

Do Grades have Absolute Meaning ?

An Experiment on Majority Judgement



Théo Delemazure **ILLC, University of Amsterdam**

Antoinette Baujard **GATE Lyon Saint-Etienne, Université Jean Monnet**

Roberto Brunetti **LEMMA, Université Paris Panthéon Assas**

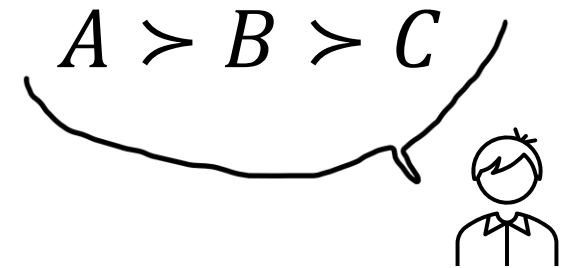
Sylvain Bouveret **LIG, Université Grenoble Alpes**

Measuring things

Physical phenomena can be measured using specific units. These measures *are universally consistent*.



Preferences are *not straightforwardly measurable* and comparable across individuals. The use of rankings in common voting rules could distort the measurement of individual preferences.



Grades

Grades are a common device for assessments used in education, performance reviews, or elections.

Could they be used as a universal language for measuring preferences?

- Excellent
- Very good
- Good
- Fairly good
- Acceptable
- Insufficient
- To reject

Majority Judgement (Balinski & Laraki, 2007)

Under **Majority Judgement (MJ)**, voters assess every candidate using verbal grades.

The candidate with the **best median grade** wins the election.

Pour cette élection,
je juge que les candidats suivants sont:

	A REJETER	INSUFFISANT	PASSABLE	ASSEZ BIEN	BIEN	TRÈS BIEN	EXCELLENT
☐							
○							
△							

Marjolaine Leray (2017)

Motivated by the absolute meaning of grades:

“scales or measures constitute common languages of words that have absolute meanings, clearly understood by those who use them.”

(Balinski & Laraki, 2011)

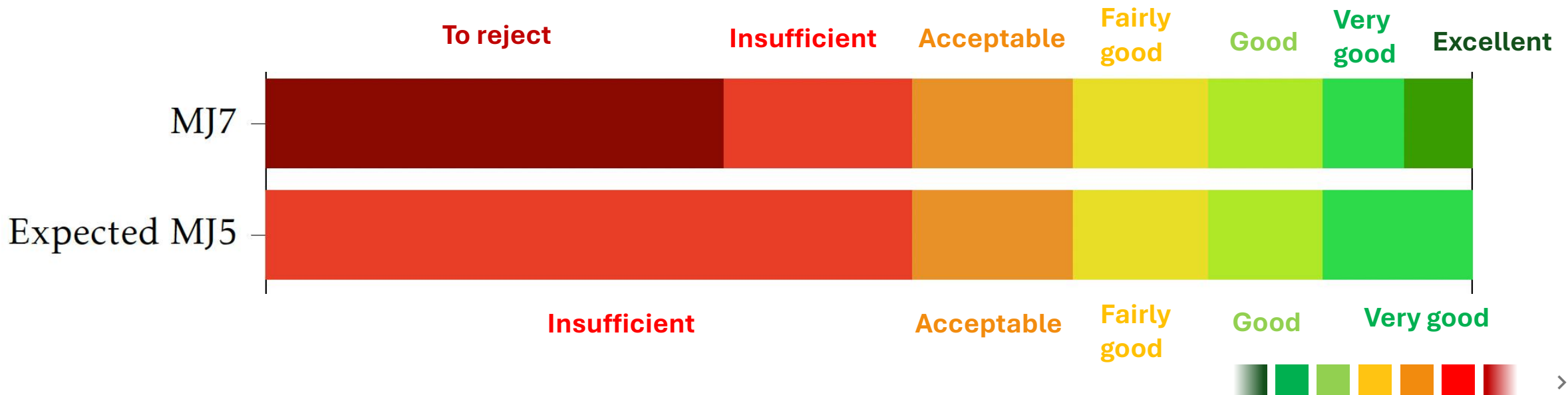
Assumption of absolute meaning

Informally, we would like to know whether a given grade (e.g. “Acceptable”)

- (1) have the same meaning for everyone (*inter-personal*), and
- (2) for any individual, it always have the same meaning** (*intra-personal*).

Assumption of absolute meaning

Consider a given grading scale where both the lowest and highest grades are removed. After this reduction, a voter shifts only the former lowest and highest grades to the new extremes, **leaving the rest of the grade distribution unchanged.**



Related works

- 1 >> **Social choice literature on the properties of alternative voting rules.**
- 1a > Theoretical properties of different rules (Sen 1995, Brams & Fishburn 2002...), and MJ in particular (Balinski & Laraki 2007,2011,2014,2020).
- 1b > Challenge of MJ's properties and assumptions (Felsenthal & Machover 2008, Brams 2011, Laslier 2019, Fabre 2021).
- 1c > Empirical testing of MJ are only few (Balinski & Laraki 2010, Baujard & al., 2024).

