University of Cyprus
Department of Economics

Working Paper 12-2019

Trimming Extreme Opinions in Preference Aggregation

Philippos Louis, Matías Núñez and Dimitrios Xefteris

Department of Economics, University of Cyprus, P.O. Box 20537, 1678 Nicosia, Cyprus
Tel.: +357-22893700, Fax: +357-22895028, Web site: http://www.ucy.ac.cy/econ/en

Trimming extreme opinions in preference aggregation

Philippos Louis*and Matías Núñez†and Dimitrios Xefteris‡

September 12, 2019

Abstract

The use of trimmed mean mechanisms in collective decision-making is motivated by the perception that they constitute a remedy for strategic misreporting. This work focuses on the strategic calculus of voting under such mechanisms and –contrary to the above presumption– it demonstrates both formally and experimentally that: a) voters persistently resort to strategic polarization for all but the most extreme levels of trimming and b) the outcome is more extreme and closer to the ideal policy of the median voter compared to when trimming does not take place. These so far uncharted properties of trimming provide novel insights –and call for caution– regarding its implementation.

Keywords: trimmed mean; equilibrium; experiment; collective decisions; facility location problem

JEL codes: D71, D72

1 Introduction

The need for aggregation is ubiquitous in organizations: groups aggregate preferences to reach collective decisions; websites aggregate product reviews to inform consumers; committees aggregate experts' information to give recommendations; courts aggregate judgements to reach verdicts. Of course, there is no unique way of performing such an aggregation. Choosing a mechanism to do so in each of these instances is guided by experience and often supplemented by the analysis of the experts in aggregation: statisticians. After all, estimation, which lies at the heart of statistical inference, is typically the aggregation of

*University of Cyprus
†CNRS & CREST, Ecole Polytechnique.
‡University of Cyprus

observations in a sample into a single value. Statistical theory has a lot to say about the properties of different estimators and can tell us, for instance, when using the median of a distribution instead of the mean will make a difference (see for example Lehmann and Casella, 2006).

Some caution is required, though, when applying statistical theory to solve aggregation problems as the ones described. The data generating process giving rise to a statistical sample is typically independent from the estimation process and generally unaffected by its result. However, such independence does not hold in many aggregation problems. Consider, for instance, the problem of a jury that needs to aggregate jurors' judgments to reach a verdict as first studied by Condorcet (1785). If we assume that there is an underlying correct verdict, a decision is taken by majority, and each juror is more likely to have identified the correct verdict than not, then –by the law of large numbers– it follows that the probability of a correct majority verdict converges to one as the number of jurors increases. Austen-Smith and Banks (1996) noted however that this result relies on the seemingly innocuous assumption that jurors vote sincerely –for what they believe is the correct verdict– and, famously, argued that sincere voting is rarely rational in such settings. Therefore, it is not unerring to directly apply standard statistical techniques to understand the asymptotic likelihood of a correct majority verdict, and a proper equilibrium analysis is warranted.

Notice that in the jury setting the problem arises even when jurors are assumed to have a common interest in reaching the right decision. This suggests that statistical arguments relying on sincere behavior may be even less robust in environments where participants in the aggregation process can have conflicting interests regarding its outcome. A growing literature at the interface of statistics, economics and computer science studies the properties of common estimation processes, such as linear regression, in the presence of this kind of "strategic noise" (see for instance Cai et al., 2015; Caragiannis et al., 2016, and references therein). One emblematic case where this issue arises is in the one-dimensional preference aggregation problem (Moulin, 1980).[1] Such problems are not only ubiquitous, but also plagued by the issue of "strategic noise". It suffices to think of the problem of setting the air conditioner temperature in a common office space. If the decision coincides with the mean of the workers' requests, the colleagues may exaggerate their preferences to achieve a temperature closer to what they wish for. Indeed, in such cases simple averaging "assigns voting power to cranks in proportion to their crankiness" (Galton, 1907). But if taking an average does not work, is there a simple way to overcome this problem?

The use of *trimmed means* has been proposed as a remedy for these issues. It involves the calculation of the mean after discarding given parts of a sample at the high and low end, typically an equal amount of both. For instance, the Olympic mean only discards the

[1] This is also known in the literature as the one-dimensional "facility location problem".

highest and the lowest value.[2] The interquartile mean discards the lowest 25% and the highest 25% and is used extensively in the computation of important financial benchmarks such as the LIBOR and the EURIBOR. From a statistical perspective, trimming can improve an estimator's efficiency, especially for the case of fat-tailed distributions, as the estimate becomes more robust to outliers.[3] It is believed that this statistical property also renders the trimmed mean immune to manipulation (see for example Eisl et al., 2017). So, for instance, in figure skating a single judge cannot manipulate the score by giving an extremely high or low score. While this idea is intuitive, surprisingly, there is no study that characterizes the equilibrium outcomes of trimmed mean mechanisms while also testing their properties in empirically relevant settings.[4]

In this paper we take the first step in this direction, and look at the class of trimmed mean aggregation mechanisms from a strategic perspective. Namely, we consider the game in which several players submit some value in the unit interval, the outcome coincides with the trimmed mean and the players' payoffs depend negatively on the distance between their peaks/ideal points and the trimmed mean. Our approach is both theoretical and experimental. It proceeds as follows:

On the theory side of this work, we show that, each trimmed mechanism leads, essentially, to a unique equilibrium. Its outcome can be fully characterized by the players' peaks and the degree of trimming (i.e., how many reports are trimmed in each extreme of the distribution). In this equilibrium all players, but possibly one, polarize: they submit extreme reports independently of whether some reports are trimmed or not. Intriguingly, the equilibrium outcome becomes *more extreme as the degree of trimming becomes higher* for every possible vector of ideal policies. That is, if the players' ideal policies are drawn from a certain well-behaved distribution that is symmetric about the center of the unit interval, then the equilibrium outcome of a trimmed mean mechanism is farther from the population mean (i.e., the center of the policy space) compared to the simple mean mechanism. In fact, the more we trim, the closer to the extremes the equilibrium outcome gets. Moreover, and somewhat less surprisingly, as the degree of trimming increases the equilibrium outcomes approaches

[2]This is used in some sports such as figure skating and in Farm Commodity Programs in the US. One is referred to Schnitkey (2012) for a detailed discussion.

[3]One is referred to Rothenberg et al. (1964), Bickel (1965) and Huber (1972) for seminal theoretical contributions and to Andrews and Hampel (2015), Stigler (1977), Hill and Dixon (1982) and Bryan and Cecchetti (1994) for empirical findings.

[4]The literature that considers the implications of trimming in a strategic setting is scant. Among them, the most recent ones are Hurley and Lior (2002) and Rosar (2015). Hurley and Lior (2002) use Monte Carlo simulations to compare the effect of trimming assuming that strategic voting occurs with a positive probability. Rosar (2015) compares the median and the average rule in model with interdependent preferences and incomplete information; yet, the focus of the paper is not on trimming even though some lessons on trimming with a large number of players are drawn. See also, Bassett Jr and Persky (1994) for a model with trimmed means, honest voting and Monte Carlo simulations and Yaniv (1997) for a description of heuristics in judgment aggregation.

the ideal policy of the median voter. Indeed, when trimming becomes extreme the outcome literally coincides with the median report, and hence a gradual transition of the equilibrium outcome to the median peak seems as a reasonable consequence as the degree of trimming increases.

This stark difference regarding the theoretical properties of trimming between statistical and strategic settings calls for careful empirical investigation. The idea that trimming should disincentivize instrumental extreme reports seems compelling and, hence, our theoretical finding that the degree of strategic behavior is unaffected by the degree of trimming might not prove relevant in contexts of applied interest. In fact, if real individuals are substantially discouraged from submitting extreme reports when such reports are trimmed away, then it might be very likely that trimming leads to more moderate outcomes than the simple mean mechanism (as possibly desired by the mechanism designer). For this reason we conduct a laboratory experiment in which groups of five subjects are asked to make collective choice following three alternative mechanisms: the Simple Mean (no trimming), the Olympic Mean (the highest and lowest reports are trimmed), and the Median (the two highest and the two lowest reports are trimmed).

The results point clearly in favour of our main theoretical prediction: for given preferences of the group, the outcome of the decision procedure becomes more extreme and moves closer to the ideal policy of the median voter as the degree of trimming becomes higher. The difference is very strong between any pair of mechanisms, and this establishes in a robust manner that in strategic environments trimming pushes the outcome in the predicted direction. Importantly, though, the experiment also weakly justifies the common perception that trimming mitigates strategic misreporting: individuals always misreport, but, conditional on being extremists, choose to report values closer to their ideal policies as the degree of trimming increases. When trimming becomes extreme (i.e., when all reports except the median one are trimmed away) then incentives to misreport vanish and subjects behave more sincerely. These additional insights –that could not be drawn from the theoretical analysis– lead to the following key observation: While trimming pushes subjects mildly towards more sincere behavior, the effect is not strong enough to counterbalance the centrifugal force that it induces on the outcome. Indeed, if non-extreme trimming (e.g., the degree of trimming employed in the Olympic mean mechanism) could induce substantially more sincere behavior then in several cases it would lead to more moderate outcomes than simple averaging (which incentivizes, unambiguously, strategic behavior). As we find though, as long as the degree of trimming remains non-extreme the strategic forces continue to dominate, subjects polarize and misreport broadly to the same extent, and the outcome becomes more extreme as trimming increases.

