where *mi* = *n*-1)

Individual DRI (DRI*Ind*) is then calculated by transforming into a -1 to 1 scale using the theoretical maximum average distance for the group—or Lambda (*λ* = ; where λ = or 0.71).

Lambda, which is graphically indicated in Figure C.1.1, provides the upper boundary for average distance. The tendency for the maximum average distance to converge (on an asymptote) on a maximum value less than 1.41—i.e., it is not possible for all values to fall on the extreme corners of the intersubjective consistency plot—is because of the interrelationships between different pairs of individuals. For example, in Figure C.1.1, if individual A transforms their position, decreasing their consistency with C toward one of the corners of the graph, this will also transform their consistency with the remainder of the group. There is a very strong chance that their consistency with D, which previously fell outside λ, will correspondingly decrease.[[1]](#footnote-2)

The tendency to converge on an asymptote on Lambda has been tested by generating hypothetical cases using Monte Carlo simulation. Convergence occurs very quickly, where average distance asymptotes on λ as n→∞, once the items capturing considerations and preferences and number of individuals are all greater than 2.

Taking Lambda into account, DRI for individual *i* is calculated using:

In Step 4 DRI is calculated for the group by averaging DRI*Ind*:

for all *i* (*i*…*n*)

## Data and Calculations

Calculation of individual DRI can be performed using the dataset provided in the attached Excel file using the R code. The first worksheet in the file provides the pre and post deliberative responses to the respective consideration and preference surveys for each of the cases. This is the raw data needed to calculate DRI. The first set of R code calculates all sets of intersubjective correlations (which are also provided in the second worksheet in the Excel file), and then individual DRI (third worksheet).

The MLM analysis is then performed using the second set of R code using the combined data provided in the Excel file—which includes individual DRI values for all individuals, along with the available demographic data and case study variables.

# information AND OTHER EFFECTS

## Information vs Formal Deliberation

The results in Table D.1.1 report the relative impact of “information” and “formal deliberation” phases for the ten cases where data was obtained mid-process, after much of the information was provided but before formal deliberation commenced—though there was nothing to stop participants talking to each other during the information phase, so some of the change in DRI that we observe for that phase could be due to informal deliberation.

Table D.1.1: Information vs Formal Deliberation Effects

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | DRI Change  (Phase) | |  |
| Case # | Case | Information | Formal Deliberation | n† | |
| 1 | Uppsala Speaks Group Briefing | 0.09\*\* | 0.10\*\* | 22 (23) | |
| 2 | Uppsala Speaks Group Building Plus | 0.12\*\*\* | 0.15\*\*\* | 24 (27) | |
| 6 | Forest ERA Lay Citizens | 0.06 | 0.09 | 11 (12) | |
| 7 | Forest ERA Stakeholders | -0.07 | 0.21\*\*\* | 10 (86) | |
| 10 | WA Biobank Citizens | 0.17 | 0.12\* | 15 (16) | |
| 11 | WA Biobank Stakeholders | -0.10 | 0.04 | 16 (26) | |
| 14 | Energy Futures WA | 0.12\*\* | 0.15\*\*\* | 21 (23) | |
| 15 | Energy Futures NSW | 0.16\* | 0.06 | 14 (18) | |
| 16 | Energy Futures VIC | 0.23\*\*\* | 0.15\*\*\* | 15 (19) | |
| 18 | GBR Futures | 0.08 | 0.04 | 7 (7) | |

Note: Two-tailed Wilcoxon Test: p<0.1\*, p<0.05\*\*, p<0.01\*\*\*; †Number of available data points. Number of participants for each case study (*N*) is shown in parenthesis.

## Group Building and Deliberative Phases

Typically, we see improvements in deliberative reason during both phases, with a slightly greater overall improvement during the formal deliberation phase compared to the information phase. There is considerable variation in these terms among the cases reported in Table D.1.1. The ForestERA Citizens case study and WA Biobank Stakeholders stand out. In both, DRI initially decreases, with a dramatic improvement during formal deliberation in the former case.

It is conceivable that this variation in information versus formal deliberation effects explains the absence of a clear overall “duration” effect in our analysis. Deliberation may still take time (Curato et al. 2017), but the impact of duration depends on deliberative design and the division of time between information and formal deliberation.