In this paper, we run a large online experiment to study how votes cast under MJ depend on grade scales, in order to **confirm or reject the claim that grades have absolute meaning.**

Related works

- 2 >> **Observed individual preferences can vary with the used voting rule.**
- 2a > **Strategic voting** is unavoidable for preference orderings, hence expressed individual choices differ from genuine preferences in strategic context. (Gibbard 1973 ; Satterthwaite 1975). Even MJ is manipulable (Laslier 2019, Baujard et al. 2024).
- 2b > **Framing effects** due to scales: Names of the labels or scales matter, e.g. in political or happiness surveys (Schwarz et al 1991, 2012 ; Fleurbaey & Blanchet 2013), or in evaluative voting rules (Baujard et al. 2018, 2021 ; Darmann et al. 2019).

In this paper, the hypothetical nature of the experimental vote does not incentivize strategic behaviors. Hence, **we focus on the framing effect** due to varying scales, abstracting from the strategic voting confounder.



Related works

3 >> Literature contrasting **epistemic vs. electoral democracy**.

Among the distinct views of democracy (e.g. Elster 1986, Girard 2014, Landemore 2020), **electoral democracy** consists in aggregating individual subjective preferences, while **epistemic democracy** aims at finding the best procedure to uncover an absolute assessment of options, a ground truth (e.g., in a jury).

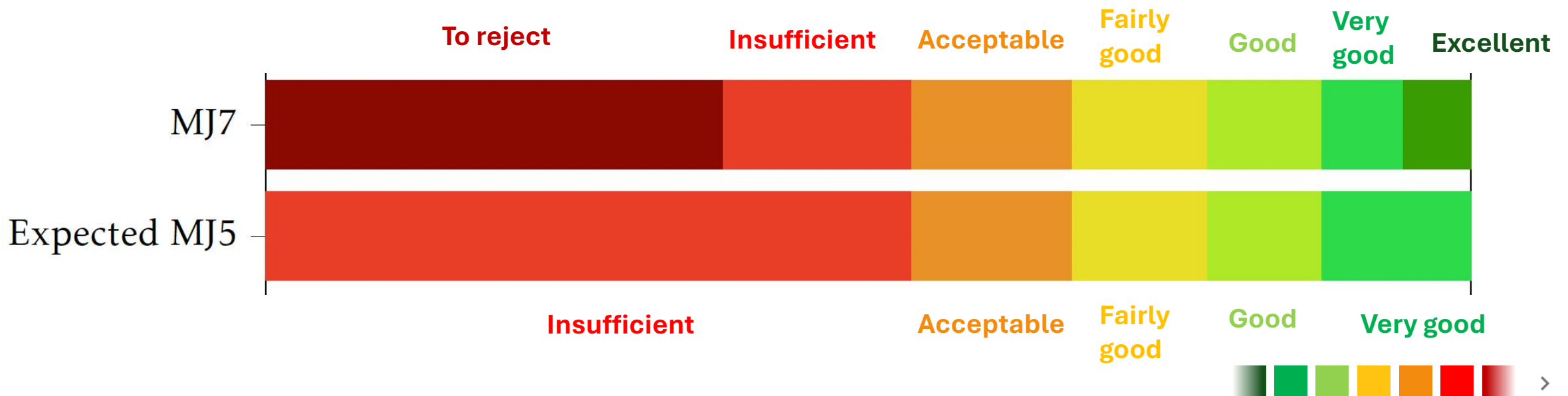
Arguments for MJ are based on the epistemic approach : **voting inputs reveal absolute assessments of candidates**

In this paper, we question whether the MJ procedure is more adapted to epistemic or to electoral democracy.

› **Experimental Design**

Assumption of absolute meaning

Consider a given grading scale where both the lowest and highest grades are removed. After this reduction, a voter shifts only the former lowest and highest grades to the new extremes, **leaving the rest of the grade distribution unchanged.**



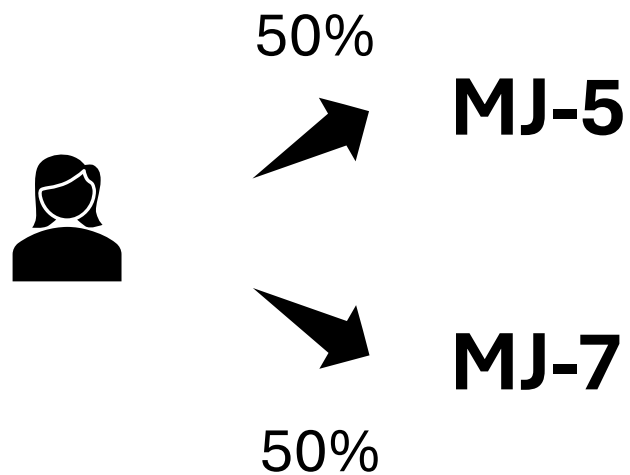
Within or between subject

Within subject experiment



Could lead to carry-on effects

Between subject experiment



The context: “Un Autre Vote” (2022 presidential election)



Online experiment at vote.imag.fr

Participants could simulate a vote for one president among the 12 different official candidates, testing different voting rules.