Overall, we consider that these first results provide a solid groundwork for further analy-

sis of these popular mechanisms and pin down important and, so far, unidentified properties of trimming in strategic contexts. Apart from the academic debate, our findings also call for caution when it comes to deciding whether a trimmed mean mechanism should be employed in real-life settings. An outcome-oriented mechanism designer who wants to extract information regarding a value of interest (i.e., the true quality of the performance of an athlete) should choose the mechanism that –taking in account the potential strategic behavior of agents– leads to the most accurate estimate. Indeed, disincentivizing strategic misreporting seems a natural target from this perspective. But as we find, this is only a simplistic view of the problem: if sincerity increases slowly with the degree of trimming, then the effects on the outcome might be completely opposite than desired. The loss of information due to trimming seems to be in several cases much higher than the induced increase in sincere behavior, and this leads, both in theory and in the laboratory, to more extreme outcomes and potentially less accurate estimates of the target values. Hence, it seems plausible that the best way to induce the desired outcome (e.g., to moderate the outcome as much as possible) might be by not trimming extreme reports at all.

After developing our theoretical analysis in Sections 2 and 3, and illustrating some key facts in Section 4, Section 5 presents the experimental design and the hypotheses to be tested. Results are presented in Section 6.

2 Theoretical Setting

Let $A := [0,1]$ denote the set of alternatives and $N := \{1,\dots,n\}$ the set of players with $n = 2k+1$ for some positive integer k. That is, we focus on the case where n is odd but a similar analysis can be conducted for the case in which n is even. Each player i has utility function u_i in U, the set of single-peaked preferences on the set of alternatives, with $u_i(x)$ the utility of player i when $x \in A$ is implemented. The player's utility function, u_i, reaches its maximum at its unique peak, $p_i \in A$, so that $u_i(x') < u_i(x'')$ when $x' < x'' \leq p_i$ and when $p_i \leq x'' < x'$. A social choice function is a function $f : U^n \to A$ that associates every $u = (u_1,\dots,u_n) \in U^n$ with a unique alternative $f(u)$ in A.

Ordered vectors and generalized medians. For any positive integer z and any finite collection of *points* $x = (x_1,\dots,x_z)$ in $[0,1]^z$, we let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_z)$ denote the ordered profile associated to x with $\tilde{x}_1 \leq \tilde{x}_2 \leq \dots \leq \tilde{x}_z$. Note that in case of ties in x, the element with the lowest index in x is associated the lowest index in the ordered profile $\tilde{x}$. Given these specifications, the ordered profile $\tilde{x}$ is uniquely defined for each collection x. The midpoint of $\tilde{x}$ is the left median of x (the smallest x_k for which $\#\left\{\ell \mid x_\ell \leq x_k\right\} \geq \frac{z}{2}$) and we denote it by $m(x)$. A social choice function is a generalized median rule (GMR) if there is some collection

of points $\kappa_1, \ldots, \kappa_{n-1}$ in $[0,1]$ such that, for each $u \in U^n$, $f(u) = m(p_1, \ldots, p_n, \kappa_1, \ldots, \kappa_{n-1})$. We refer to $\kappa = (\kappa_1, \ldots, \kappa_{n-1})$ as the phantom vector or vector of calibration parameters of the GMR and to $p = (p_1, \ldots, p_n)$ as the peak vector. Essentially, the outcome of the GMR with phantom vector κ is the ideal policy of the median voter of the group composed of the n players with peaks in p plus $n-1$ phantom/artificial voters with peaks in κ.

Nash and strong Nash equilibria. A mechanism is a function $\theta : S^n \to A$ that assigns to every $s \in S^n$, a unique element $\theta(s)$ in A, where S is the strategy space of player i. Given a mechanism $\theta : S^n \to A$, the strategy profile $s \in S^n$ is a Nash equilibrium of θ at $u \in U^n$, if $u_i\left(\theta(s_i, s_{-i})\right) \geq u_i\left(\theta(s_i', s_{-i})\right)$ for all $i \in N$ and any $s_i' \in S$. Similarly, given a mechanism $\theta : S^n \to A$, the strategy profile $s \in S^n$ is a strong Nash equilibrium if there is no $C \subseteq N$ with $u_i\Big(\theta(s^C, s^{N\setminus C})\Big) > u_i\Big(\theta(s)\Big)$ with $s^C = (s_i^C)_{i\in C}$ and $s^{N\setminus C} = (s_i)_{i\in N\setminus C}$.

Trimmed Mean Mechanisms. We consider trimming mechanisms that (i) request each player to announce one alternative and (ii) select a single alternative as an outcome. The different trimming mechanisms differ only on one dimension: the number of trimmed reports. The degree of trimming is denoted by ω and since we consider symmetric trimming (the same number of trimmed reports from below and from above) and $n = 2k+1$, it follows that ω belongs to $\{0, \ldots, k\}$. The trimmed mean of degree ω, denoted θ_ω, drops the ω highest and the ω lowest reports and implements the average of the remaining values. It follows that, in each such mechanism, the strategy space for each player equals $S = [0,1]$ and therefore $\theta_\omega : [0,1]^n \to [0,1]$ associates to any strategy profile $s \in [0,1]^n$ the outcome:

$$\theta_\omega(s) = \text{Average}(\tilde{s}_{\omega+1}, \ldots, \tilde{s}_{n-\omega}).$$

This family of trimming mechanisms includes, among others, the Mean Mechanism, the Olympic Mean and the Median Mechanism. The Mean mechanism corresponds to the case without trimming ($\omega = 0$) since it simply implements the average of the reports. The Olympic mean mechanism trims the highest and the lowest value so that $\omega = 1$. The Median mechanism ($\omega = k$) selects the median of the reports, which is equivalent to the average of the unique value that remains after trimming the ω lowest and highest reports so that $\theta_k(s) = m(s)$ for each $s \in [0,1]^n$.

3 Equilibria with Coalitional Deviations

Our theoretical analysis of these mechanisms is based on the concept of strong (Nash) equilibrium (Aumann, 1959). This equilibrium concept refines the classical notion of Nash equilibrium. In a strong equilibrium, no collective profitable deviation exists for any group of

agents (or coalition) whereas in a Nash equilibrium no individual profitable deviation exists. One of the reasons of our focus on strong equilibria is that Nash equilibrium has almost no predictive power in our setting, as the next lemma shows.

Lemma 1. *For every admissible preference profile, any outcome can be sustained in a Nash equilibrium of the mechanism θ_ω as long as $\omega > 0$.*

The logic of this indeterminacy of Nash equilibria is immediate: as long as all players announce the same value, this constitutes a Nash equilibrium since no unilateral deviation can alter the outcome. This, in turn, triggers the abundance of equilibrium outcomes (see the proof in the appendix) since any alternative can be implemented in a Nash equilibrium.

Of course, not all Nash equilibria are equally plausible. In fact, this equilibrium multiplicity is largely a theoretical artifact: on the one hand it requires a high degree of coordination –i.e., all players can somehow infer that everybody else will announce $x \in A$– but at the same time players that could profit from a mutual deviation (e.g., players with peaks to the left of x) cannot effectively coordinate. For this reason, like Moulin (1980), we turn to solution concepts that are robust to communication and coordination attempts. Indeed, the contrast with strong equilibria is steep: each of trimming mechanisms under consideration admits a strong equilibrium and its outcome is unique as will be discussed in the rest of the section.

Theorem 1. *For each trimming degree ω:*

1. *the game-form associated to the mechanism θ_ω admits a strong equilibrium for every admissible preference profile.*

2. *every strong equilibrium s of this game-form with peak profile p satisfies $\theta_\omega(s) = m(p, \kappa_1^\omega, \ldots, \kappa_{n-1}^\omega)$ with, for each $j = 1, \ldots, n-1$ and each $\omega = 0, \ldots, k$:*

$$\kappa_j^\omega = \min\left\{\max\{0, \frac{j-\omega}{n-2\omega}\}, 1\right\}.$$

3. *in every strong equilibrium s and each $\omega < k$, each player with $p_i < \theta(s)$ plays 0 and each player with $p_i > \theta(s)$ plays 1.*

Theorem 1 presents in our view a very appealing property of trimmed mean mechanisms: the existence of strong equilibria and the uniqueness of its outcomes.

As far as existence is concerned, it is well known that strong equilibria seldom exist since they impose no restriction on how coalitions choose their profitable deviation. In our games, this criticism does not apply since a strong equilibrium exists for *every* possible peak

specification. However, we are aware that strong Nash equilibria are often criticized by imposing no restriction on how coalitions choose their profitable deviation and that, often, the concept of coalition-proof Nash equilibria is deemed superior. Yet, following the results of Yamamura (2011), one can prove that in each of the games under consideration, the sets of strong Nash and Coalition-Proof Nash equilibria coincide (Bernheim and Peleg, 1987), so that the results identified here do not hinge on the specific details of communication between agents.

Regarding the uniqueness of the equilibrium outcome, Theorem 1 shows that this outcome coincides with the median of the peak profile p and of the calibration vector κ^ω. Using implementation jargon, the trimmed mean mechanism θ_ω implements the GMR with phantom vector κ^ω in strong equilibrium.[5] In the specific case in which $\omega = 0$ our result aligns with Renault and Trannoy (2005) and Yamamura and Kawasaki (2013) who have characterized the Nash equilibrium of the mean mechanism without trimming.

Finally, note that our prediction is that individual behavior will be polarized in every trimmed mechanism θ_ω, with the possible exception of the median mechanism, where $\omega = k$. This is a very crucial point of our theoretical analysis as it directly contradicts the common perception that trimming mitigates the incentives for strategic misreporting and postulates that unless it is extreme (i.e., unless all but the median report are trimmed), players have strong incentives to polarize and report opinions far from their true ones.

In the case of extreme trimming, where $\omega = k$, voting for one's peak is a weakly dominant strategy. Therefore, this mechanism admits multiple (strong) equilibria. Nevertheless, in all equilibria the median voter votes for her peak and all equilibria lead to the same unique strong Nash equilibrium outcome for this mechanism, which is the median voter's peak.

