Online Appendix

When should the majority rule? Experimental evidence for Madisonian
judgments in five cultures

**Table of Contents**

[Appendix A – Sample characteristics and country differences 2](#_Toc33379194)

[Appendix B – Experimental materials 5](#_Toc33379195)

[Dinner Scenario 5](#_Toc33379196)

[Activity Scenario 5](#_Toc33379197)

[Company Scenario 5](#_Toc33379198)

[On the selection of four decision rules 6](#_Toc33379199)

[Appendix C – Additional analyses of participants’ choice of decision rule 7](#_Toc33379200)

[Preference for voting and levels of democracy 7](#_Toc33379201)

[Choice of decision rule broken down by scenario 8](#_Toc33379202)

[Multilevel multinomial logistic regression for choice of decision rule 8](#_Toc33379203)

[Appendix D – Additional analyses of appropriateness ratings 10](#_Toc33379204)

[Appropriateness of decision rule broken down by scenario 11](#_Toc33379205)

[Multilevel linear regression models for the appropriateness of voting and consensus 13](#_Toc33379206)

[Appendix E – Main analysis without excluding comprehension failures 14](#_Toc33379207)

# Appendix A – Sample characteristics and country differences

Table A1 shows the demographics of the five samples. All samples are relatively young and slightly left leaning (on a 1-7 scale). Overall, the samples provide considerable variation in national, political, and cultural contexts, hence they allow us to test whether groups of people show similar Madisonian intuitions despite radically different social, cultural, and political backgrounds. Admittedly, these samples are not representative of the national populations, hence we cannot (and do not) make any statements about the relationship between sample estimates and the latent population parameters.

Table A1. Sample characteristics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Country | Sampling | *n* | Female | Age (*SD*) | Ideology (*SD*) |
| Denmark[[1]](#footnote-1) | University students | 139 | 55% | 20.7 | (1.3) | 3.76 | (1.27) |
| Hungary | University students | 113 | 48% | 21.7 | (1.9) | 4.04 | (1.06) |
| India | Mechanical Turk | 115 | 32% | 33.1 | (9.1) | 3.57 | (1.78) |
| Russia | University students | 209 | 72% | 22.1 | (4.6) | 3.31 | (1.30) |
| USA | Mechanical Turk | 114 | 40% | 34.1 | (10.5) | 3.23 | (1.85) |

 Table A2 reports detailed information on sampling methods. All data was collected between May and October of 2017. Attrition rates (proportion of participants who started but have not finished the study) range from 13% in the US to 52% in Russia. Bounce rates (participant getting invited but deciding not to start the survey) are between 10-15% for the two MTurk samples. The same rate in Denmark is virtually 0% as the survey was embedded in a mandatory in-class survey for recently enrolled students, who have been warned in advance to bring their own computer for the occasion. Unfortunately, we do not have good estimates on the number of participants invited to complete the study in Russia and in Hungary, but admittedly the bounce rates were much higher than in the other three countries. This can be partially explained by the recruitment and reward method and in the case of our Russian sample, by the topic of our study that could have been perceived as politically sensitive. Several approached lecturers were suspicious of the goal of the survey and refused to help with data collection. Participants might have dropped out form the study for the same reason.

Table A2. Sampling methods

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Final N* | *N recruited* | *Inclusion criteria* | *Compensation scheme* |
| Denmark | 139 | 299 | Enrolled in "Methods 1" 1st year Political Science BA course at a large Danish public university | Course credit |
| India  | 113 | 211 | Amazon's Mechanical Turk users with following attributes:Country: India, HITs approved: 1000+, Approval rate: 95%+ | $1 upon completion |
| Hungary | 115 | 151 | Enrolment in one of the three largest public universities in Budapest. Participants were recruited either by their instructors in class or via email or by getting approached on campus by one of the authors. | Enrolment in a lottery for 10,000HUF (~$34) |
| Russia  | 209 | 438 | Participants were recruited with the help of a network of lecturers at various universities across Russia. Most of participants learnt about the survey from their lecturers in class and some from social networks (via friends and online). The majority of our participants live in Irkutsk and Krasnoyarsk. Other locations include Moscow, Yekaterinburg, and Tomsk. | Enrolment in a lottery for 4000RUB (~$60) |
| United States  | 114 | 132 | Amazon's Mechanical Turk users with following attributes:Country: India, HITs approved: 1000+, Approval rate: 95%+ | $0.8-1 upon completion |