April 8th 2022 – May 7th 2022

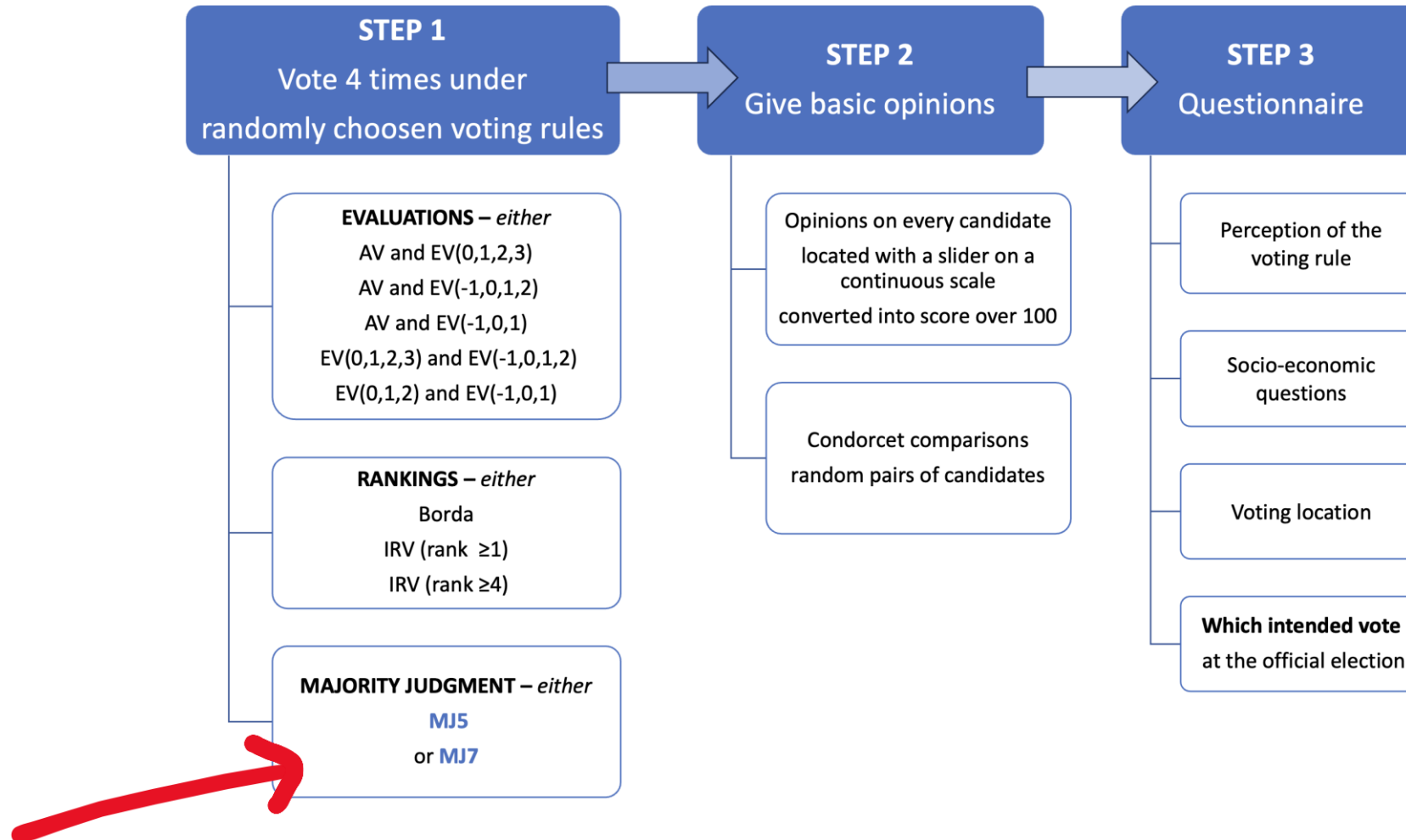


2 308 participants



No monetary incentives

The experiment design




Focused Samples

Each participant randomly tests **one out of two scales**, MJ5 or MJ7.

 **Not representative**

 **Highly homogeneous**

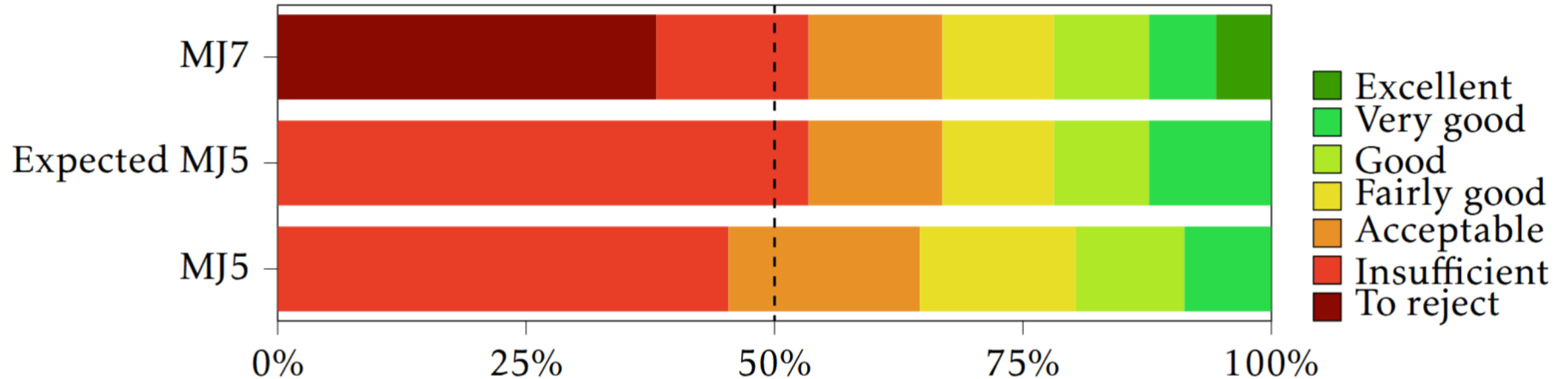
 We conducted **robustness check** with post-stratification weights.