## Group Composition

In the main paper we said that any tendency to conformity in a forum would suppress Group DRI. This effect appears in one of our cases. The three Energy Futures cases share the same forum design and topic, except for the NSW case (cases 15), which allocated participants to breakout discussion groups based on similarity of position (i.e., likely conformity). This case yielded roughly half the level of DRI improvement during the formal deliberation phase compared to its counterparts.

## The Role of Expert Presentation

Our observations suggest we do not have to worry about domination by expert framings. Were that to be the case we would expect DRI to change much more strongly during the information phase than is observed, in comparison to change during subsequent formal deliberation.

## Other Design Features

Other design features that could influence level of deliberative reason include diversity of initial opinions, in-person versus online, decision rule (which in most of our cases could be determined by the participants themselves, or left implicit, rather than forced on the forum), and composition (for example, random selection from the public versus stakeholders). Further research could focus on, for example, differences between stakeholders and lay citizens when it comes to the effects of decision rule (e.g. Grönlund et al. 2010).

### Number of Participants

We do not expect overall number of participants in a forum to have an effect on DRI, given that much group reasoning occurs in smaller breakout discussion groups. This was tested, and the results provided in the tables below rather than in the main text, due to limited space. For the analysis we defined Size such that number of participants n<15 is coded as Size=1; 15<n<30, Size =2; and n>30, Size=3. The results are reported in Table D.5.1.1 below. Model 10 shows the results when Size is added to the MLM (the first nine models reported in Table 4 in the main text). As expected, there is no effect. Models 13 and 14 show the effect of adding Size to interaction model, resulting in similar effects to those reported in Table 5 in the main text.

Table D.5.1.1: Multilevel regression results including number of participants (Size)

|  |  |  |  |
| --- | --- | --- | --- |
| Dependent variable: DRIInd Change | | | |
|  | (10) | (I3) | (I4) |
| Deliberation Per Se | 0.116\*\*\*  (0.014) | 0.116\*\*\*  (0.014) | 0.116\*\*\*  (0.014) |
| Individual Level Variables | | | |
| Age | -0.004  (0.004) | -0.004  (0.004) | -0.004  (0.004) |
| Gender | 0.012  (0.014) | 0.012  (0.014) | 0.012  (0.014) |
| Education | -0.004  (0.009) | -0.005  (0.009) | -0.005  (0.009) |
| Case Level Variables | | | |
| Group building | 0.078\*\*  (0.033) | 0.058\*\*  (0.024) | -0.074  (0.066) |
| Complexity | 0.006  (0.054) | 0.431\*\*\*  (0.130) | 0.266\*  (0.134) |
| Duration | -0.072  (0.074) | -0.045  (0.052) | 0.001  (0.048) |
| Decision Impact | -0.001  (0.032) | 0.363\*\*\*  (0.108) | 0.402\*\*\*  (0.093) |
| Size | -0.010  (0.053) | -0.046  (0.039) | -0.033  (0.033) |
| Intercept | 0.316  (0.266) | -0.875\*  (0.389) | -0.720\*  (0.335) |
| Interactions | | | |
| Decision Impact × Complexity |  | -0.127\*\*\*  (0.037) | -0.136\*\*\*  (0.031) |
| Group Building × Complexity |  |  | 0.056\*  (0.027) |
| Observations | 630 | 630 | 630 |
| Individuals | 315 | 315 | 315 |
| Cases | 16 | 16 | 16 |
| Pseudo-R2 (Fixed Effect) | 0.235 | 0.363 | 0.433 |
| Group RE σ | 0.12 | 0.08 | 0.06 |
| Residual RE σ | 0.17 | 0.17 | 0.17 |
| ICC | 0.302 | 0.18 | 0.12 |

Notes: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

### Stakeholder vs Citizen Deliberation

Our analysis includes two cases (7 and 11, see Table D.5.2.1 below) that involve deliberation among specifically recruited stakeholders, rather than randomly selected citizens. Table 3 in the main text shows that these cases produce a lower level of DRI improvement than their randomly selected citizen counterparts (cases 6 and 10). The WA Biobank Stakeholder case (11) resulted in a decline in DRI (of 0.07), compared to a strong improvement (0.28) for its lay citizen counterpart.