Having characterized the equilibrium outcome as a function of the trimming degree, we now turn our attention to some important comparative statics. Indeed, since for each preference profile and each trimming degree ω the outcome is unique, this allows us to understand how trimming affects the distance of the equilibrium outcome from the center of the policy space and from the ideal policy of the median voter. For each $p \in [0,1]^n$ and each $\omega = 0, \ldots, k$, the equilibrium outcome associated to the mechanism θ_ω is henceforth denoted by :

$$m_\omega(p) = m(p_1, \ldots, p_n, \kappa_1^\omega, \ldots, \kappa_{n-1}^\omega).$$

As Theorem 2 shows, the higher the trimming degree (the higher ω), the closer the outcome $m_\omega(p)$ is to the median of the peaks $m(p)$ and the further away from the midpoint of the outcome space. More formally,

[5] For general results regarding implementation in strong Nash equilibrium one is referred to Maskin (1978) and Dutta and Sen (1991).

Theorem 2. *Take any pair of Trimmed Mean mechanisms θ_ω, θ_ϕ with $\omega < \phi$. Letting $\|\cdot\|$ stand for the Euclidean distance, the associated strong equilibrium outcomes satisfy:*

a.) the lower the trimming degree, the closer to the midpoint of A so that:

$$\|m_\omega(p) - \frac{1}{2}\| \leq \|m_\phi(p) - \frac{1}{2}\|.$$

b) the higher the trimming degree, the closer to the median peak so that:

$$\|m_\phi(p) - m(p)\| \leq \|m_\omega(p) - m(p)\|.$$

Theorem 2 gives a clear illustration of the effect of trimming in strategic contexts. To understand the importance of this result, one should contrast it with the effect of trimming when reports are sincere: in such cases trimming, by removing outliers, may lead to more or less moderate outcome depending on the exact peak distribution (we present examples in thee next section). Our results show that this ambiguity disappears in a strategic setting. Indeed, among the different trimmed mechanisms, the one for which the outcome $m_\omega(p)$ is closer to the center of the interval is the mean rule, that is the only mechanism of this family in which no trimming occurs. In reality, the above theorem is even stronger since it applies for each pair of mechanisms: the less trimming, the more centered the decision becomes.

Finally, our results also underline a different effect: the more trimming, the closer the outcome shifts towards the median of the peaks. This is somewhat more intuitive since among the mechanisms under consideration, the one that exhibits the highest degree of trimming is the median mechanism ($\omega = k$) and its unique equilibrium outcome coincides with the median of the peaks.

4 The Mean, the Olympic Mean and The Median

In this section we present two examples of peak distributions that explain why trimming has ambiguous effect on the outcome when behavior is sincere and why this ambiguity disappears when players are strategic, and we also provide a link between our theoretical and our experimental analysis. As argued, for each peak profile p and each trimming degree ω, the (strong) equilibrium outcome is unique. Thus, the discussion deals simply with two parameters:

$\theta_\omega(p)$: Sincere voting outcome of trimming

$m_\omega(p)$: Strategic voting outcome of trimming.

For the sake of clarity and consistency with our experimental analysis, the present discussion focuses on a five-member committee. Therefore, we denote by $N = \{1, \ldots, 5\}$ the set of agents and consider three mechanisms: the simple mean mechanism SM ($\omega = 0$, i.e., no trimming), the Olympic mean mechanism OM ($\omega = 1$, i.e., trimming the highest and the lowest report) and the median mechanism M ($\omega = 2$, i.e., trimming the two highest and the two lowest reports). In all mechanisms, each player i submits a report $s_i \in [0, 1]$ and the mechanism selects an alternative in $[0, 1]$. Under sincere voting, $s_i = p_i$, and under strategic voting we have, generically, $s_i \neq p_i$. Table 1 summarizes the outcome of the different mechanisms under sincere and strategic (strong Nash equilibrium) voting.

Mechanism	ω	**Calibration vector**	**Sincere voting outcome** $\theta_\omega(s)$	**Strategic voting outcome** $m_\omega(p)$
SM	0	$\kappa^0 = (.2, .4, .6, .8)$	$Average(s_1, \ldots, s_5)$	$m(p, \kappa^0)$
OM	1	$\kappa^1 = (0, \frac{1}{3}, \frac{2}{3}, 1)$	$Average(s_2, s_3, s_4)$	$m(p, \kappa^1)$
M	2	$\kappa^2 = (0, 0, 1, 1)$	$Average(s_3) = s_3$	$m(p, \kappa^2) = m(p)$

Table 1: **Mechanisms and outcomes.** For presentation purposes it is assumed here that the vector of peaks p is such that $p_1 \leq p_2 \leq p_3 \leq p_4 \leq p_5$. Note that when voting is sincere $s_i = p_i$.

In order to clarify the effect of trimming, we now analyze two examples that are representative of two diverse classes of cases.

Example 1: aligned effects of trimming. Consider the following profile of peaks:

$$p_1 = .05, p_2 = .15, p_3 = .25, p_4 = .65 \quad \text{and} \quad p_5 = .85.$$

This profile is left-biased since the median of the peaks is located to *the left* of the center of the interval. Table 2 summarizes the outcomes that obtain under different mechanisms under either sincere or strategic voting. For strategic voting it also shows the strategic reports submitted by voters. These are examples of strong equilibrium profiles sustaining the equilibrium outcome as shown in Theorem 1.

Note that the effect of trimming on the outcome goes in the same direction for both sincere and strategic voting. Indeed, under both assumptions, the outcome shifts from right to left and gets as close as possible to .25, the median of the peaks. For strategic voting

			Peaks					
			p_1	p_2	p_3	p_4	p_5	
			.05	.15	.25	.65	.85	
Mechanism	ω	**Sincere outcome**	**Strategic reports**					**Strategic outcome**
			s_1	s_2	s_3	s_4	s_5	
SM	0	.39	0	0	0	1	1	.4
OM	1	.35	0	0	0	1	1	.33
M	2	.25	0	0	.25	1	1	.25

Table 2: **The effects of trimming in Example 1**. Under both sincere and strategic voting, trimming shifts the outcome towards the median peak.

this is exactly what is predicted by Theorem 2. For sincere voting it is a result of the particular peak profile used here. The next example demonstrates that for sincere voting the comparative statics depend on the peak profile, something which is not true under strategic voting.

Example 2: misaligned effects of trimming. Consider the following profile of peaks:

$$p_1 = 0, p_2 = 0, p_3 = .3, p_4 = .36 \quad \text{and} \quad p_5 = 1.$$

This peak profile is again left-biased since the median of the peaks is located to *the left* of the center of the interval. As one can observe in Table 3, the effect of trimming does not push the outcome in the same direction when considering sincere and strategic behavior.

Indeed, under strategic behavior, the outcome shifts from right (i.e., .36) to left (i.e., .33) and gets as close as possible to the median of the peaks as ω increases. With sincere behavior, the outcome starts at .332, then gets more extreme (i.e., .22 for the Olympic mean) and then becomes more moderate again for the median mechanism.

Broadly speaking, trimming does not have the same effects under sincere and strategic voting. With sincere voting the trimming may have a non-monotonic effect and this will always depend on the particular peak profile. Under strategic voting the effect is always in the same direction as dictated by Theorem 2. Whether real subjects will choose to employ one or the other behavioral rule is obviously an empirical question, and we will try to address it in the sections that follow by the means of a laboratory experiment.

			Peaks					
			p_1	p_2	p_3	p_4	p_5	
			0	0	.3	.36	1	
Mechanism	ω	**Sincere outcome**	**Strategic reports**					**Strategic outcome**
			s_1	s_2	s_3	s_4	s_5	
SM	0	.332	0	0	0	.8	1	.36
OM	1	.22	0	0	0	1	1	.33
M	2	.3	0	0	.3	1	1	.3

Table 3: **The effects of trimming in Example 2**. Here, under sincere voting the effect of trimming is not monotonic. For strategic voting the predictions of Theorem 2 hold under any peak profile.

5 Experimental Design and Hypothesis

5.1 Experimental Design

The aim of the experimental design is to test the theoretical predictions concerning the effect of trimming on voter's behavior and the outcome. We use a between-subject design with three treatments. In each treatment, subjects make collective decisions using one of the three mechanisms described in the previous section: the Simple Mean (SM), the Olympic Mean (OM) and the Median (M). The decision rules are such that the outcome is unique for each admissible profile, allowing us to compare the effect of trimming since the mechanisms differ only on the degree of trimming.

The experiment took place at the University of Cyprus Lab of Experimental Economics (UCY-LExEcon). A total of 135 subjects, all students of the University of Cyprus participated in 9 equally sized sessions, with 3 sessions per treatment. Recruitment was done using ORSEE (Greiner, 2015). The experiment was computerized, and the software was programmed and run using zTree (Fischbacher, 2007). An outline of the design is presented in Table 4.

Timing and the experimental task. In all three treatments, subjects receive written instructions after entering the lab. These are also read aloud to establish common knowledge. In each round, subjects are placed in a group of five. Each group needs to choose collectively an integer between 1 and 100 as the group's destination. Each group member has an individual starting point, that is, a different integer between 1 and 100. The payoff in each period is then 100 points minus the distance between the destination and the subject's

Treatment	Rule	Group size	Subjects per session	Sessions	N	Rounds
SM	Simple mean	5	15	3	45	80
OM	Olympic mean	5	15	3	45	80
M	Median	5	15	3	45	80*

Table 4: **Experimental design.** *Due to a technical problem in the first session of the M treatment, it was only possible to conduct the first 76 rounds. For the remaining sessions subjects played all 80 rounds.

starting point. Starting points are common knowledge and are different for every subject in each round. Groups are reshuffled randomly in each round, and subjects do not know the identity of the other group members.

Parameter selection. The only parameters that are different between subjects and rounds were the subjects' starting points, which determine their payoffs. Nevertheless, the exact same set of parameters was used across all nine sessions. That is, for any combination of starting points used for a group in a specific round of a session, there was another group in all other sessions with the same starting points in the same round. Furthermore, the exact same sequence of parameters was assigned to subjects in all sessions.