	Our sample		General population
	JM7 (N=965)	JM5 (N=990)	
<i>Gender</i>			
Male	0.68	0.68	0.48
Female	0.32	0.32	0.52
<i>Age</i>			
18-29	0.44	0.42	0.17
30-39	0.27	0.30	0.15
40-49	0.18	0.15	0.16
50+	0.11	0.13	0.51
<i>Education</i>			
High school diploma	0.05	0.05	0.68
Higher education	0.95	0.95	0.32
<i>Vote</i>			
Jean-Luc Mélenchon (Far left)	0.61	0.64	0.22
Anne Hidalgo (Left)	0.02	0.02	0.02
Emmanuel Macron (Center)	0.12	0.11	0.28
Valérie Pécresse (Right)	0.00	0.01	0.05
Marine Le Pen (Far right)	0.02	0.02	0.23
Abstained	0.03	0.02	0.26
Blank vote	0.01	0.01	0.01

› **Results Pt.1**

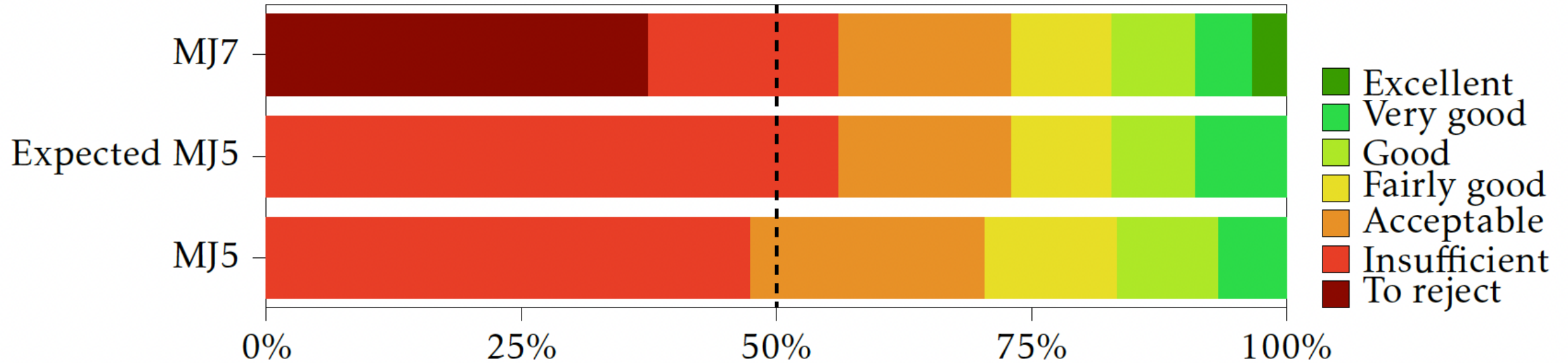
› **Descriptive Analysis**

Grade distributions



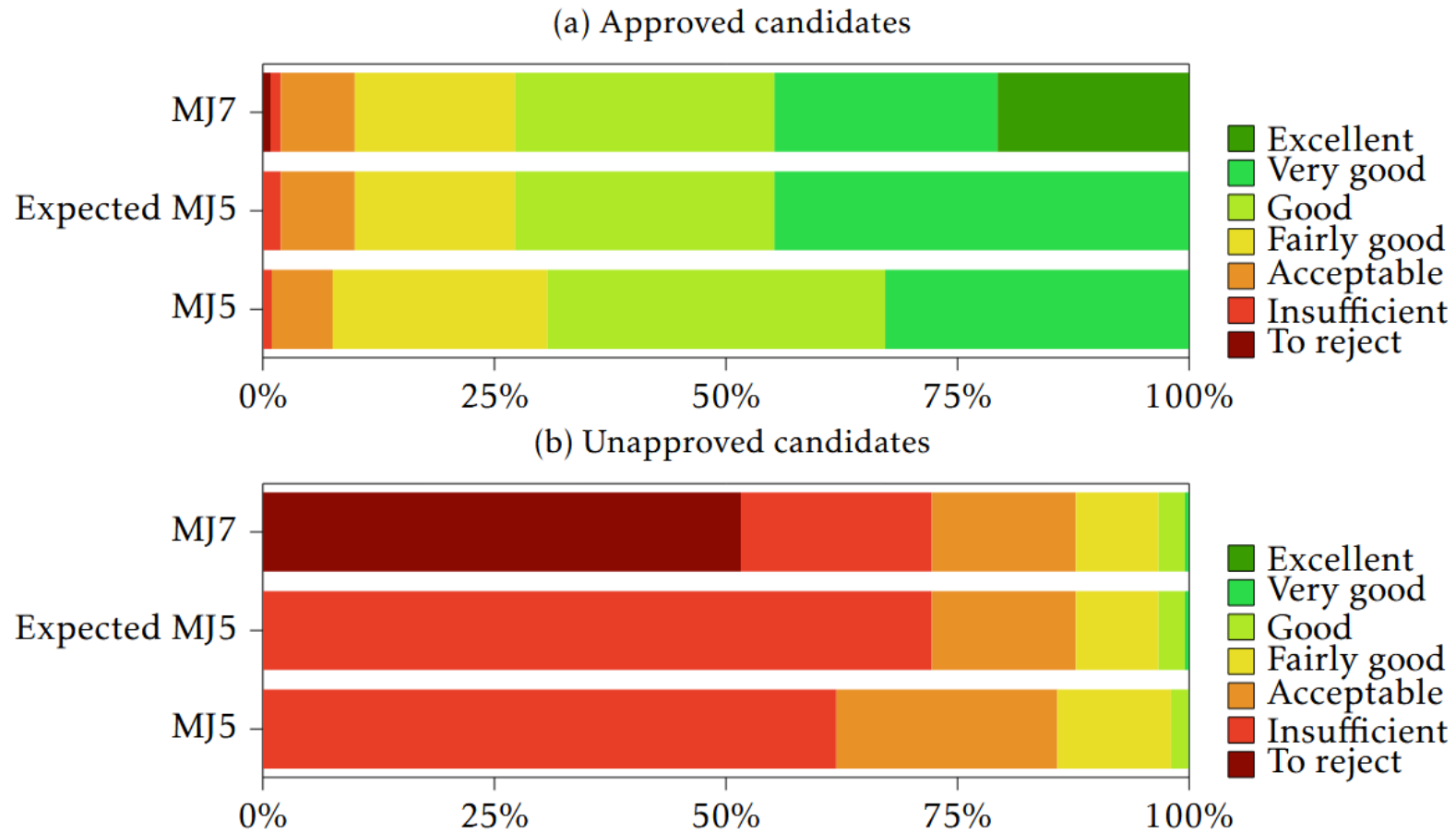
When we remove the two extreme grades (**excellent** and **to reject**), participants use more often the intermediary grades (**acceptable**, **fairly good** and **good**).