This suggests that, as a whole, citizen deliberation tends to produce greater DRI improvement (compared to stakeholders), which also appears to result from more than catching up from a lower starting point. Our tentative explanation is that lay citizens are more likely to activate their deliberative capabilities and engage in greater integration than their more vested counterparts, even if they begin with less knowledge. However, more data is required properly to explore the effect here. We have only two stakeholder processes, not nearly enough to include in the analysis.

We have, however, noted in the main text that the inclusion of the stakeholder processes does not affect our overall results. Table D.5.2.1 and Table D.5.2.2 below reproduce the results reported in Table 4 and Table 5 in the main text, but without including the two stakeholder cases. A comparison between the tables and their counterparts in the main text reveals only a small impact on the results, without any implications for the conclusions that we draw from the analysis.

Table D.5.2.1: Multilevel regression results, without Stakeholder cases

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Dependent variable: *DRIInd* Change | | | | | | | | |
|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Deliberation Per Se |  | 0.122\*\*\*  (0.013) | 0.117\*\*\*  (0.013) | 0.117\*\*\*  (0.013) | 0.117\*\*\*  (0.014) | 0.117\*\*\*  (0.014) | 0.117\*\*\*  (0.014) | 0.117\*\*\*  (0.014) | 0.117\*\*\*  (0.014) |
| Individual Level Variables | | | | | | | | | |
| Age |  |  | -0.005  (0.004) | -0.005  (0.004) | -0.005  (0.004) | -0.005  (0.004) | -0.005  (0.004) | -0.005  (0.004) | -0.005  (0.004) |
| Gender |  |  |  | 0.020  (0.013) | 0.018  (0.014) | 0.019  (0.014) | 0.019  (0.014) | 0.019  (0.014) | 0.019  (0.014) |
| Education |  |  |  |  | -0.005  (0.009) | -0.005  (0.009) | -0.005  (0.009) | -0.005  (0.009) | -0.005  (0.009) |
| Case Level Variables | | | | | | | | | |
| Group building |  |  |  |  |  | 0.090\*\*\*  (0.026) | 0.091\*\*\*  (0.027) | 0.093\*\*\*  (0.027) | 0.091\*\*  (0.030) |
| Complexity |  |  |  |  |  |  | -0.010  (0.036) | 0.012  (0.041) | 0.010  (0.045) |
| Duration |  |  |  |  |  |  |  | -0.071  (0.061) | -0.069  (0.066) |
| Decision Impact |  |  |  |  |  |  |  |  | -0.004  (0.029) |
| Intercept | 0.376\*\*\*  (0.032) | 0.315\*\*\*  (0.032) | 0.338\*\*\*  (0.038) | 0.309\*\*\*  (0.043) | 0.330\*\*\*  (0.050) | 0.023  (0.097) | 0.048  (0.132) | 0.212  (0.192) | 0.228  (0.231) |
| Obs | 718 | 718 | 698 | 698 | 606 | 606 | 606 | 606 | 606 |
| Cases | 17 | 17 | 16 | 16 | 15 | 15 | 15 | 15 | 15 |
| Pseudo-R2 (Fixed Effect) | 0 | 0.076 | 0.070 | 0.072 | 0.070 | 0.292 | 0.292 | 0.294 | 0.288 |
| ΔR2 (fixed) |  | 0.076 | -0.006 | 0.002 | -0.002 | 0.222 | 0.000 | 0.002 | -0.006 |
| Group RE σ | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.10 | 0.10 | 0.10 | 0.11 |
| Residual RE σ | 0.18 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 | 0.17 |