The values for the starting points are chosen in a way that allows us to better detect differences across treatments. For instance, for any profile of starting points where the median voter lies close to the center of the policy space there is no difference in the outcome across the three mechanisms we use. Choosing starting points randomly would result in a large number of such profiles, eroding the power of our experimental design to detect differences in the outcome across mechanisms. Instead, we make a selection of profiles that allows us to cover all cases where there should be differences across at least two treatments.

In particular, for the design we require 240 different starting point profiles. The theoretical difference in the (strong) Nash equilibrium outcomes in treatments SM vs. OM can lie between zero and seven points, while for OM vs. M it goes from zero to 33, with values above 24 being more rare. We therefore chose 200 profiles such that the differences in equilibrium outcomes cover the $[0, 7] \times [0, 24]$ surface uniformly and another 40 profiles that have equilibrium outcomes with OM vs. M differences above 24. Given these desiderata, the selection of the particular profiles used in the experiment was random.

Treatments. As mentioned earlier, a different decision rule is used in each treatment to select the group's destination. In all three treatments subjects vote for one location choosing an integer from 1 to 100. The collective outcome is determined with one of the following rules:

Simple Mean: the collective outcome is the mean of all five votes.

Olympic Mean: the collective outcome is the mean of the three central votes (after dropping the highest and the lowest votes).

Median: the collective outcome is the median vote.

Voting and time limit. Voting in each round in all three treatments lasts for $20+x$ seconds, where x is a number between one and five, chosen randomly in each round and not known to the subjects. During this time, each subject is informed about her and others' starting points and can enter her vote. She can also observe the votes entered by other group members in real time. At any given point in time, the software calculates the destination and the payoffs for each subject. These are shown on the screen as a clock counts down from 20 seconds. After 15 seconds, a text starts blinking indicating that time is almost up. After the initial 20 seconds have passed it turns red for the remaining x seconds and indicates that voting may finish at any moment. The destination for the period is determined by the votes entered when the $20+x$ seconds finish. Finally, a screen appears informing subjects about the results of the voting: the votes and the payoffs for each subject and the final destination. The round finishes and a new round begins.

Payments. Ten rounds are chosen randomly and payoffs in these rounds are used to determine the subject's payment for the experiment. Subjects receive 1 € for every 80 points earned in the selected round, plus an additional 5€ as a participation fee. Subjects earned 15.10 € on average across all sessions.

5.2 Hypotheses

The theoretical results in the previous section suggest a few hypotheses to test using our experimental design. These can be classified as pertaining to either individual voting behavior or the aggregate outcome of these. We start with the latter.

Theorem 1 gives a unique prediction about the voting outcome in each treatment: the strong Nash equilibrium (sNE). For SM and OM the sNE differs from the outcome that obtains if all voters vote sincerely. For M the two coincide as they both give the median's

starting point as the outcome.

Hypothesis 1. In treatments SM and OM the outcome will be close to the sNE and far away from the sincere outcome. In treatment M the outcome will be the median voter's preferred point, as predicted by both the sNE and sincere play.

From Theorem 2 it follows that as we move to treatments with a higher degree of trimming, we expect to see the outcome move closer to the median voter's starting point. In fact, in treatment M, the former should be identical to the latter. But Theorem 2 also predicts that outcomes will lie between the median and the policy space mid-point. Thus, outcomes closer to the median are also further away from the center. These comparative static results are reflected in the following hypotheses.

Hypothesis 2a. As we move from treatment SM to OM and then to M, the outcome moves closer to the median voter's starting point.

Hypothesis 2b. As we move from treatment SM to OM and then to M, the outcome moves further away from the center of the policy space.

Regarding individual voting behavior we again have a strong prediction for treatments SM and OM. There we mostly expect most voters to move to the extremes of the policy space, with the exact number depending each time on the exact profile of starting points in a group. In treatment M there are multiple equilibria that all lead to the same outcome. There is always one in which all voters (except the median one) vote for one of the extremes and the median one votes for her preferred outcome. Everyone voting sincerely is also an equilibrium in this treatment. Given the multiplicity we expect a less polarized distribution of votes in this treatment.

Hypothesis 3. Individual votes in the SM and OM treatments will concentrate on the extremes of the voting space. The distribution of votes in M will be less polarized.

As discussed in the introduction, the motivation to use trimmed mean mechanisms in practice is to moderate incentives for participants to misrepresent their preferences. Our theory of strategic play essentially predicts that at least four out of five players will always choose an extreme (and, hence, largely insincere) report in SM and in OM. Therefore, one substantial difference is that sincerity should be higher in M compared to the other two treatments. But if one investigates non-equilibrium dynamics one can see that in SM strategic incentives to misreport one's preferences are somewhat more salient compared to the ones in OM. To see this consider a strategy profile where no player chooses an extreme policy and the outcome coincides with the median voter's ideal policy. If the most leftist player has chosen her ideal policy, then, it is easy to validate that under OM this player has no incentives to move farther from her ideal policy and towards zero, while under SM she

does. Of course, in both cases more central players might have incentives to deviate. Still this example demonstrates that the extremists' drive to exaggerate might be dampened under trimming when players deviate from perfect equilibrium behavior. Finally, recall that in M sincere voting is a weakly dominant strategy. So while multiple equilibria exist, all yielding the same outcome, we do expect to observe higher levels of sincerity in this treatment.

Anticipating some deviations from sNE, and given the above discussion, we can conjecture that individual votes will be closer to the ideal policies of the corresponding players as trimming increases.

Hypothesis 4. Individual votes will be closer to their respective ideal points in treatment M, followed by OM and then SM.

6 Results

6.1 Voting outcome

We first compare the outcomes across treatments to different theoretical benchmarks. These are summarized in Table 5.

We start with the prediction of the strong Nash equilibrium. Overall, the sNE does a good job predicting the voting outcome. The average deviation from sNE in both SM and OM is statistically significant but rather small in magnitude. In M we do not find any significant difference from sNE. More importantly, and in line with our theoretical results, it seems that in SM and OM the sNE does a better job predicting the outcome compared to what is expected given sincere behavior. The distance from the sincere outcome is significantly larger in OM. For M, the sNE and sincere outcomes coincide.

A regression of the group outcome on the sNE or sincere prediction for M results on a coefficient that is not statistically different from one. This confirms that when using the median rule it is the median's ideal point that entirely determines the group's outcome. For treatments SM and M, regressing the outcome on both the sNE and sincere behavior predictions reveals that the group outcome can be viewed as a convex combination of these two variables. The largest weight, 77%, is put on the sNE and the remaining 23% is put on the sincere outcome.

Result 1. *We find support for Hypothesis 1. Group outcomes are largely determined by the sNE prediction in all treatments. In treatments SM and OM where the sincere behavior prediction differs, it has some predictive power, albeit much less than the sNE.*

Treatment	Distance from Nash	Distance from sincere	Distance from median	Distance from center	Distance from mean
SM	4.63	8.45	16.95	15.63	8.45
	$=$	$<^{*}$	$>^{**}$	$<^{**}$	$=$
OM	6.30	9.41	13.26	20.28	8.21
	$>^{*}$	$>^{**}$	$>^{**}$	$<^{**}$	$>^{**}$
M	2.30	2.30	2.30	31.05	11.24

Table 5: **Outcomes vs. Benchmarks.** The numbers indicate the average distance of outcomes from the specified benchmark in each treatment. We compare treatments by regressing the absolute distance on treatment dummies with robust st. errors and test using the wild bootstrap, clustering on the session level (see Roodman et al., 2019). Stars indicate significance levels as follows: p-val<.05: *, p-val<.01: **. Comparing M with OM and SM yield similar results. The distance from Nash, sincere and median in M is not significantly different from zero. All other distances are.

From the fourth and fifth columns of Table 5 we see what is predicted by Theorem 2. On average, outcomes are closest to the median's ideal point in treatment M, followed by OM, with those in SM being the furthest away. At the same time, the outcomes lie between the median and the center of the policy space in the reverse order. This is also evident in Figure 1, despite the noise in the data. Outcomes in the SM treatment tend to be closer to the horizontal line in the center, while the ones in M are closer to the 45 degree line. The outcomes in OM lie mostly between the two. One can conclude that even if the outcomes do not fully conform to the sNE predictions, the comparative statics are robust.