The case of Emmanuel Macron



The median grade for Emmanuel Macron shifts from **Insufficient** with MJ7 to **Acceptable** with MJ5.

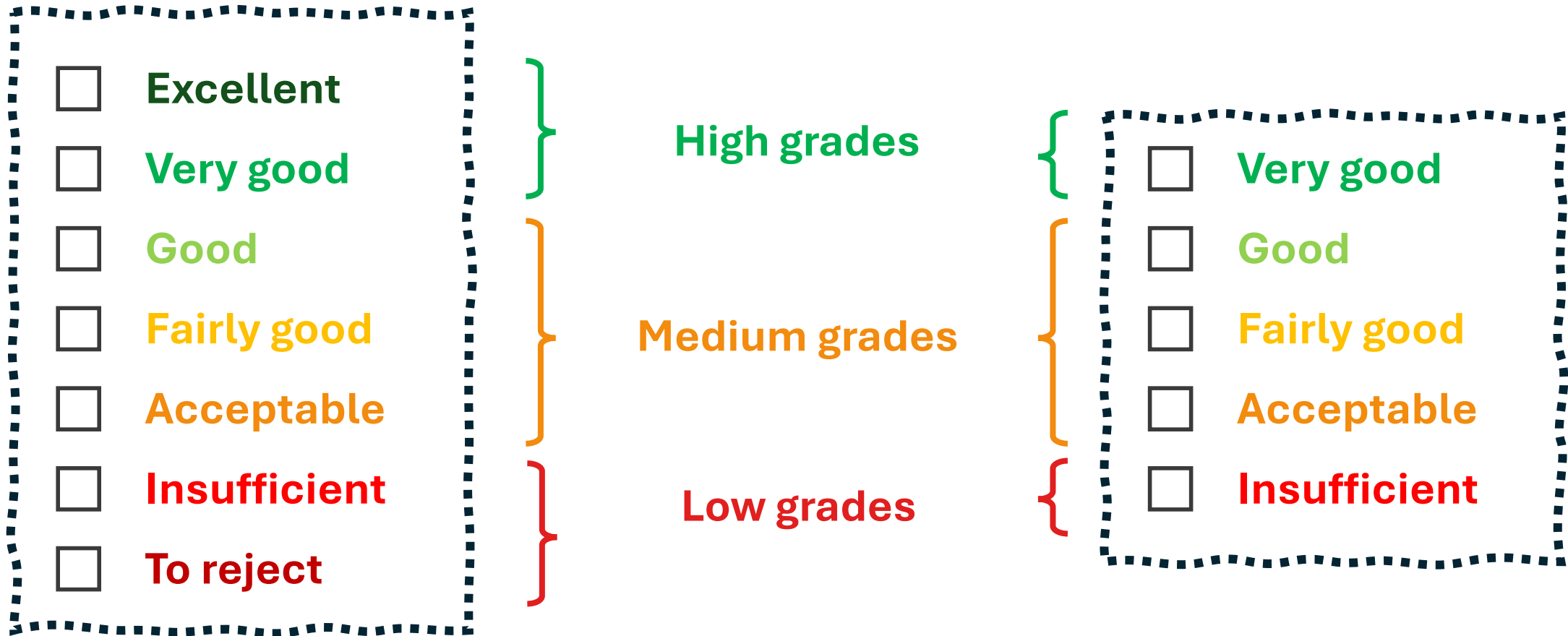
Disapproved candidates



› **Results Pt.2**

› **Regression Analysis**

Grade categories



The regression model

$$y_{ij} = \kappa + \beta \cdot \text{MJ5}_i + \gamma \cdot \mathbf{X}_i + \epsilon_{ij}$$

y_{ij} is 1 iff voter i uses a **medium** grade for candidate j (or **low**, or **high**).