Notes: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

Table D.5.2.2: Interaction effects, without Stakeholder cases

|  |  |  |
| --- | --- | --- |
| Dependent variable: *DRIInd* Change | | |
|  | (I1) | (I2) |
| Deliberation Per Se | 0.117\*\*\*  (0.014) | 0.117\*\*\*  (0.014) |
| Individual Level Variables | | |
| Age | -0.005  (0.004) | -0.005  (0.004) |
| Gender | 0.019  (0.014) | 0.019  (0.014) |
| Education | -0.005  (0.009) | -0.006  (0.009) |
| Case Level Variables | | |
| Duration | -0.04  (0.051) | 0.004  (0.048) |
| Group building | 0.074\*\*  (0.024) | -0.060  (0.070) |
| Complexity | 0.343\*\*  (0.122) | 0.209  (0.122) |
| Decision Impact | 0.292\*\*  (0.106) | 0.354\*\*\*  (0.095) |
| Interactions | | |
| Decision Impact × Complexity | -0.104\*\*  (0.036) | -0.120\*\*\*  (0.032) |
| Group Building × Complexity |  | 0.054\*  (0.027) |
| Intercept | -0.775\*  (0.389) | -0.660\*  (0.335) |
| Obs | 606 | 606 |
| Cases | 15 | 15 |
| Pseudo-R2 (FE) | 0.377 | 0.440 |
| Group RE σ | 0.08 | 0.06 |
| Residual RE σ | 0.17 | 0.17 |

Notes: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01.

# COMPLEXITY Coding

## Coding for Complexity

We conceptualize complexity in terms of the task facing deliberators, rather than in terms of just the issue per se. Task complexity has three dimensions: remit breadth, informational burden, and geographical scale. Each dimension is coded on a 1-4 ordinal scale. The scores on each dimension were then aggregated.

### Remit breadth

1. The task is restricted to acceptance or rejection of a single option.
2. The task involves choice across two or more well-defined options.
3. The participants can devise options of their own in the context of a well-defined problem.
4. The participants can extend the agenda/range of problems to be addressed.

### Informational Burden

1. Process uses only lay/ordinary knowledge, no technical or scientific expertise.
2. Process uses expert knowledge from a single discipline.
3. Process uses expert knowledge from more than one discipline.
4. Process uses expert knowledge that is controversial and contested.

### Geographical scale

Scale here refers to the characteristics of the issue, rather than (for example) the jurisdiction in which participants were recruited, or the level of government receiving recommendations.

1. Local
2. Regional
3. National
4. International

## Complexity Coding Results and Aggregation

Four coders used information available for each of the cases at their respective Participedia sites (see Table A.1.2) producing the scores provided in Table E.2.1, with an inter-coder reliability score (Cronbach’s Alpha) of 0.94. Aggregation was performed using median component score as per Lindstädt et al. (2020).

Table E.2.1: Case Study Complexity Coding

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Study | Remit scope | Informational Burden | Geo scale | Final score |
| 1 | Uppsala Speaks | 2 | 1.5 | 1.5 | 1.5 |
| 2 | FNQCJ | 2 | 3 | 1.5 | 2 |
| 3 | Australian Citizens' Parliament | 4 | 2 | 3 | 3 |
| 4 | Sydney CC Adaption | 3.5 | 4 | 2 | 3.5 |
| 5 | Forest ERA | 2 | 3 | 1 | 2 |
| 6 | Biobanks | 3 | 3 | 2 | 3 |
| 7 | Fremantle Bridge | 2 | 1 | 1 | 1 |
| 8 | CCPS ACT | 3.5 | 4 | 2.5 | 3.5 |
| 9 | CSIRO Energy Futures | 2 | 4 | 3 | 3 |
| 10 | Valsamoggia | 1 | 1 | 1 | 1 |
| 11 | GBR Futures | 3 | 4 | 2.5 | 3 |
| 12 | Human Genome Editing CJ | 3 | 3 | 3 | 3 |

# MLM Analysis Notes

As outlined in the main text, Multi-Level Modelling (MLM) on DRIInd across our 19 cases was performed using two levels:

* the effect of participants’ individual-level characteristics (level 1); and
* the effect of the case level features discussed above (level 2).

The ability of MLM to deal with nested hierarchical data structures (Tasca et al. 2009) is particularly important for our analysis. It corrects for violation of independence (Peugh and Heck 2017), which would otherwise happen given the intersubjective nature of the data. It also accommodates clustering within our dataset, observed via high interclass correlation coefficient (ICC; 0.31) within the null model (i.e., model with no predictors) (Kreft and De Leeuw 1998). The analysis estimates general relationships via fixed effects, and uses random effect estimates of heterogeneity to assess variation across deliberative forums.