Result 2. *We find strong support for Hypotheses 2a and 2b. Group outcomes lie mostly between the median and the center of the policy space. They are closest to the former in treatment M and to the latter in treatment SM. Outcomes in OM are on average between the other two.*

6.2 Individual votes

We now take a look at how individual subjects vote across treatments. In each treatment, a voter's behavior will largely depend on her position with respect to the other voters in the group. Even before looking at what theory predicts, it is intuitive to think that the median voter is likely to behave differently than the two voters on the extremes or the remaining moderate voters. The theoretical prediction, as discussed previously, is more nuanced as it depends on a voter's position with respect to the mechanism's predicted equilibrium outcome, i.e., the median of all ideal points and the mechanism's "phantom voters". For SM and OM,

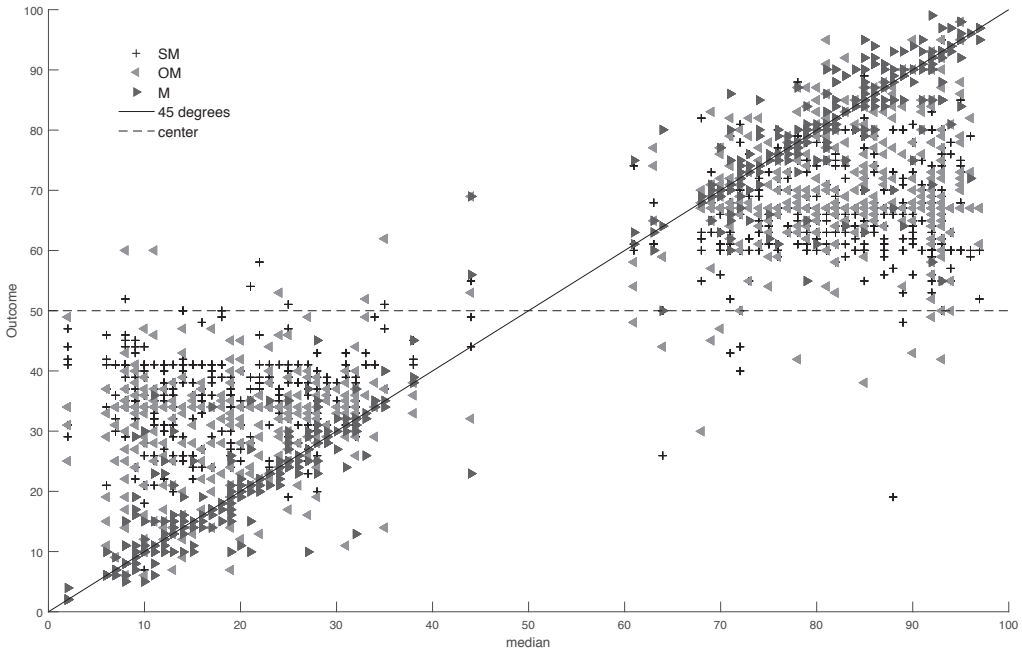

Figure 1: **Outcome vs. median.** Each point in the above scatter plot corresponds to the collective outcome of a group with a given median voter. Crosses correspond to the SM treatment, left-pointing triangles to OM, and right-pointing triangles to M. The solid diagonal line is the 45 degree line. The dashed line indicates the center of the policy space.

in the vast majority of cases, a voter will vote for one of the extremes of the policy space. This will not happen if her ideal policy coincides with the predicted outcome of the mechanisms, in which case she votes in the interior of the policy space. Doing the same in M is also an equilibrium but not the unique one.

The solid black line in each graph in Figure 2 indicates the distribution of the distance of votes from the center of the policy space if voters behave as predicted by the sNE (for M where there are multiple sNE's with a unique outcome it assumes non-median voters will adopt the one with the most extreme behavior). In most cases it is predicted that almost the entirety of votes will lie at one of the extremes of the policy space. One obvious exception is the case of median voters in the M treatment. In any sNE these voters vote for their ideal point and that is the collective outcome of the mechanism. Another special case is that of moderate voters in SM. Due to the particular selection of peak profiles we did for the experiment (see previous section) it is often the case in this mechanism that the equilibrium outcome coincides with a moderate voter's peak. As a result, these voters vote sincerely and do not polarize. As can be seen from the top middle graph in Figure 2, this happens in about a third of cases.

The distribution of actual votes in the experiment is indicated in each graph of Figure 2 by the grey solid line. The first thing to note is that in treatments SM and OM, the majority of votes lie on one of the extremes of the policy space, i.e., have the maximum distance from the center of the policy space. This is not the case for treatment M, but here we observe that median voters' votes are distributed almost identically to what is predicted by the sNE. When this is the case, any vote by moderates and extremists that lies on the same side of the median as their ideal point is a best response. In fact, the dashed line shows the distribution of voters' ideal points' distance from the center of the policy space and this would be the distribution of the distance of votes if everyone voted for his ideal point. We can see that for moderates and extremists in M the distribution of voted is not very different from that.

While the majority of votes does lie on the extremes, there is a significant portion of them that does not. This exceeds what is expected in equilibrium and indicates that while sNE has a lot of power in explaining individual votes, equilibrium play is not the only factor explaining voting behavior.

Result 3. *We find support for Hypothesis 3. Groups tend to be polarized in treatments SM and OM, but less so in M. The mass of votes in the first two treatments lies on the extremes of the policy space.*

Results 1 to 3 demonstrate the power of strong Nash equilibrium theory to predict behavior and outcomes in this environment. Still, there are deviations from equilibrium behavior.

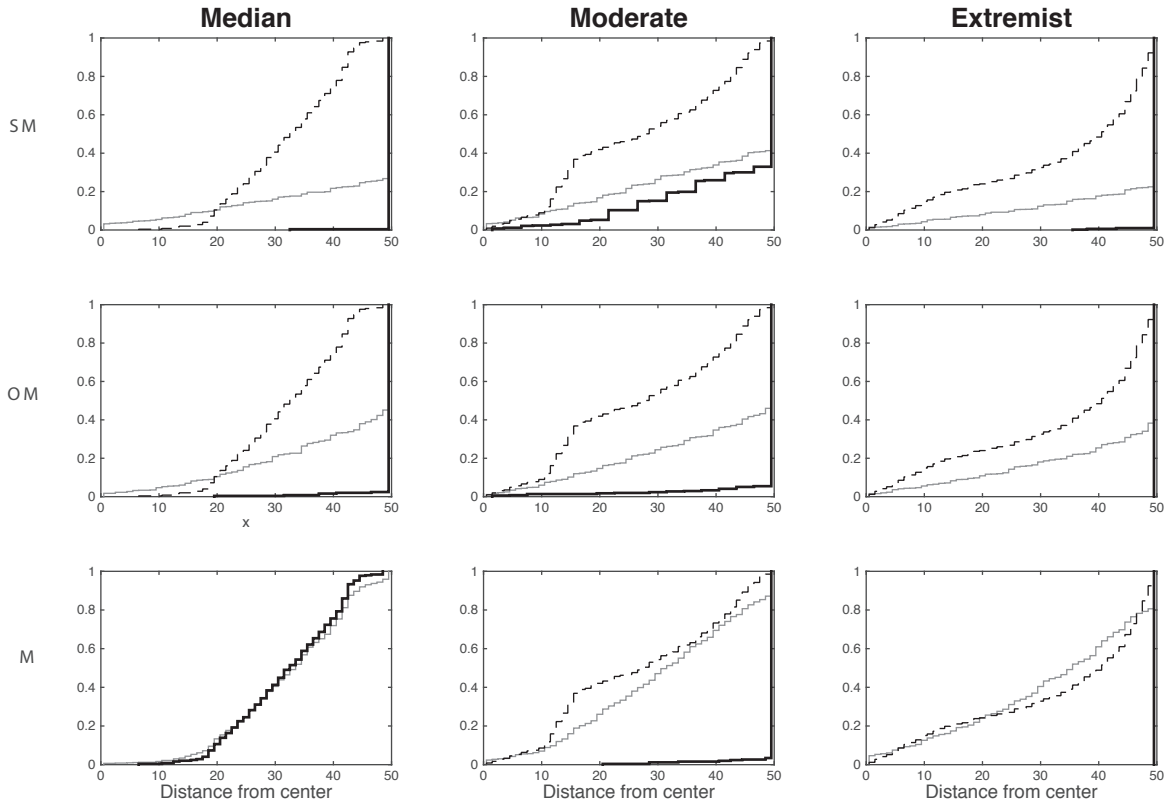

Figure 2: **The distance of individual votes from the center of the policy space.** The grey line in each graph is the empirical cumulative distribution function (CDF) for the distance of votes from the center of the policy space for different types of individual voters (medians, moderates, extremists) and treatments (SM, OM and M), as observed in the experimental data. The solid black lines indicate the CDF that would result from subjects voting according to the sNE. For treatment M we show the CDF for the sNE with the highest degree of polarization. The dashed lines correspond to the CDF of ideal points for each type of voter. This is the same in all treatments and coincides with the distribution of votes under sincere voting.

Such deviations are necessary for our last hypothesis to have some support, as in contrast to the others it does not rely on equilibrium theory.

Hypothesis 4 is motivated by the intuition that higher levels of trimming disincentivize subjects from casting extreme votes. In Table 6 we show the average distance of individuals' votes from their respective ideal points across treatments. According to our hypothesis, this should be lower for higher levels of trimming, namely in treatment M, followed by OM and being highest in SM. Overall, the ranking does conform to the one hypothesized, albeit, the differences between SM and OM are not significant. Looking more closely at the behavior of different types of voters reveals a clear support for the hypothesis for median voters. For moderate voters the average distance is significantly smaller in M compared to the other

Treatment	Overall	Median	Moderate	Extreme
SM	20.28	16.98	22.49	18.68
	$=$	$>^{*}$	$=$	$>^{**}$
OM	19.53	14.27	23.57	16.79
	$>^{**}$	$>^{**}$	$>^{**}$	$=$
M	12.54	2.78	10.40	18.60

Table 6: **Average distance of vote from ideal point.** The numbers indicate the average distance of an individual's vote from her ideal point. The first column shows aggregates per treatment across all voters. The remaining three columns show averages across each type of voter. We compare treatments by regressing the absolute distance on treatment dummies with robust st. errors and test using the wild bootstrap, clustering on the session level (see Roodman et al., 2019). Stars indicate significance levels as follows: p-val<.05: *, p-val<.01: **. Tests comparing SM to M yield the same results as OM vs. M, except for extreme voters where the average distance in SM is not significantly different than the one in M.

treatments. For these voters the highest average distance is observed in OM, although the difference from SM is not significant. To some degree this can be explained by the selection profiles in our experiment: in many cases in SM it is one of the moderate voters determining the outcome in equilibrium, which means that she should not vote one of the extremes (see upper middle panel in Figure 2). If one controls for the distance of the Nash equilibrium prescribed vote from the ideal point the treatment effects are aligned with what we hypothesize, but still the difference between SM and OM is not significant for moderates. Extreme voters cast votes furthest away from their ideal points in SM, but are now followed by their counterparts in M. The difference between the two is not significant. One potential explanation here is that given the mechanism in M, it is highly unlikely for extreme voters to be able to influence the outcome in their favor. This can induce noisy behavior, as voting for anything can be deemed as a best response. The average distance in OM is significantly lower than in SM in line with the reasoning backing this hypothesis.