Table A3 below reports power analyses based on the estimates reported in DeScioli and Bokemper’s (2018) Study 3. In particular, first we calculate the sample sizes required to find a statistically significant (p < 0.05) effect for the vulnerable minority manipulation (versus the control condition) with 80% probability. We base our effect size estimates on the chi-square tests for the changes in preferred decision rule and the t-tests for the appropriateness of voting. Following the logic of the published estimates, we calculate estimates separately for each scenario. The sample size estimates in Table A3 demonstrate that samples between 32 and 187 are required to achieve 80% power for the change in the preferred decision rule, whereas the same power could be achieved with as few as 29 participants for the vote appropriateness.

We chose to collect samples of 120 in each country (and not 187 participants, which would have been the most conservative decision) for two reasons. First, by having sufficient power (by large margins) to test five of the six statistical tests, it appeared to be an optimal way to spend our resources. Second, being primarily interested in the overall treatment effect, we knew that we will be able to increase the precision of our estimates (compared to DeScioli and Bokemper 2018) by a) averaging over scenarios as in the main text, and b) by relying on partial pooling through multilevel regression modelling as in the robustness checks reported below. Therefore, we believe our paper has a very low risk of having noisy, unreliable, or false positive estimates.

|  |
| --- |
| Table A3. Power analyses |
|   |   | **Dinner** | **Day trip** | **Company** |
| **Sample size for power = 0.8** | Choice of decision rule | 32 | 105 | 187 |
|  | Vote appropriateness | 9 | 29 | 29 |
|  |  |  |  |  |

 Table A4 summarizes key political and social differences between the five countries based on four indicators. The democracy score reflects the democratic institutions in society. The economic development of a country may affect the vulnerability of different groups in society. Ethnic fractionalization indicates the potential for vulnerable minorities based on ethnicity. The number of political parties indicates the potential for political groups that could become marginalized, vulnerable minorities.

Overall, the table shows considerable variation. Politically, four countries are democratic, while Russia is a non-democratic regime. Furthermore, all countries except Denmark have witnessed the rise of charismatic, populist leaders in the years preceding data collection (PM Orbán in Hungary, PM Modi in India, President Putin in Russia and President Trump in the USA). Economically, Denmark and the USA are rich countries, whereas Hungary and Russia are relatively poorer, and India is considerably poorer. Ethnically, Denmark and Hungary are homogeneous, whereas India and the USA are heterogeneous, with Russia in the middle. In terms of political parties, there are large differences in party systems ranging from the House of Representatives in the US, which has a two-party system to India with a multiparty system.

|  |
| --- |
| Table A4. Political and social differences between samples |
|   | Democracy score (2017)**a** | GDP PPP (2016)**b** | Ethnic fractionalization (2003)**c** | Effective number of parties at the electoral level (year)**d** |
| Denmark | 97 |  $49,496  | 0.08 | 5.86 (2015) |
| Hungary | 76 |  $26,681  | 0.15 | 3.22 (2014) |
| India | 77 |  $6,572  | 0.42 | 6.82 (2014) |
| Russia | 20 |  $23,163  | 0.25 | 3.14 (2016) |
| USA | 89 |  $57,467  | 0.49 | 2.14 (2016) |