MJ5_i is 1 iff voter i is assigned to the group voting with **MJ5**.

\mathbf{X}_i represents the **control variables** (gender, age, education level, and political preferences)

The regression model

$$y_{ij} = \kappa + \beta \cdot \text{MJ5}_i + \gamma \cdot \mathbf{X}_i + \epsilon_{ij}$$




We want to see if $\beta = 0$.

y_{ij} is 1 iff voter i uses a **medium** grade for candidate j (or **low**, or **high**).

MJ5_i is 1 iff voter i is assigned to the group voting with **MJ5**.

\mathbf{X}_i represents the **control variables** (gender, age, education level, and political preferences)

Results of OLS regression

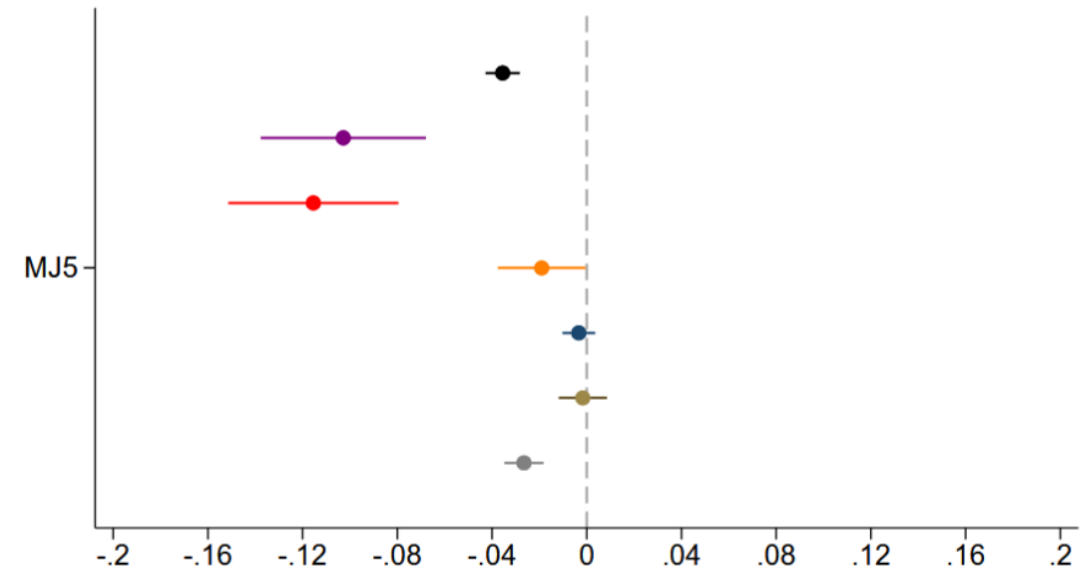
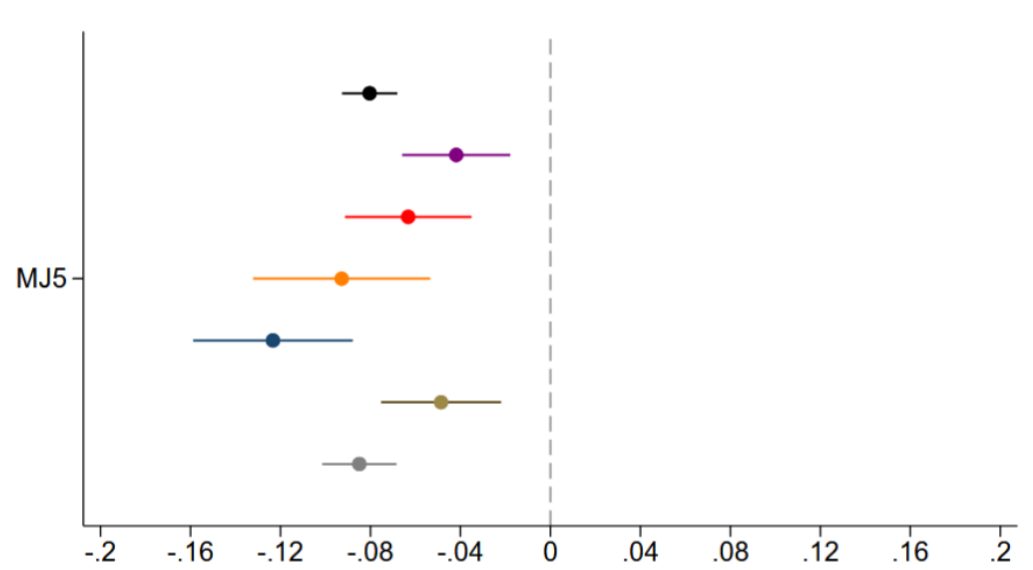
	(1)	(2)	(3)
	Low grades	High grades	Intermediate grades
MJ5	-0.080*** (0.006) 	-0.036*** (0.004) 	0.12*** (0.007) 
Dep. var. mean in MJ7	0.53	0.12	0.34
Observations	23460	23460	23460
Clusters	1955	1955	1955
Controls	✓	✓	✓
R-squared	0.010	0.011	0.015

The meaning of grades depends on the relative position on the scale used (every effects statistically significant at $p < 0.01$).