## Commensurability across Studies

Structuring the analysis using MLM also overcomes any issues associated with commensurability of DRI scores across cases from different studies, which use different survey instruments. MLM accommodates the fact that DRI measurement is issue and context specific by adopting a model that considers cases as random effect and other predictors as fixed in order to hold transformation effects constant, thus enabling us to determine whether changes to DRIInd can be attributed to deliberation per se, demographic (level 1) or group/case (level 2) variables.

## Robustness and Sample Size

As discussed in the main text, a conservative approach has been adopted for the MLM analysis to ensure robustness of the results given the relatively small sample size, adopting specific measures outlined by Kenward and Roger (1997) to the calculation of degrees of freedom. We also followed Kenward and Roger (1997) in adopting a restricted maximum likelihood estimation model.

## Choice of Model

In deciding which model to use in our analysis we have had to balance competing considerations of theoretical ideal and methodological fidelity. An ideal statistical solution would be to use a random slope for deliberation— longitudinal random intercept and slope model. However, this approach requires more than two points of measurement, beyond the pre-and and post-deliberation (PRE and POST) that we analyze. Although we do have three data points for ten cases, this would dramatically reduce our sample size, below what is required to conduct a reliable analysis.

A compromise to our theoretical ideal in the model that we report in the main text is to adopt a random intercept model, which incorporates the random effects of each deliberative forum and overcomes the variation across pre-deliberative DRI scores. Using this approach, the scores on the dependent variable for each individual observation are predicted by the intercept, which randomly varies across each cluster—i.e. DRI levels are different across the forums.

The effect of the deliberative process across our cases is accounted for by having data points both prior to deliberation (DRI pre) and post deliberation (DRI post)—i.e., we are modelling DRI at two time points, hence the number of observations used in the model is twice the number of individuals. These two points were entered into the model using a dummy variable for which 0 represents the fact that the time is pre deliberation (PRE), and 1 represents the fact that the time is post deliberation (POST).

## Dataset

A conservative approach was adopted to accommodate variation in availability of individual-level data across cases, which meant missing data for some analyses. Missing data is handled via listwise deletion—where cases are missing data used in a particular step they are removed for the remainder of the analysis.[[2]](#footnote-3) The level of model fit is assessed using pseudo-R2 (Nakagawa and Schielzeth 2013).

## Multicollinearity

Multicollinearity between variables has been checked by calculating the Variance Inflation Factors (VIF) across all variables, see Table F.5.1.

Table F.5.1: Multicollinearity Report

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Models | Deliberation  (VIF) | Age  (VIF) | Gender  (VIF) | Education  (VIF) | Group Building  (VIF) | Complexity  (VIF) | Duration  (VIF) | Decision Impact  (VIF) |
| (3) | 1 | 1 |  |  |  |  |  |  |
| (4) | 1 | 1.001 | 1.001 |  |  |  |  |  |
| (5) | 1 | 1.002 | 1.013 | 1.013 |  |  |  |  |
| (6) | 1 | 1.002 | 1.012 | 1.013 | 1 |  |  |  |
| (7) | 1 | 1.003 | 1.012 | 1.013 | 1.028 | 1.028 |  |  |
| (8) | 1 | 1.003 | 1.012 | 1.013 | 1.031 | 1.031 | 1.30 |  |
| (9) | 1 | 1.003 | 1.013 | 1.015 | 1.262 | 1.486 | 1.387 | 1.402 |

In all cases the VIF is less than 2.5, which means the standard error for the coefficient of each variable is less than 1.6 (√2.5) times what its value would be if it had a correlation of 0 with the other variable.

## Sampling Bias Test

We tested for implications of our relatively small sample size using both the Begg Rank Test, and the Egger Regression Test, which are drawn from meta-analytical techniques to check for publication bias. There is some debate in the literature regarding which is the correct test, so we applied both to check whether our observations of DRI are biased—underestimate or overestimate because of sample size—either in terms of number of cases, or number of participants. The Begg ranked test specifically tests for independence of variance and the effect size within our sample to determine whether there is effect size bias considering the number of observations within each case. Similarly, the Egger tests for inverse sample size the strength of our sample to test the relationship between DRI effect size, given the sample size of our cases. The results for both tests reported below in Table F.6.1.