1. Freedom House’s Freedom in the World 2017 report. Aggregate scores reflecting both political rights and civil liberties: 0 = least free, 100 = most free. <https://freedomhouse.org/report/fiw-2017-table-country-scores> accessed [2018.07.28.]
2. World Bank’s estimates of per capita Gross Domestic Product at purchasing power parity [https://en.wikipedia.org/wiki/List\_of\_countries\_by\_GDP\_(PPP)\_per\_capita](https://en.wikipedia.org/wiki/List_of_countries_by_GDP_%28PPP%29_per_capita) accessed [2018.07.28.]
3. Ethnic fractionalization index developed by Alberto Alesina; et al. (2003). "Fractionalization". Journal of Economic Growth. 8: 155–194. doi:10.1023/a:1024471506938. The numbers reflect the probability that two randomly drawn individuals from a country are not from the same group. Data retrieved from [https://en.wikipedia.org/wiki/List\_of\_countries\_ranked\_by\_ethnic\_and\_cultural
\_diversity\_level](https://en.wikipedia.org/wiki/List_of_countries_ranked_by_ethnic_and_cultural_diversity_level) accessed [2018.07.28.]
4. Gallagher, Michael, 2018. Election indices dataset at [http://www.tcd.ie/Political\_Science/
people/michael\_gallagher/ElSystems/index.php](http://www.tcd.ie/Political_Science/people/michael_gallagher/ElSystems/index.php) accessed [2018.07.28.]

# Appendix B – Experimental materials

The treatment differences are shown with colored font, blue for the control condition and red in brackets for the vulnerable minority condition.

## Dinner Scenario

A group of ten people are deciding where to have a dinner event. Some people prefer to have sushi at the event and others prefer to have the event at an Italian restaurant instead. [Seven people want to have the event at a Japanese sushi restaurant. Three people cannot eat sushi because they have fish allergies and they want to have the event at an Italian restaurant instead.] They have discussed this issue for a while but haven’t come to a conclusion. How should the group decide what to do?

## Activity Scenario

A group of ten people rented a boat and they are deciding where to go for a day trip. Some people prefer to go to the beach, while other people prefer to go to the waterfall. [Seven people want to go down the river to a beach. Three people do not like the beach because they sunburn very easily, and they want to go up the river to a waterfall instead.] They have discussed the issue for a while but haven’t come to a conclusion. How should the group decide what to do?

## Company Scenario

A group of ten people are selling their software company and deciding how to divide the profits. All ten people contributed equal investments to start the company, but some people worked more hours than other people to grow the company. Some people think that they should divide the profits equally, and some people think they should divide the profits based on how much work each person contributed. [three of the people did all of the work creating and selling the software. The seven who invested without working think the profits should be divided equally. The three who did the work think they should receive a larger share of the profits.] They have discussed the issue for a while but haven’t come to a conclusion. How should the group decide what to do?

**Dependent measures**

1. In your opinion, which of the following four decision rules is best for this situation?

* **A leader should decide.** One person should take a leadership role and make the decision for the group.
* **The group should debate until a consensus is reached.** All group members should debate the options until everyone agrees on which option to choose.
* **The group should vote.** The group should vote on the options and choose the option that receives the most votes.
* **The group should use a chance process.** The group should use a random process such as flipping a coin, rolling dice, drawing straws, or picking names out of a hat.

2. For the next four items, rate the appropriateness of each decision rule for solving the group’s problem. (1 = very inappropriate; 7 = very appropriate)

* A leader should decide
* The group should discuss until consensus is reached
* The group should vote
* The group should use chance

**Comprehension Question**

Which of the following occurred in at least one of the scenarios that you read?

* A baby was crying for help
* A car was speeding out of control
* A couple went on a romantic date
* All of the Above
* None of the Above

## On the selection of four decision rules

Previous research found that in free responses, participants spontaneously proposed voting, consensus, leadership, and chance as procedures for making a variety of group decisions; no other decision rule was mentioned by a substantial portion of participants (>5%; DeScioli and Bokemper, in press). Although our primary focus is on majority-rule voting, and to a lesser extent consensus, we included all four decision rules to ensure that preferences for voting or consensus were not due to limited response options.

# Appendix C – Additional analyses of participants’ choice of decision rule

## Preference for voting and levels of democracy

Figure C1 shows participants’ choice of voting (%, y axis) in each country by the level of democracy in the country (x axis), based on the Freedom House’s 2017 democracy scores. To focus on general preferences, the figure includes only the control condition in which the scenarios did not have a vulnerable minority. Although a small number of data points warrants caution, they appear to be correlated, *r*(3) = 0.83, *p* = 0.08.



Figure C1. Participants’ choice of voting (%) by their country’s level of democracy.

## Choice of decision rule broken down by scenario

Figure C2 shows participants choices broken down by scenario.