Regression for each candidate

Low grades

High grades

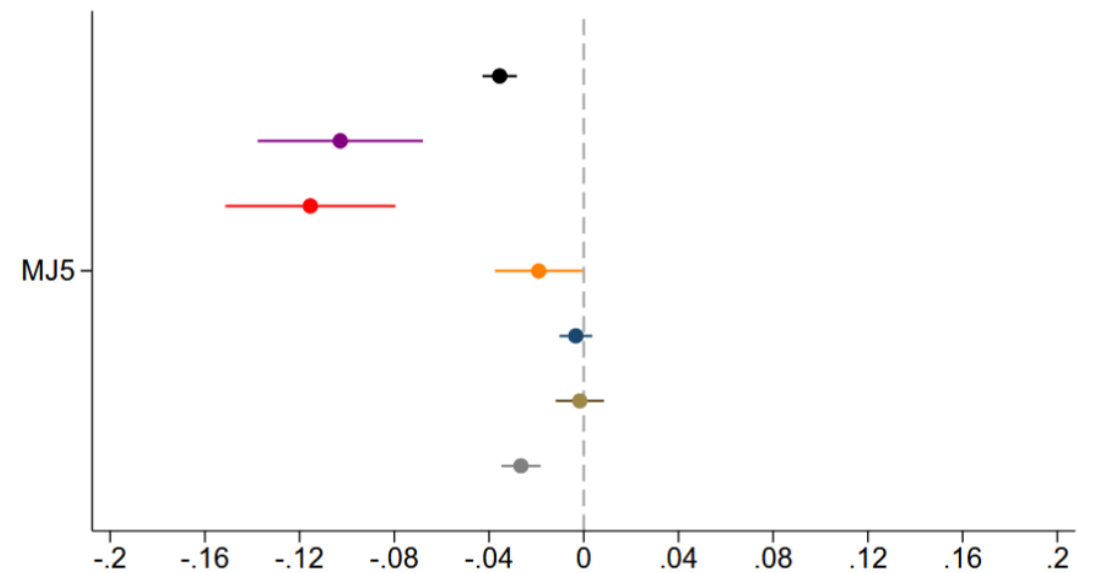
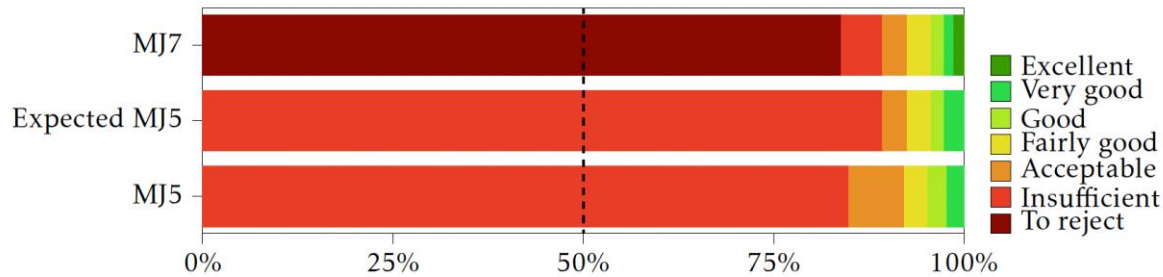


- All candidates
- Jadot (Left)
- Pecresse (Right)
- Others
- Mélenchon (Far left)
- Macron (Center)
- Le Pen (Far right)