Result 4. *We find some support for Hypotheses 4. On average, voters in M cast votes closer to their ideal points, followed by voters in OM and voters in SM being the furthest. The effect of trimming on vote sincerity is more pronounced for median voters across all treatments, for moderate voters in M and for extreme voters in OM.*

7 Concluding remarks

In this paper we have provided a full equilibrium analysis of trimmed mean mechanisms and we have also tested several of the theoretical predictions by the means of a laboratory investigation. The experimental test showed that the formal results described the players behavior and decisions well, but also unveiled patterns that have not been pinned down by the formal analysis. Indeed, real subjects are more complex than assumed by the rational choice model and best-responding to the choices of the other players, while prevalent, is not the only aspect that shapes their behavior. Additional empirical analysis from field experiments and observational data seems a natural next step so that further insight can be gained.

Our results were derived in a context of complete information regarding players' preferences and the aggregating process. Despite this being the obvious first step, an analysis of environments of incomplete information is evidently also relevant since in many cases players are not exposed to the biases of the other group members (especially when groups are composed of a large number of individuals), and the aggregation process itself may not be even fully transparent (e.g., when voting is secret and the averaging process is conducted by a third party). Another potentially interesting route is to consider common values since in many cases players do not only need to aggregate their preferences but also their information. This is particularly the case when the committee is composed of experts who need to aggregate their pieces of possibly conflicting evidence to a unique policy proposal. Finally, asymmetries in the weights of the involved parties also make sense to be explored since they become more and more frequent in settings of applied interest (e.g., voting in E.U. decision making bodies). While all these extensions and generalizations are beyond the scope of the present study, they all represent promising research directions for the future and would nicely complement the present analysis.

References

Andrews, D. and F. Hampel (2015): *Robust estimates of location: Survey and advances*, vol. 1280, Princeton University Press.

Aumann, R. (1959): "Acceptable points in general cooperative n-person games," *Contributions to the Theory of Games (AM-40)*, 4, 287–324.

Austen-Smith, D. and J. Banks (1996): "Information aggregation, rationality, and the Condorcet jury theorem," *American political science review*, 90, 34–45.

Bassett Jr, G. and J. Persky (1994): "Rating Skating," *Journal of the American Statistical Association*, 89, 1075–1079.

Bernheim, B. D. and M. Peleg, B.and Whinston (1987): "Coalition-proof nash equilibria i. concepts," *Journal of Economic Theory*, 42, 1–12.

Bickel, P. (1965): "On some robust estimates of location," *The Annals of Mathematical Statistics*, 36, 847–858.

Bryan, M. and S. Cecchetti (1994): "Measuring core inflation," in *Monetary policy*, The University of Chicago Press, 195–219.

Cai, Y., C. Daskalakis, and C. Papadimitriou (2015): "Optimum statistical estimation with strategic data sources," in *Conference on Learning Theory*, 280–296.

Caragiannis, I., A. Procaccia, and N. Shah (2016): "Truthful univariate estimators," in *International Conference on Machine Learning*, 127–135.

Condorcet, N. J. A. (1785): *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix. Par M. le marquis de Condorcet....*, de l'Imprimerie Royale.

Dutta, B. and A. Sen (1991): "Implementation under strong equilibrium: A complete characterization," *Journal of Mathematical Economics*, 20, 49–67.

Eisl, A., R. Jankowitsch, and M. G. Subrahmanyam (2017): "The manipulation potential of Libor and Euribor," *European Financial Management*, 23, 604–647.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental economics*, 10, 171–178.

Galton, F. (1907): "One vote, one value," *Nature*, 75, 414.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

Hill, M. and W. Dixon (1982): "Robustness in real life: a study of clinical laboratory data." *Biometrics*, 38, 377–396.

Huber, P. (1972): "The 1972 wald lecture robust statistics: A review," *The Annals of Mathematical Statistics*, 43, 1041–1067.

Hurley, W. and D. Lior (2002): "Combining expert judgment: On the performance of trimmed mean vote aggregation procedures in the presence of strategic voting," *European Journal of Operational Research*, 140, 142–147.

Lehmann, E. L. and G. Casella (2006): *Theory of point estimation*, Springer Science & Business Media.

Maskin, E. (1978): "Implementation and strong Nash equilibrium," .

Moulin, H. (1980): "On Strategy-proofness and Single Peakedness," *Public Choice*, 35, 437–455.

Renault, R. and A. Trannoy (2005): "Protecting Minorities through the Average Rule," *Journal of Public Economic Theory*, 7, 169–199.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019): "Fast and wild: Bootstrap inference in Stata using boottest," *The Stata Journal*, 19, 4–60.

Rosar, F. (2015): "Continuous decisions by a committee: median versus average mechanisms." *Journal of Economic Theory*, 159- Part A, 15–65.

Rothenberg, T., F. M. Fisher, and C. B. Tilanus (1964): "A note on estimation from a Cauchy sample," *Journal of the American Statistical Association*, 59, 460–463.

Schnitkey, G. (2012): "Simple versus Olympic Averages in Prices used in Farm Commodity Programs," *Farmdoc Daily*, 2.

Stigler, S. (1977): "Do robust estimators work with real data?" *The Annals of Statistics*, 1055–1098.

Yamamura, H. (2011): "On coalitional stability and single peakedness," Tech. rep., mimeo, Kobe University.

Yamamura, H. and R. Kawasaki (2013): "Generalized Average Rules as stable Nash mechanisms to implement generalized median rules," *Social Choice and Welfare*, 40, 815–832.

Yaniv, I. (1997): "Weighting and trimming: Heuristics for aggregating judgments under uncertainty," *Organizational behavior and human decision processes*, 69, 237–249.

A Proofs

Proof of Lemma 1

Proof. Let $n = 5$ (the proof can be extended to any number of agents) and take some profile $x = (c, \ldots, c)$ where every player announces c. It follows that $\theta_\omega(x) = c$ for each $\omega = 0, 1, 2$. If $\omega > 0$, no player has a profitable deviation since any unilateral deviation is removed and hence does not modify the outcome. More precisely, for each $i \in N$, any $s_i' \neq x$ leads to $\theta_\omega(s_i', c, \ldots, c) = c$. Thus, the strategy profile x is an equilibrium for any collection of the peaks. □

Proof of Theorem 1

Proof. The claim is immediate for $\omega = 0$ since the mechanism θ_0 (the Average rule) admits a unique Nash equilibrium outcome $m(p_1, \ldots, p_n, \kappa_1^0, \ldots, \kappa_{n-1}^0)$, as shown by Yamamura and Kawasaki (2013) and any strong Nash equilibrium is a Nash equilibrium by definition. In the sequel, take some mechanism θ_ω with $\omega = 1, \ldots, k$. Consider any collection of peaks $p = (p_1, \ldots, p_n)$ with w.l.o.g. $p_1 \leq p_2 \leq \ldots \leq p_n$. Let $f(p) = m(p_1, \ldots, p_n, \kappa_1^\omega, \ldots, \kappa_{n-1}^\omega)$ the generalized median rule with calibration vector $(\kappa_1^\omega, \ldots, \kappa_{n-1}^\omega)$. To simplify notation, we write κ_j rather than κ_j^ω for each $j = 1, \ldots, n-1$.

The rest of the proof is divided into two steps. Step 1 shows that the mechanism θ_ω admits a strong Nash equilibrium with outcome $f(p)$. Step 2 shows that this outcome is the unique one induced by strong Nash equilibria.

Step 1: The mechanism θ_ω admits a strong Nash equilibrium s with $\theta_\omega(s) = m(p_1, \ldots, p_n, \kappa_1, \ldots, \kappa_{n-1})$.

There are two possible cases: either there is no $i \in N$ for which $p_i = f(p)$ or there is some $i \in N$ with $p_i = f(p)$.

Step 1.a.: There is no $i \in N$ with $p_i = f(p)$. In this case, there is some κ_j with $f(p) = \kappa_j$. Note that $f(p) = \kappa_j$ implies[6] that $\#\{i \in N \mid p_i < \kappa_j\} = n - j$ and $\#\{i \in N \mid p_i > \kappa_j\} = j$. Consider the profile $x \in S^n$ with

$$x_i = 0, \ \forall i \in \{1, \ldots, n-j\} \ \text{ and } x_i = 1, \forall i \in \{n-j+1, \ldots, n\}.$$

[6]Since $(p_1, \ldots, p_n, \kappa_1, \ldots, \kappa_{n-1})$ has $2n-1$ components and κ_j is its median, there are $n-1$ components lower than κ_j and $n-1$ higher than κ_j. Since by construction, $\kappa_j > \kappa_1, \ldots, \kappa_{j-1}$ and $\kappa_j < \kappa_{j+1}, \ldots, \kappa_{n-1}$, it follows that $\#\{i \in N \mid p_i < \kappa_j\} + (j-1) = n-1$ and $\#\{i \in N \mid p_i > \kappa_j\} + (n-1-j) = n-1$. Therefore, $\#\{i \in N \mid p_i < \kappa_j\} = n-j$ and $\#\{i \in N \mid p_i > \kappa_j\} = j$, as wanted.

In the profile x, each player with a peak lower than κ_j plays 0 and each player with peak higher than κ_j plays 1. Moreover, note that the construction of x implies that the following equality holds:

$$\theta_\omega(x) = \text{Average}(\tilde{x}_{\omega+1}, \ldots, \tilde{x}_{n-\omega}) = \frac{j-\omega}{n-2\omega} = \kappa_j,$$

which ensures that the strategy profile x implements the alternative κ_j.

We now prove that x is a Strong Nash equilibrium. Assume by contradiction that there is some coalition of agents $C \subseteq N$ with a profitable deviation.