**

Figure C2. Participants’ choice of decision rule by country and scenario.

## Multilevel multinomial logistic regression for choice of decision rule

We also analysed participants’ choices with multilevel multinomial logistic regression. Multinomial logistic regression models are an extension of simple logistic regressions allowing to regress categorical (as opposed to binary) dependent variables on multiple independent variables. They are analogous to conducting a series of logistic regressions picking one level as the reference category and comparing its likelihood to all other levels of the dependent variable. Meanwhile, multilevel or hierarchical models extend simple regressions by assuming that regression coefficients themselves could be modelled as a function of higher order variables. This helps to relax standard regression assumptions about the independence of error terms, which are violated by within-participant manipulations and clustering by countries.

 Accordingly, our multilevel multinomial logistic regression model estimates the overall effect of the vulnerable minority treatment on the probability of choosing each decision rule while accounting for the fact that participants completed multiple scenarios and were clustered by country. The model includes varying intercepts for individuals and both varying intercepts and slopes for the 15 country × scenario combinations. In other words, the model assumes that the baseline preferences for the four decision rules vary across participants, and that both the baseline preferences and the magnitude of the minority effect vary across scenarios and countries.

Because we are most interested in voting, we used it as the reference category in the analysis so that each effect is interpretable as the difference between a given rule (consensus, leadership, or chance) and voting. Thereby, the model estimates the baseline preference for voting and the minority effect on the preference for voting in comparison with each of the other decision rules. We used the *brms* package in R to run the analysis, and the code is in the paper’s OSF repository (https://osf.io/4tp9b).

Table C1 shows the main results. All estimates are standard log odds with negative estimates corresponding to events less likely than the reference group or category and positive estimates to events more likely. Specifically, the intercepts estimate the likelihood of each choice compared to voting in the control condition. The negative coefficients (and CIs) indicate that consensus, leadership and chance are all chosen significantly less often than voting in the control condition without a vulnerable minority.

The vulnerable minority effects estimate the change in the likelihood of choosing each option relative to voting when the group included a vulnerable minority. The positive coefficients indicate that participants were more likely to choose other options besides voting when there was a vulnerable minority. Finally, we note that although the minority effects for consensus and leadership were similar in size, participants were most likely to choose consensus in absolute terms, since the baseline preference for consensus was greater (i.e., the added intercept and minority effect is greatest for consensus). In short, this multilevel multinomial logistic regression model reaffirms the findings from the main text.

|  |
| --- |
| **Table C1. Multilevel multinomial regression model of choice of decision rule** |
|   | Consensus vs. Vote | Leader vs. Vote | Chance vs. Vote |
| Intercept | -0.92 | -2.24 | -2.74 |
|   | (-1.58, -0.33) | (-3.12, -1.42) | (-3.58, -2.02) |
| Vulnerable minority | 1.73 | 1.62 | 0.85 |
|   | (1.13, 2.35) | (0.91, 2.36) | (0.34, 1.39) |
| Note: N = 1931. The model includes random intercepts for participants and random intercepts and slopes for scenario and country. The 95% confidence intervals are in parentheses.  |

# Appendix D – Additional analyses of appropriateness ratings

We first examine participants’ ratings of appropriateness for each decision rule separately (see Figure D1). The appropriateness ratings provide more detail about participants’ judgments of each rule, rather than only their relative preferences. We combine the three scenarios by averaging participants’ ratings of voting across them, separately for the vulnerable minority condition and the control condition.

Generally, participants’ support for voting varied by country as expected, *F*(4, 973) = 6.14, *p* < .001. More to the point, participants rated voting as less appropriate when there was a vulnerable minority (*M* = 0.11) than in the control condition (*M*= 1.50), *t*(1794) = 17.0, *p* < .001. Indeed, we found a substantial and significant decrease in the appropriateness of voting in all five countries: Denmark, *ΔM* = -1.59, *t*(98)= 6.37, *p* < .001; Hungary, *ΔM* = -1.54, *t*(86)= 6.48, *p* < .001; India, *ΔM* = -0.97, *t*(113)= 4.27, *p* < .001; Russia, *ΔM* = -1.48, *t*(201)= 8.68, *p* < .001; and USA, *ΔM* = -1.42, *t*(93)= 6.09, *p* < .001.