Table F.6.1: Begg Test and Egger Test Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Test | Intercept | z | CI | Kendall τ | P value |
| Begg rank test |  |  |  | 0.24 | 0.16 |
| Egger Regression | 0.383  (0.491) | 1.03 | -0.58, -1.34 |  | 0.30 |

The resulting P values for both tests (both >0.05) indicates that there is no effect size bias across the sample of cases.

# Multi-Level Modelling—Interaction Analysis

The interaction displayed by models 11 and 12 (see table 5, main text) can be visualized in terms of the effect that variation in one variable has on the relationship between the second variable and DRI. Because of limitations with this analysis, using ordinal data, the location of the lines cannot be determined with any precision. The uncertainty regarding the location of these relationships is indicated by the 95% probability contours shown in the figures provided below.

## Decision Impact and Complexity

As reported in the main text, the relationship between complexity and DRI changes as decision impact increases. Figure G.1.1 shows that, where decision impact is low, DRI is lower compared to high decision impact. However, the situation reverses once complexity reaches a level of 3, where the effect of increasing complexity on DRI changes from positive to negative once decision impact reaches level 4. The uncertainty regarding the precise location of these relationships is indicated by the 95% probability contours shown in the figure.

Figure G.1.1: Interaction Between Complexity and Decision Impact



Note: The model from which this interaction is drawn can be found in Table 5 in the main text (model I2).

Table G.1.1 reports the conditional predicted values for the interaction between Decision Impact and complexity that have are incorporated into figure G.1.1.

Table G.1.1: Conditional effects for interaction Decision Impact\*Complexity (Model I2)

|  |  |  |  |
| --- | --- | --- | --- |
| Decision Impact (Level) | Complexity (Level) | Predicted DRI | 95% CI |
| 1 | 1 | -0.22 | [-0.53, 0.10] |
| 1.5 | -0.05 | [-0.30, 0.19] |
| 2 | 0.11 | [-0.06, 0.28] |
| 3 | 0.43 | [0.36, 0.51] |
| 3.5 | 0.59 | [0.49, 0.70] |
| 2 | 1 | 0.03 | [-0.18, 0.24] |
| 1.5 | 0.13 | [-0.03, 0.29] |
| 2 | 0.23 | [0.12, 0.34] |
| 3 | 0,43 | [0.37, 0.48] |
| 3.5 | 0.53 | [0.45, 0.60] |
| 3 | 1 | 0.28 | [0.16, 0.41] |
| 1.5 | 0.32 | [0.23, 0.41] |
| 2 | 0.35 | [0.29, 0.42] |
| 3 | 0.42 | [0.37, 0.48] |
| 3.5 | 0.46 | [0.38, 0.53] |
| 4 | 1 | 0.53 | [0.42, 0.65] |
| 1.5 | 0.50 | [0.42, 0.59] |
| 2 | 0.47 | [0.41, 0.54] |
| 3 | 0.42 | [0.34, 0.49] |
| 3.5 | 0.39 | [0.28, 0.49] |
| 5 | 1 | 0.78 | [0.58, 0.98] |
| 1.5 | 0.69 | [0.54, 0.84] |
| 2 | 0.60 | [0.48, 0.71] |
| 3 | 0.41 | [0.30, 0.52] |
| 3.5 | 0.32 | [0.17, 0.46] |

Table G.1.2 reports the mean predicted values for the interaction between Decision Impact and Complexity that have been considered to build Figure G.1.1.

Table G.1.2: Mean conditional effect interaction Decision Impact\*Complexity (Model I2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Complexity\*Decision Impact | | | | | |
| Statistic | N | Mean | St. Dev. | Min | Max |
| Complexity | 25 | 2.200 | 0.946 | 1.000 | 3.500 |
| Predicted DRI | 25 | 0.366 | 0.230 | -0.215 | 0.783 |
| Standard Error | 25 | 0.063 | 0.033 | 0.027 | 0.162 |
|  | Confidence Interval | | | | |
| Lower | 25 | 0.242 | 0.274 | -0.532 | 0.584 |
| Upper | 25 | 0.490 | 0.197 | 0.101 | 0.982 |

## Group Building and Complexity

Figure G.2.1 illustrates the interaction between Complexity and Group Building. As reported in the main text, as complexity increases, so too does the relative impact of group building in producing higher DRI. The positive role of group building increases once complexity reaches around 1.5 (hence the negative, but not significant, result for the group building coefficient in model I2 in Table 5 in the main text).