If such a coalition exists, it cannot include both agents with a peak higher than κ_j and agents with a peak lower than κ_j. To see this, consider first a coalition C that induces an outcome η with $\eta < \kappa_j$. Each player with a peak p_j higher than κ_j is worse-off with this deviation since $\|p_j - \eta\| > \|p_j - \kappa_j\|$ and player j's utility decreases with the distance between his peak and the outcome. Thus, the coalition C can only include agents with a peak lower than κ_j. A symmetric argument applies if the coalition C induces an outcome η with $\eta > \kappa_j$ and shows that such coalition can only include agents with a peak higher than κ_j. It follows that if there is a coalition C with a profitable joint deviation, either $C \subseteq \{i \in N \mid p_i < \kappa_j\}$ or $C \subseteq \{i \in N \mid \kappa_j < p_i\}$.

Consider a deviation by agents in some coalition $C \subseteq \{i \in N \mid p_i < \kappa_j\}$, a symmetric argument applies to coalitions of agents with peaks larger than κ_j. Assuming the coalition $C \subset \{1, \ldots, n-j\}$ has cardinal $m \leq n-j$, we have that:

$$\theta_\omega(x^C, x^{N \setminus C}) = \theta_\omega(x_1^C, x_2^C, \ldots, x_m^C, \underbrace{0, \ldots, 0}_{n-j-m \text{ times}}, \underbrace{1, \ldots, 1}_{j \text{ times}}), \tag{1}$$

where x_i^C denotes the deviation of player i for each $i \in C$. The objective of the coalition C is to select a deviation x_C that minimizes the value of (1) since each player $i \in C$ has a peak $p_i < \kappa_j$. Yet, the minimum of (1) equals $\frac{j-\omega}{n-2\omega}$ and is reached for $x_1^C = x_2^C = \ldots = x_{n-j}^C = 0$. However, since $\theta_\omega(x) = \frac{j-\omega}{n-2\omega}$ by construction, it follows that C has no deviation that induces an outcome lower that $\theta_\omega(x)$, which shows that x is a SNE and finishes the claim for Step 2.a.

Step 1.b: There is some $i \in N$ with $p_i = f(p)$. In this case, there is some $i \in N$ with $p_i = f(p)$. We pick κ_j and p_i with $\kappa_j \leq p_i = f(p)$ such that $j + (i-1) = n-1$ or $j = n-1$, which is possible since p_i is the median of $(p_1, \ldots, p_n, \kappa_1, \ldots, \kappa_{n-1})$. Letting $\kappa_0 = 0$ and $\kappa_n = 1$, we can write that $\kappa_{n-i} \leq p_i \leq \kappa_{n-i+1}$ since again p_i is the median of $(p_1, \ldots, p_n, \kappa_1, \ldots, \kappa_{n-1})$.

Consider the profile $x_{-i} \in S^{n-1}$ such that

$$x_k = 0, \forall k \in \{1, \ldots, i-1\}, \text{ and } x_k = 1, \forall k \in \{i+1, \ldots, n\}.$$

In this profile, each player j with peak $p_j < p_i$ plays 0 and each player j with peak $p_j > p_i$ plays 1. The argument applies verbatim if $i = 1$ by letting $x_k = 1$ $\forall k \in \{2, \ldots, n\}$ and if $i = n$ letting $x_k = 0$ $\forall k \in \{1 \ldots, n-1\}$.

For any $x_i' \in S$, the definition of the trimming mechanism θ_ω implies that

$$\theta_\omega(0, x_{-i}) = \kappa_{n-i} \leq \theta_\omega(x_i', x_{-i}) \leq \kappa_{n-i+1} = \theta_\omega(1, x_{-i}),$$

which jointly with the continuity of θ_ω on a player's message implies that there is some $x_i^* \in S$ with $\theta_\omega(x_i^*, x_{-i}) = p_i$. We now show that $x = (x_i^*, x_{-i})$ is a Strong Nash equilibrium.

In order to do this, we need to show that there is no coalition with a profitable deviation. As in Step 1.a., there is no coalition with agents with peaks both higher and lower than p_i since their objective is opposed. It follows that if there is a coalition C with a profitable joint deviation, either $C \subseteq \{i \in N \mid p_i < \kappa_j\}$ or $C \subseteq \{i \in N \mid p_i > \kappa_j\}$. Consider a deviation by agents in some coalition $C \subseteq \{i \in N \mid p_i < \kappa_j\}$, a symmetric argument applies to coalitions of agents with peak larger than κ_i. The same contradiction as the one described in Step 1.a applies: the minimum of θ_ω that can be reached by the coalition C of agents occurs when $x_i = 0$ for each $i \in C$. Thus, there is no profitable coalitional deviation concluding the proof.

We now prove that the unique outcome that can be reached in a SNE is the median of the peaks and the calibration parameters $\kappa_1, \ldots, \kappa_{n-1}$.

Step 2: The mechanism θ_ω admits $m(p_1, \ldots, p_n, \kappa_1, \ldots, \kappa_{n-1})$ as a unique Strong Nash equilibrium outcome.

Suppose by contradiction that there is some collection of peaks p and some SNE $x = (x_1, \ldots, x_n)$ such that $\theta_\omega(x) > m(p_1, \ldots, p_n, \kappa_1, \ldots, \kappa_{n-1}) = f(p)$. Letting $V \equiv \{i \in N \mid p_i \leq f(p)\}$ and $W \equiv \{j \in \{1, \ldots, n-1\} \mid \kappa_j \leq f(p)\}$, we can then select i^*, j^* in N such that $V = \{1, 2, \ldots, i^*\}$, $W = \{1, 2, \ldots, j^*\}$ with $f(p) = \max\{p_{i^*}, \kappa_{j^*}\}$.

We first show that in any strong equilibrium x, $x_i = 0$ for each player $i \in V$. Assume by contradiction that $x_i > 0$ for some $i \in V$. Note that for each player $i \in V$, $p_i \leq f(p) < \theta_\omega(x)$. As argued in Step 1, each player with peak $p_i \in V$ strictly prefers an outcome η with $p_i < \eta < \theta_\omega(x)$. Thus, each coalition $C \subseteq V$ has a profitable deviation x^C with $x_i^C = 0$ for each $i \in C$ since this strategy uniquely minimizes the value of $\theta_\omega(x^C, x^{N \setminus C})$ with $\omega < k$[7]

[7]Note that when $\omega < k$, the outcome corresponds to the average of $n - 2\omega$ values. Since $\{i \in N \mid p_i < f(p)$ and $s_i \in \tilde{s}_{\omega+1}, \ldots, \tilde{s}_{n-\omega}\}$ is not empty for any $\omega < k$, then any coalition in V must announce 0 in a

This establishes that $x_i = 0$ for each player $i \in V$.

Given that $x_i = 0$ for each player $i \in V$, we can now establish the following inequality

$$\theta_\omega(x) = \theta_\omega(x^V, x^{N\setminus V}) = \theta_\omega(\underbrace{0, \ldots, 0}_{i' \text{ times}}, x^{N\setminus V}) \leq \kappa_{n-i'} \leq f(p). \tag{2}$$

The first inequality in (2) comes from the fact that

$$\max_{x_{N\setminus V} \in S^{n-i'}} \theta_\omega(\underbrace{0, \ldots, 0}_{i' \text{ times}}, x^{N\setminus V}) = \kappa_{n-i'},$$

which is reached when $x_i^{N\setminus V} = 1$ for each $i \in N \setminus V$. The second inequality in (2) stems from the observation that $\#V + \#W = i^* + j^* \geq n \iff n - i^* \leq j^*$ which implies that $\kappa_{n-i^*} \leq \kappa_{j^*} \leq \max\{\kappa_{j^*}, p_{i^*}\} = f(p)$. We have therefore proved that $\theta_\omega(x) > p$ implies that $\theta_\omega(x) \leq f(p)$, entailing the desired contradiction. A similar contradiction arises if we assume that some strong equilibrium x satisfies $\theta_\omega(x) \geq f(p)$. Hence, the unique outcome in a strong equilibrium is $f(p)$, as required. □

Proof of Theorem 2

Proof. For each vector $\alpha = (\alpha_1, \ldots, \alpha_{n-1}) \in [0,1]^{n-1}$, pick the generalized median rule that associates to each profile $p = (p_1, \ldots, p_n)$ the alternative $m(p, \alpha)$. Letting $\tilde{v}$ be the ordered profile of the values in (p_{-i}, α) with $\tilde{v} = (\tilde{v}_i)_{i=1}^{2n-2}$ and $\tilde{v}_1 \leq \tilde{v}_2 \leq \ldots \leq \tilde{v}_{2n-2}$, the outcome of $m(p_i, p_{-i}, \alpha)$ equals:

$$m(p, \alpha) = m(p_i, \tilde{v}_{n-1}, \tilde{v}_n). \tag{3}$$

Equipped with this restatement of the median rule, take two trimming mechanisms θ_ω and $\theta_{\omega'}$ with $\omega < \omega'$ and associated vectors κ^ω and $\kappa^{\omega'}$. It can be checked that, for each $j = 1, \ldots, n-1$,

$$\kappa_j^\omega < \frac{1}{2} \Longrightarrow \kappa_j^{\omega'} \leq \kappa_j^\omega \text{ and } \kappa_j^\omega > \frac{1}{2} \Longrightarrow \kappa_j^{\omega'} \geq \kappa_j^\omega.$$

Proposition 1 proves that for any pair $m(p, \alpha)$ and $m(p, \beta)$ of generalized medians with vectors α and β such that, for each $j = 1, \ldots, n-1$,

$$\beta_j \leq \frac{1}{2} \Longrightarrow \alpha_j \leq \beta_j \text{ and?} \beta_j \geq \frac{1}{2} \Longrightarrow \alpha_j \geq \beta_j,$$

sNE.