*Figure D1.* Participants’ appropriateness ratings of voting and consensus by condition (control or vulnerable minority) and country. The error bars are standard errors.

In tandem, participants rated consensus to be more appropriate when there was a vulnerable minority (*M* = 1.35) compared to the control condition (*M* = 0.63), *t*(1924) = 9.3, p < .001. These differences also occurred across all five countries: Denmark, *ΔM* = 0.78, *t*(123)= 2.98, *p* < .01; Hungary, *ΔM* = 1.12, *t*(108)= 4.10, *p* < .001; India, *ΔM* =0.73, *t*(93)=3.10, *p* < .01; Russia, *ΔM* =0.71, *t*(205)= 4.27, *p* < .001; and USA, *ΔM* = 0.21, *t*(110)= 0.94, *p = 0.35*.

## Appropriateness of decision rule broken down by scenario

We also present participants’ appropriateness ratings broken down by scenario; Figure D2 shows the appropriateness of voting and consensus, and Figure D3 shows leadership and chance.

*Figure D2. Participants’ ratings of appropriateness for voting and consensus. Participants generally rated voting (black) as less appropriate when there was a vulnerable minority compared to the control condition, and they rated consensus (grey) as more appropriate when there was a vulnerable minority. Error bars are standard errors.*



*Figure D3. Participants’ ratings of appropriateness for leader and chance. Error bars are standard errors.*

## Multilevel linear regression models for the appropriateness of voting and consensus

We also analyzed appropriateness ratings with multilevel linear regression models. We conducted a multilevel regression model of the appropriateness of voting with a predictor for whether there was a vulnerable minority, while allowing for varying intercepts for individuals as well as varying slopes and intercepts for country × scenario combinations (as before with participants’ choices). Table D1, Model 1 shows the results. As in our analysis of choices, participants rated voting as highly appropriate in the control condition without a vulnerable minority. The negative effect of the vulnerable minority shows that participants rated voting as less appropriate when there was a vulnerable minority.

 We conducted the same regression analysis for consensus ratings (Table D1, Model 2). As before, the vulnerable minority increased the appropriateness of consensus.

|  |
| --- |
| **Table D1. Multilevel models for the appropriateness of voting and consensus** |
|  |
|  | Vote | Consensus |
|  | (1) | (2) |
|  |
| Intercept | 1.548\*\*\* | 0.625\*\*\* |
|  | (1.244, 1.853) | (0.372, 0.877) |
|  |  |  |
| Vulnerable minority | -1.385\*\*\* | 0.712\*\*\* |
|  | (-1.748, -1.022) | (0.439, 0.985) |
|  |  |  |
|  |
| Observations | 1,931 | 1,931 |
| Log Likelihood | -3,726.380 | -3,638.012 |
| *Note:* \*\*\*p<0.01, The model includes random intercepts for participants and random intercepts and slopes for scenario and country. The 95% confidence intervals are in parentheses. |

# Appendix E – Main analysis without excluding comprehension failures

As planned in advance, in the main analysis we excluded participants who failed the comprehension check, which was a simple question about the scenarios they read (see Appendix B). A relatively high rate of failure (15%) probably reflects low motivation to participate in the study (particularly in Russia and Hungary) or linguistic difficulties (in India).

Figure E1 shows the main results without excluding any participants. The overall treatment effects are unaffected by exclusions: pooling across countries and scenarios, participants chose voting significantly less often in the vulnerable minority condition (*M* = 23%) than the control condition (*M* = 51%), *t*(775) = 13.8, *p* < .001. Further, more participants in the vulnerable minority condition (*M* = 48%) selected consensus compared to the control condition (*M* = 26%), *t(787*) = 10.5, *p* < .001. Overall, these results indicate that shortcomings in sample quality may have introduced noise to our estimates, but are unlikely to bias our findings.



Figure E1. Participants’ choice of decision rule in each country without excluding comprehension failures

1. The Danish sample differs from the other four samples in that participants were randomly presented with two of the three scenarios due to time constraints. This feature of the data affects none of the analyses reported in the paper or this appendix. [↑](#footnote-ref-1)