Figure G.2.1: Interaction Between Complexity and Group Building (model I3)



Note: The model from which this interaction is drawn can be found in Table 5 in the main text (model I2)

Table G.2.1 reports the conditional predicted values for the interaction between Group Building and Complexity that have been considered to build Figure G.1.1.

Table G.2.1: Conditional effects for interaction Group Building\*Complexity (Model I2)

|  |  |  |  |
| --- | --- | --- | --- |
| Group Building (Level) | Complexity (Level) | Predicted DRI | 95% CI |
| 1 | 1 | 0.59 | [0.36, 0.83] |
| 1.5 | 0.7 | [0.30, 0.65] |
| 2 | 0.35 | [0.24, 0.47] |
| 3 | 0.11 | [0.00, 0.22] |
| 3.5 | -0.01 | [-0.17, 0.16] |
| 2 | 1 | 0.57 | [0.40, 0.74] |
| 1.5 | 0.48 | [0.36, 0.61] |
| 2 | 0.39 | [0.31, 0.48] |
| 3 | 0.21 | [0.14, 0.29] |
| 3.5 | 0.12 | [0.01, 0.24] |
| 3 | 1 | 0.55 | [0.43, 0.68] |
| 1.5 | 0.49 | [0.40, 0.58] |
| 2 | 0.43 | [0.37, 0.49] |
| 3 | 0.32 | [0.25, 0.38] |
| 3.5 | 0.26 | [0.17, 0.34] |
| 4 | 1 | 0.53 | [0.42, 0.65] |
| 1.5 | 0.50 | [0.42, 0.59] |
| 2 | 0.47 | [0.41, 0.54] |
| 3 | 0.42 | [0.34, 0.49] |
| 3.5 | 0.39 | [0.28, 0.49] |
| 5 | 1 | 0.51 | [0.36, 0.67] |
| 1.5 | 0.51 | [0.40, 0.63] |
| 2 | 0.52 | [0.42, 0.61] |
| 3 | 0.52 | [0.40, 0.63] |
| 3.5 | 0.52 | [0.37, 0.67] |

Table G.2.2 reports the mean predicted values for the interaction between Group Building and Complexity that have been considered to build Figure G.2.1.

Table G.2.2: Mean conditional effect interaction Group Building\*Complexity (Model I2)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Complexity\*Group Building | | | | | | |
|  | N | Mean | St. Dev. | Min | Max |
| Complexity | 25 | 2.200 | 0.946 | 1.000 | 3.500 |
| Predicted DRI | 25 | 0.410 | 0.158 | -0.007 | 0.594 |
| Std Error | 25 | 0.059 | 0.022 | 0.031 | 0.121 |
|  |  | Confidence Interval | | | |
| Lower | 25 | 0.295 | 0.156 | -0.175 | 0.431 |
| Upper | 25 | 0.526 | 0.171 | 0.161 | 0.832 |

The progressive effect of the different increments of group building in overcoming how complexity impedes DRI appears relatively smooth—i.e. there are no threshold effects and incremental improvements in group building lead to improvement in mitigating the effect of complexity.

# References

Ashworth, Peta, Anna Littleboy, Paul Graham, and Simon Niemeyer. 2010. “Turning the heat on: Public engagement in Australia’s energy future.” In Renewable Energy and the Public: From NIMBY to Participation, edited by Patrick Devine-Wright. London: Earthscan.

Brown, Steven R. 1970. “On the use of variance designs in Q methodology.” *Psychological Record* 20: 179–189.

Carson, Lyn, John Gastil, Janette HartzKarp, and Ron Lubensky, eds. 2013. The Australian Citizens' Parliament and the Future of Deliberative Democracy. University Park, Penn.: Pennsylvania State University Press.

Curato, Nicole, John S. Dryzek, Selen Ercan, Carolyn M. Hendriks, and Simon John Niemeyer. 2017. “Twelve Key Findings in Deliberative Democracy Research.” *Daedalus* 146 (3): 28–38.