then:

$$m(p,\beta) \leq 1/2 \Longrightarrow m(p,\alpha) \leq m(p,\beta) \text{ and } m(p,\beta) \geq 1/2 \Longrightarrow m(p,\alpha) \geq m(p,\beta).$$

Yet, this directly concludes the proof since it implies that

$$m(p,\kappa^{\omega}) \leq 1/2 \Longrightarrow m(p) \leq m(p,\kappa^{\omega'}) \leq m(p,\kappa^{\omega}) \leq \frac{1}{2} \text{ and}$$
$$m(p,\kappa^{\omega}) \geq 1/2 \Longrightarrow m(p) \geq m(p,\kappa^{\omega'}) \geq m(p,\kappa^{\omega}) \geq \frac{1}{2},$$

where $m(p) = m(p,\alpha)$ with $\alpha_j = 0$ for $i = 1,\ldots,\frac{n-1}{2}$ and $\alpha_j = 1$ for $i = \frac{n-1}{2},\ldots,n-1$ (n is odd and hence $\frac{n-1}{2}$ is well-defined). □

Proposition 1. *Let α and β be two vectors in $[0,1]^{n-1}$ such that for each $j = 1,\ldots,n-1$,*

$$\beta_j \leq \frac{1}{2} \Longrightarrow \alpha_j \leq \beta_j \text{ and } \beta_j \geq \frac{1}{2} \Longrightarrow \alpha_j \geq \beta_j.$$

Then, for any $p \in [0,1]^n$:

$$m(p,\beta) \leq 1/2 \Longrightarrow m(p,\alpha) \leq m(p,\beta) \text{ and } m(p,\beta) \geq 1/2 \Longrightarrow m(p,\alpha) \geq m(p,\beta).$$

Proof. For each $p_i \in [0,1]$ and each $p_{-i} \in [0,1]^{n-1}$, consider the vectors $\tilde{v}$ and $\tilde{w}$ the ordered vectors associated to (α, p_{-i}) and to (β, p_{-i}) respectively. Using (3), it suffices to show that for each $i \in N$, each $p_i \in [0,1]$ and each $p_{-i} \in [0,1]^{n-1}$,

$$m(p_i,\tilde{w}_{n-1},\tilde{w}_n) \leq \frac{1}{2} \Longleftrightarrow m(p_i,\tilde{v}_{n-1},\tilde{v}_n) \leq m(p_i,\tilde{w}_{n-1},\tilde{w}_n).$$

Since by assumption, the vectors α and β satisfy $\beta_j \leq \frac{1}{2} \Longrightarrow \alpha_j \leq \beta_j$ and? $\beta_j \geq \frac{1}{2} \Longrightarrow \alpha_j \geq \beta_j$ for each $j = 1,\ldots,n-1$, it follows that each $h = 1,\ldots,2n-1$, the following implications hold:

$$\tilde{w}_h \leq \frac{1}{2} \Longrightarrow \tilde{v}_h \leq \tilde{w}_h \text{ and } \tilde{w}_h \geq \frac{1}{2} \Longrightarrow \tilde{v}_h \geq \tilde{w}_h. \tag{4}$$

Suppose that $m(p,\beta) \leq 1/2$, the same logic applies if $m(p,\beta) \geq \frac{1}{2}$ by symmetry.

a. If $\tilde{w}_n \leq \frac{1}{2}$, then $\tilde{w}_{n-1} \leq \frac{1}{2}$ since $\tilde{w}$ is an ordered vector. Therefore, (4) implies both that $\tilde{v}_{n-1} \leq \tilde{w}_{n-1}$ and $\tilde{v}_n \leq \tilde{w}_n$. Thus, Lemma 1 implies that $m(p_i,\tilde{v}_{n-1},\tilde{v}_n) \leq m(p_i,\tilde{w}_{n-1},\tilde{w}_n) \Longleftrightarrow m(p,\alpha) \leq m(p,\beta)$ as wanted.
b. If $\tilde{w}_{n-1} \leq \frac{1}{2} \leq \tilde{w}_n$, then (4) implies that $\tilde{v}_{n-1} \leq \tilde{w}_{n-1} \leq \frac{1}{2} \leq \tilde{w}_n \leq \tilde{v}_n$. It follows that according to Lemma 2, $m(p_i,\tilde{v}_{n-1},\tilde{v}_n) \leq m(p_i,\tilde{w}_{n-1},\tilde{w}_n) \Longleftrightarrow m(p,\alpha) \leq m(p,\beta)$ as required.

c. Finally, if $\frac{1}{2} \leq \tilde{w}_{n-1} \leq \tilde{w}_n$, $m(p_i, \tilde{w}_{n-1}, \tilde{w}_n) \geq \frac{1}{2}$ so that there is a contradiction with $m(p, \alpha) \leq \frac{1}{2}$. This concludes the proof. □

Lemma 1: For each pair of vectors $(\alpha_1, \alpha_2), (\beta_1, \beta_2) \in [0,1]^2$ with $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ and $0 \leq \beta_1 \leq \beta_2 \leq 1$, if $0 \leq \alpha_1 \leq \beta_1$ and $0 \leq \alpha_2 \leq \beta_2 \Longrightarrow m(x, \alpha_1, \alpha_2) \leq m(x, \beta_1, \beta_2)$.

Proof. There are two cases: either $\alpha_1 \leq \alpha_2 \leq \beta_1 \leq \beta_2$ or $\alpha_1 \leq \beta_1 \leq \alpha_2 \leq \beta_1$.

Assume first that $\alpha_1 \leq \alpha_2 \leq \beta_1 \leq \beta_2$. In this case, the claim is immediate since $\min_{x\in[0,1]} m(x, \beta_1, \beta_2) \geq \max_{x\in[0,1]} m(x, \alpha_1, \alpha_2)$.

Assume now that $\alpha_1 \leq \beta_1 \leq \alpha_2 \leq \beta_1$. In this case, we write for each $x \in [0,1]$,

$$m(x, \alpha_1, \alpha_2) = \begin{cases} \alpha_1 & \text{if } x < \alpha_1 \\ x & \text{if } \alpha_1 \leq x \leq \alpha_2 \\ \alpha_2 & \text{if } x > \alpha_2 \end{cases} \text{ and } m(x, \beta_1, \beta_2) = \begin{cases} \beta_1 & \text{if } x < \beta_1 \\ x & \text{if } \beta_1 \leq x \leq \beta_2 \\ \beta_2 & \text{if } x > \beta_2 \end{cases}$$

which implies that

$$m(x, \alpha_1, \alpha_2) - m(x, \beta_1, \beta_2) = \begin{cases} \alpha_1 - \beta_1 & \text{if } x < \alpha_1 \\ x - \beta_1 & \text{if } \alpha_1 \leq x \leq \beta_1 \\ 0 & \text{if } \beta_1 \leq x \leq \beta_2 \\ \alpha_2 - x & \text{if } \alpha_2 \leq x \leq \beta_2 \\ \alpha_2 - \beta_2 & \text{if } x > \beta_2 \end{cases}$$

The previous statement of the function $m(x, \alpha_1, \alpha_2) - m(x, \beta_1, \beta_2)$ directly implies that $m(x, \alpha_1, \alpha_2) \leq m(x, \beta_1, \beta_2)$ for each $x \in [0,1]$, as wanted. □

Lemma 2: For each pair of vectors $(\alpha_1, \alpha_2), (\beta_1, \beta_2) \in [0,1]^2$ with $0 \leq \alpha_1 \leq \beta_1 \leq \frac{1}{2} \leq \beta_2 \leq \alpha_2 \leq 1$:
· if $m(x, \beta_1, \beta_2) \leq \frac{1}{2}$ then, for each $x \in [0,1]$, $m(x, \alpha_1, \alpha_2) \leq m(x, \beta_1, \beta_2)$
· if $m(x, \beta_1, \beta_2) \geq \frac{1}{2}$ then, for each $x \in [0,1]$, $m(x, \alpha_1, \alpha_2) \geq m(x, \beta_1, \beta_2)$.

Proof. By the definition of the median function,

$$m(x, \alpha_1, \alpha_2) = \begin{cases} \alpha_1 & \text{if } x < \alpha_1 \\ x & \text{if } \alpha_1 \leq x \leq \alpha_2 \\ \alpha_2 & \text{if } x > \alpha_2 \end{cases} \text{ and } m(x, \beta_1, \beta_2) = \begin{cases} \beta_1 & \text{if } x < \beta_1 \\ x & \text{if } \beta_1 \leq x \leq \beta_2 \\ \beta_2 & \text{if } x > \beta_2 \end{cases}$$

Since $0 \leq \alpha_1 < \beta_1 < \frac{1}{2} < \beta_2 < \alpha_2 \leq 1$, it follows that

$$m(x, \alpha_1, \alpha_2) - m(x, \beta_1, \beta_2) = \begin{cases} \alpha_1 - \beta_1 & \text{if } x < \alpha_1 \\ x - \beta_1 & \text{if } \alpha_1 \leq x \leq \beta_1 \\ 0 & \text{if } \beta_1 \leq x \leq \beta_2 \\ x - \beta_2 & \text{if } \beta_2 \leq x \leq \alpha_2 \\ \alpha_2 - \beta_2 & \text{if } x > \alpha_2 \end{cases}$$

Note that, given the conditions of the parameters, if $m(x, \beta_1, \beta_2) \leq \frac{1}{2}$ then $x \leq \frac{1}{2}$ and thus $m(x, \alpha_1, \alpha_2) \leq m(x, \beta_1, \beta_2)$. Similarly, if $m(x, \beta_1, \beta_2) \geq \frac{1}{2}$ then $x \geq \frac{1}{2}$ which in turn implies that $m(x, \alpha_1, \alpha_2) \geq m(x, \beta_1, \beta_2)$, as wanted. □