Regression for each candidate

High grades

Marine Le Pen's grades

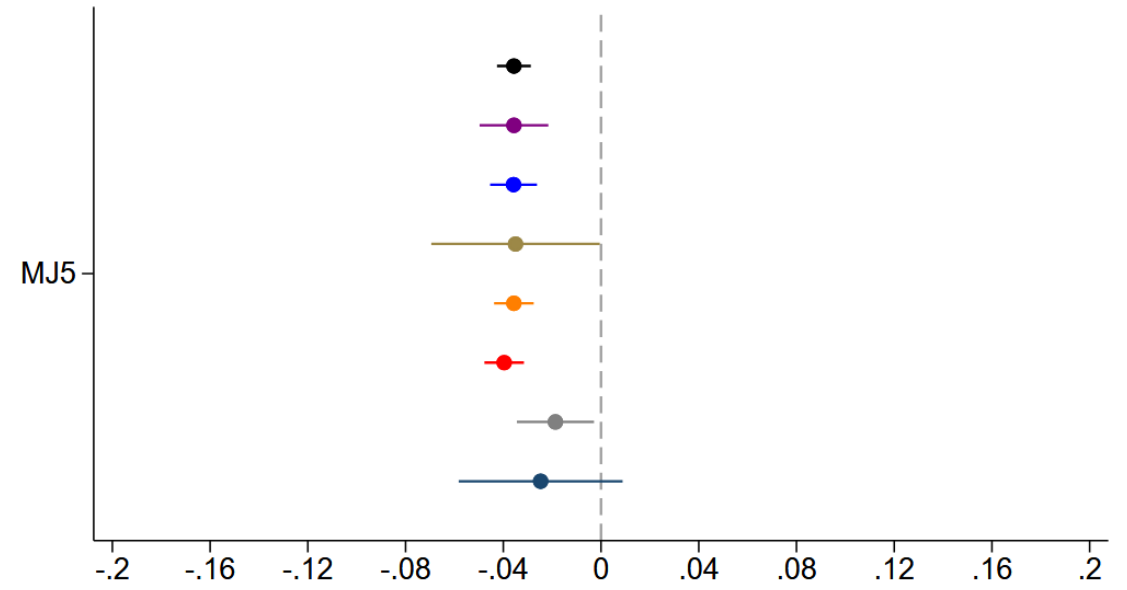
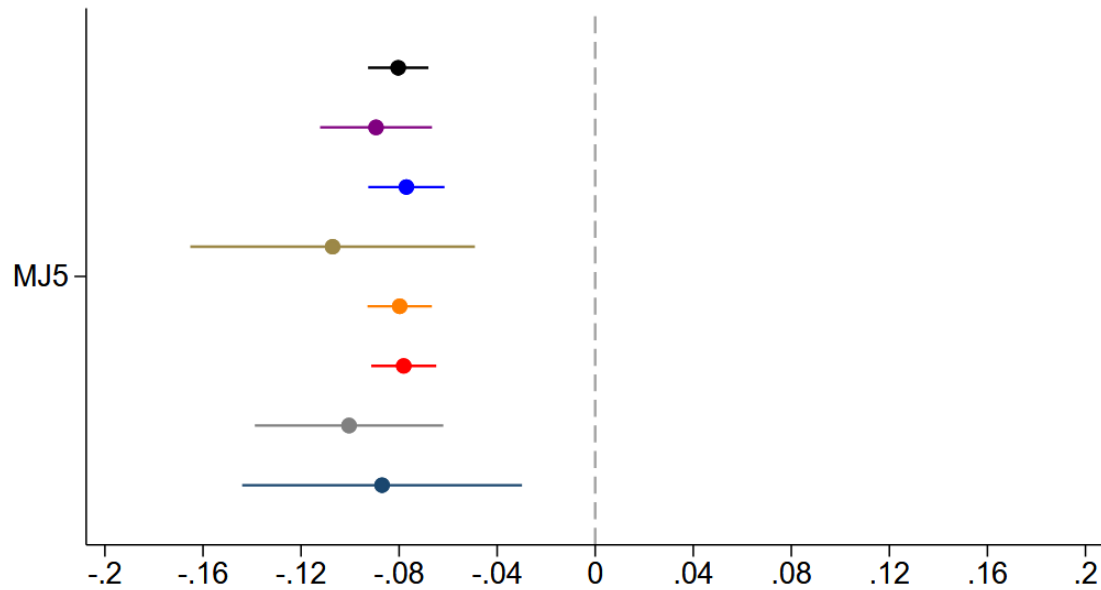


- All candidates
- Jadot (Left)
- Pecesse (Right)
- Others
- Mélenchon (Far left)
- Macron (Center)
- Le Pen (Far right)

Heterogeneity Analysis

Low grades

High grades



- Entire sample
- Female
- Male
- No higher education
- Higher education
- Left-wing
- Center
- Right-wing

Robustness checks

The self-selected and unrepresentative sample may raise concerns regarding the external validity of our results.

Thus, we reproduce regression **with post-stratification weights** based on demographics and political beliefs.

	Weights:Demographics			Weights:Political pref.		
	(1) Low grades	(2) High grades	(3) Intermediate grades	(4) Low grades	(5) High grades	(6) Intermediate grades
MJ5	-0.088*** (0.03)	-0.034*** (0.01)	0.12*** (0.03)	-0.067*** (0.03)	-0.021*** (0.006)	0.088*** (0.03)
Dep. var. mean in MJ7	0.53	0.12	0.34	0.53	0.53	0.53
Observations	23460	23460	23460	22728	22728	22728
Clusters	1955	1955	1955	1894	1894	1894
Controls	✓	✓	✓	✓	✓	✓
R-squared	0.027	0.022	0.036	0.013	0.029	0.015

Robustness checks



Other regression methods (**Probit**).



Randomization inference.



Controlling the **voting rules tried before** (for carry-on effect).

> Conclusion & Discussion

Discussion: other context changes

We showed that the meaning of the words **depend on the scale used**, and on the number of options.

Other changes in the context could also have an impact on the “absolute meaning” of words, like **the number of candidates**.

Discussion: one-shot effect

In our experiment, participants are **not used to vote with MJ** and are not familiar with the scale (one shot effect).

We can assume that if voters repeatedly vote with MJ, they become more **accustomed to the available grades** and refine their understanding of them. This could lead to a convergence toward associating certain individual assessments with specific grades and, therefore, a lower impact of framing effects.

Conclusion

We tested whether the use of grades in fictitious elections with MJ is impacted by a framing effect due to the grade scales. We have run an online experiment where participants were randomly assigned to a treatment in which they voted with either MJ5 or MJ7.

Take-away: the meaning of grades is significantly dependent on scales, even in cases in which participants had strong opinions over candidates.

These results suggest the **high sensitivity of the election outcomes' interpretation** to the scale used, and the utmost importance of choosing the appropriate language to represent citizens' views with MJ.

> Thanks for your attention!