Felicetti, Andrea, Simon Niemeyer, and Nicole Curato. 2016. “Improving deliberative participation: connecting mini-publics to deliberative systems.” *European Political Science Review* 8 (3): 427-448.

Grönlund, Kimmo, Maija Setälä, and Kaisa Herne. 2010. “Deliberation and civic virtue: lessons from a citizen deliberation experiment.” *European Political Science Review* 2 (1): 95–117.

Hartz-Karp, Janette, Patrick Anderson, John Gastil, and Andrea Felicetti. 2010, “The Australian Citizens' Parliament: forging shared identity through public deliberation.”, *Journal of Public Affairs* 10 (4): 353-371.

Hobson, Kersty Pamela. 2009. “On the modern and nonmodern in deliberative environmental democracy.” Global Environmental Politics 9 (4): 64-80.

Hobson, Kersty Pamela, and Simon John Niemeyer. 2011. “Public responses to climate change: The role of deliberation in building capacity for adaptive action.” *Global Environmental Change* 21 (3): 957–971.

Jennstål, Julia. 2018. “Deliberative Participation and Personality: The Effect of Traits, Situations and Motivation.” *European Political Science Review* 10 (3): 417-440. [4](about:blank).

---. 2019. “Deliberation and Integrative Complexity: Assessing the Development of Deliberative Norms in Minipublics.” *Swiss Political Science Review* 25 (1): 64–83.

Kenward, M. G., and J. H. Roger. 1997. “Small sample inference for fixed effects from restricted maximum likelihood.” *Biometrics* 53 (3): 983-997.

Lehman, Ann, Norm O'Rourke, Larry Hatcher, Edward J. Stepanski, and SAS Institute. 2013. JMP for basic univariate and multivariate statistics: methods for researchers and social scientists. Second edition. ed.

Lindstädt, R., S. O. Proksch, and J. B. Slapin. 2020. “When experts disagree: Response aggregation and its consequences in expert surveys.” *Political Science Research and Methods* 8 (3): 580-588.

Meade, Adam W. 2010. “Restriction of range.” Encyclopedia of Research Design, edited by Neil J. Salkind, 1279-1280. London: SAGE.

Nakagawa, Shinichi, and Holger Schielzeth. 2013. “A General and Simple Method for obtaining R2 from generalized linear mixed-effects models.” Methods In Ecology and Evolution (4): 133-142.

Niemeyer, Simon John. 2004. “Deliberation in the wilderness: displacing symbolic politics.” *Environmental Politics* 13 (2): 347–372.

Niemeyer, Simon John, Selen Ayirtman, and Janette HartzKarp. 2013. “Q methodology and Analysis of Policy Issues: The Fremantle Bridge Case Study.” *Operant Subjectivity* 36 (2): 114–134.

Schlosberg, David., Lisette Collins, and Simon John Niemeyer. 2017. “Adaptation Policy and Community Discourse: Risk, Vulnerability, and Just Transformation.” *Environmental Politics* 26 (3): 413–437.

Secko, D. M., N. Preto, S. Niemeyer, and M. M. Burgess. 2009. “Informed consent in biobank research: A deliberative approach to the debate.” *Social Science & Medicine* 68 (4): 781-789.

Stephenson, William. 1986. “The concourse theory of communication.” *Operant Subjectivity* 9 (2): 37–58.

1. To illustrate further, if we take two hypothetical individuals X and Y who are in perfectly inconsistent disagreement (e.g. plotting at 1,0 in Figure C.1.1), such that *dx,y*= 0.71, and we then add an additional individual (Z) who is in perfectly inconsistent disagreement with X (*dx,z*= 0.7), they would, by definition, be in perfectly consistent disagreement with Y (*dy,z*= 0), such that the maximum average distance (λ) declines from 0.71 to 0.47, and so on until λ→. [↑](#footnote-ref-2)
2. Alternatively, it would be possible to employ imputation techniques, for mean imputation, regression imputation or stochastic regression imputation. We decided not to adopt imputation to avoid uncertainty associated with too small a standard error. [↑](#footnote-ref-3)